Cityscape

A Journal of Policy Development and Research

Local Data for Local Action
Volume 26, Number 1 • 2024

U.S. Department of Housing and Urban Development Office of Policy Development and Research

Cityscape

A Journal of Policy Development and Research

Local Data for Local Action
Volume 26, Number 1 • 2024

U.S. Department of Housing and Urban Development
Office of Policy Development and Research

The goal of *Cityscape* is to bring high-quality original research on housing and community development issues to scholars, government officials, and practitioners. *Cityscape* is open to all relevant disciplines, including architecture, consumer research, demography, economics, engineering, ethnography, finance, geography, law, planning, political science, public policy, regional science, sociology, statistics, and urban studies.

Cityscape is published three times a year by the Office of Policy Development and Research (PD&R) of the U.S. Department of Housing and Urban Development (HUD). Subscriptions are available at no charge and single copies at a nominal fee. The journal is also available on line at huduser.gov/periodicals/cityscape.html.

PD&R welcomes submissions to the Refereed Papers section of the journal. Our referee process is double blind and timely, and our referees are highly qualified. The managing editor will also respond to authors who submit outlines of proposed papers regarding the suitability of those proposals for inclusion in *Cityscape*. Send manuscripts or outlines to cityscape@hud.gov.

Opinions expressed in the articles are those of the authors and do not necessarily reflect the views and policies of HUD or the U.S. Government.

Visit PD&R's website, huduser.gov, to find this report and others sponsored by PD&R. Other services of HUD USER, PD&R's Research and Information Service, include listservs, special interest and bimonthly publications (best practices, significant studies from other sources), access to public use databases, and a hotline (1–800–245–2691) for help with accessing the information you need.

Contents

Symposium

Local Data for Local Action

Guest Editor: Amy O'Hara

Guest Editor's Introduction

Novel Uses of Administrative Data for Policymaking

Amy O'Hara
Georgetown University

Support for this project was provided by the Robert Wood Johnson Foundation. The views expressed here do not necessarily reflect the views of the Foundation.

From property tax filings and building permits to home sales and rental listings, masses of data are collected by governments and commercial organizations every day that could inform local housing solutions. By analyzing administrative data and using them to develop place-based indicators—which measure impact within a city, rural community, neighborhood, or even ZIP Code area—communities can fill knowledge gaps and surface new findings to guide housing policy and program design.

This *Cityscape* symposium highlights innovative approaches to using administrative data in local housing policy and program design. These articles trace the causes and consequences of actions by individuals—including property owners and residents—on disinvestment in certain communities, the disappearance of housing units of a certain size, and other changes in housing stock availability. Changes in federal assistance, such as the use of vouchers and tax credits, are also captured in this issue. Case studies reveal how local communities reuse administrative data to guide housing policy decisions. Across the board, these articles shed new light on how housing, health, education, safety, and natural disasters are interrelated.

The articles in this issue are grouped into the themes defined in the following headings.

Linkages with Policy Impact

These articles link various administrative datasets to generate new findings with policy impacts. Coulton et al. link various Cuyahoga County, Ohio, and Cleveland, its county seat, data sources to profile the rental market relative to the risk of lead exposure and to assess landlord capability of meeting lead safety standards. Demonstrating the utility of Department of Housing and Urban Development (HUD) administrative data, Shcheglovitova and Lee link Federal Emergency Management Agency (FEMA) data to HUD rental assistance and Kansas City, Missouri, parcel data to assess flood risk. Garrison et al. link the National Center for Health Statistics (NCHS) with HUD data on assisted renters to examine housing as a social determinant of health. Curtis, Paulsen, and Shager describe how linkages across multiple Wisconsin state-level data increase the availability of quality data in the state. Ghorbani et al. create a Neighborhood Resource Index based on publicly available data to examine the needs for and placement of affordable housing units in New York State. These articles show how administrative data can reveal housing insights to better serve communities.

Ownership and Displacement

This section explores the impacts that displacement and disinvestment have on housing markets. Duda, Smith, and Jiao use Chicago parcel-level data to examine the loss of two- to four-unit buildings across the city, and Greenberg et al. link data on mortgage transactions, sales prices, housing maintenance violations, and marshals' evictions with affordable housing investments to analyze to measure how speculative finance affects communities and quality of life across New York City.

Evictions

This section centers on the factors that influence eviction outcomes for tenants. Griswold et al. use eviction records to examine the relationship between crime-free housing policies and evictions in four California cities. Thomas et al. describe a method that uses natural language processing to mine court record images to digitize eviction case records and a process to geocode addresses in a case study using Washington State data so they can answer questions about the characteristics of renters and units in eviction cases. Ellen, Lochhead, and O'Regan explore evictions in New York State subsidized housing using city, state, and federal administrative data sources.

Developing and Improving Datasets

Improving existing datasets through automation, analytical modeling, and machine learning can prepare new data resources that spur research across the country. Zhu, Neal, and Young analyze racial equity aspects of automated valuation models, offering a method to audit their performance based on neighborhood demographics. Deitz et al. demonstrate a method of placing subsidized housing units into tax parcels using New Jersey data to reduce overcounting subsidized units and improve knowledge about subsidy depth and duration. Torres clusters Florida property owners using tax parcel and business registry data in a graph data structure to examine spatial ownership patterns of single-family housing.

These articles will encourage greater awareness and use of administrative data for research and development of place-based indicators. Please share more administrative data linkage successes and challenges by emailing amy.ohara@georgetown.edu.

Acknowledgments

This *Cityscape* symposium is a part of the Georgetown University Massive Data Institute's Place-Based Indicators Project. Special thanks to Rosemary Rhodes and Gabriel Taylor for helping this issue come to fruition.

Guest Editor

Amy O'Hara is a research professor in the Massive Data Institute at the McCourt School of Public Policy at Georgetown University.

Linkages with Policy Impacts

Using Linked Administrative Data to Profile a City's Rental Stock and Landlords and Guide a Lead-Safe Housing Initiative

Merging Federal Flooding and Housing Data to Gain Insight into Flood Impacts on Federally Assisted Households: A Case Study in Kansas City, Missouri

The Health Status of Women with Children Living in Public and Assisted Housing: Linkage of the National Health Interview Survey to U.S. Department of Housing and Urban Development Administrative Data

Building a Transformational Data Resource to Support Housing Research: The Wisconsin Experience

Promoting Affordable Housing in Well-Resourced Neighborhoods: A Regional Approach to Assessing Neighborhood Resources in New York State

Using Administrative Data Linkage to Drive Homelessness Policy: Experiences From Wales

Using Linked Administrative Data to Profile a City's Rental Stock and Landlords and Guide a Lead-Safe Housing Initiative

Claudia Coulton
Michael Henderson
Francisca García-Cobián Richter
Jeesoo Jeon
April Urban
Michael Schramm
Robert L. Fischer
Case Western Reserve University

Abstract

By the time they enter kindergarten, an estimated 25 percent of Cleveland, Ohio, children have at least one test showing an elevated blood lead level, and to address this high rate of lead exposure at its source, the city committed to a lead-safe housing strategy. Most families with young children rent homes in the private rental market, making these properties and their owners key elements in moving forward on a lead-safe agenda. This article describes how parcel data, property tax rolls, deed and foreclosure records, housing code violations, rental registry information, building permits, evictions, and Housing Choice Voucher program records were used to evaluate lead risk in the rental housing stock and develop a typology of landlords. Deterministic and probabilistic methods were used to link the property data sources, resulting in the identification of 103,386 rental units, 54,786 rental properties, and 36,659 landlords for the analysis. More than one-third of the rental properties were found to be at high risk of failing to meet lead safety standards. A latent class analysis uncovered three classes of landlords, characterized as having different capabilities to comply with the lead safety ordinance. Small, under-resourced landlords who would likely require the highest level of support from the lead safety coalition owned approximately 25 percent of the rental properties. This study guided the lead-safe Cleveland strategy and is being updated to evaluate progress toward reducing lead hazards in rental housing.

Introduction

Prevention of lead exposure in young children requires action at the intersection of the health, housing, and regulatory systems. In older cities, much of the affordable housing stock carries a significant risk of lead exposure due to its age, deferred maintenance, and low market value (Shaw, 2004). Moreover, less than one-fourth of low-income families nationally live in public or subsidized housing units (Kingsley, 2017), but families with young children seeking housing in the low-cost private rental market face limited choices and leverage when it comes to their selection of housing units, and many families have difficulty finding affordable rental housing. African-American children are disproportionately exposed to lead in their homes, in large part due to historical patterns of redlining and discriminatory housing policy (Rothstein, 2017; Sampson and Winter, 2016). Prolonged disinvestment and lack of maintenance in the affordable housing stock are key factors contributing to persistent racial and socioeconomic disparities in lead exposure among young children.

The challenges of bringing privately owned rental housing up to health and safety standards are considerable. In older industrial cities, small buildings owned by individual rather than corporate landlords tend to dominate the affordable rental market. Families with children often seek out this individually owned segment of the rental market due to the preponderance of single- and two-family structures. Although little is known about the rental inventory and business practices of small landlords, several studies attest to the significant size of this sector and to its continued growth (Messamore, 2023). Other research suggests that personal perceptions and social networks often uniquely influence small landlords in the choices they make with respect to their properties and tenants (Garboden and Newman, 2012; Gomory, 2022; Greif, 2018; Shiffer-Sebba, 2020).

Bringing rental housing up to lead-safe standards was an important objective when leaders in Cleveland, Ohio, established a comprehensive approach to protecting children from lead exposure in their homes (City of Cleveland, 2019; Lead Safe Cleveland Coalition, 2023). Longitudinal research demonstrated the costly long-term effects of lead exposure on Cleveland children in the form of substantially higher involvement in public systems and lower educational attainment later in life compared with their peers (Coulton et al., 2023). Studies elsewhere confirmed that lead dust in the residential environment was the primary source of elevated blood lead levels in children, and the risk of exposure was greatest in low-quality rental housing (Lanphear, Hornung, and Ho, 2005). Lead exposure rates among children in Cleveland have been high for many years, and those rates were highest in neighborhoods where housing has been vacant or subject to disinvestment (Fischer, Steh, and Chan, 2018). However, little systematic information existed about the inventory and segmentation of the local rental market or the business models of the landlords in this space. Because these properties and their owners were key elements for achieving the goal of lead-safe housing in Cleveland, information on this sector was essential to guide the strategies for bringing rental housing up to lead-safe standards and ultimately eliminating the risk of lead exposure in children.

This article describes how multiple administrative data sources were organized, cleaned, and linked to characterize two units of analysis: Cleveland rental properties built before 1978 and their landlords. It reports on the resulting inventory and segmentation of the local rental market and the business models of the landlords in this space and demonstrates that, when such local data sources are combined, they have the potential to inform cross-sector initiatives, such as the one undertaken in Cleveland, to achieve lead safety for children. Drawing on these findings, this article discusses the value of such information and the strengths and limitations of such data and methods.

Methods

Prior to implementing the lead-safe housing initiative in Cleveland, it was important to gain a comprehensive understanding of rental housing and the ownership patterns of rental units. Such information was essential to establishing realistic plans for implementation, including estimating the resources needed for inspections and repairs and establishing processes for engaging landlords and community partners in the program. However, the data that were needed to profile rental properties and landlords along the many dimensions relevant to lead safety were not available from any one source. Instead, the data were spread across several agencies, each with its own record systems supporting its own administrative responsibilities.

Study Area

This article focuses on rental housing in the city of Cleveland, the principal city within Cuyahoga County, Ohio. Cleveland has a strong cultural and industrial history, and like similar metropolitan areas, deindustrialization and population loss have taken a toll. Embedded within a large metropolitan area on Lake Erie, it occupies 77 square miles of land area. The population estimate for Cleveland in 2021 was 368,006, down from a peak population in 1950 of 914,808. The relative age of Cleveland's housing stock and the limited income of residents are contributors to the prevalence of lead risk in the rental housing stock. In 2021, the poverty rate in Cleveland was estimated at 29.3 percent, and the median household income was $35,562 (U.S. Census Bureau, 2021a). Much of the housing stock predates the 1978 laws prohibiting lead-based paint, with 59.1 percent of units built before 1950 and 88.8 percent built before 1980, according to the 2021 American Community Survey (U.S. Census Bureau, 2021b). Rental housing predominates in the city, with 60.9 percent of occupied units being rentals.

Data Sources and Preparation

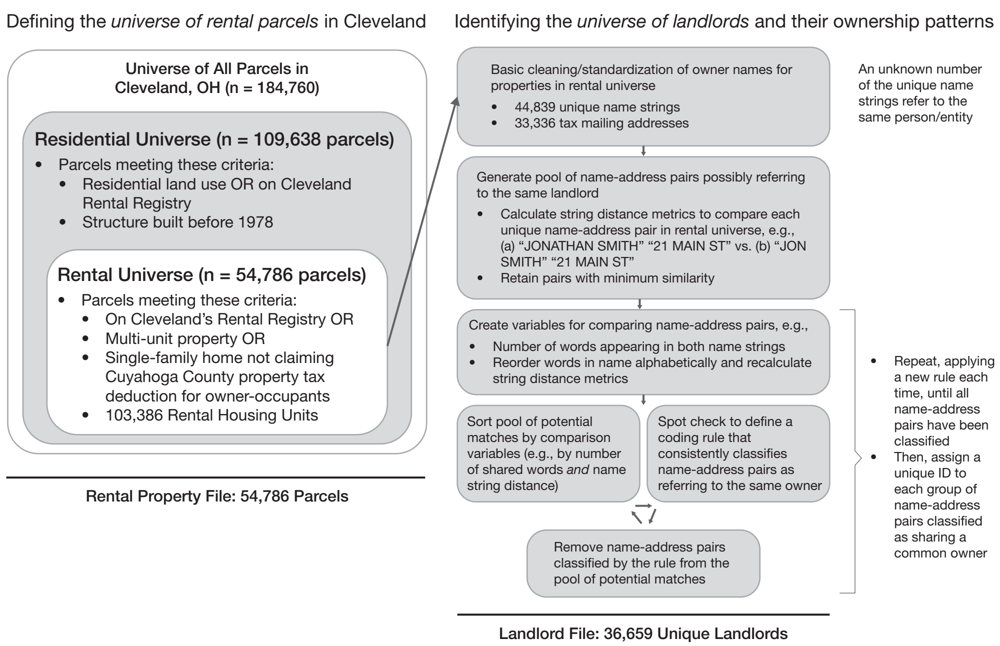
This study focuses on privately owned rental housing in Cleveland that, by being built before 1978 when lead paint was outlawed, presents a potential risk for lead exposure. It also looks at the owners of rental properties in Cleveland to determine the locations and characteristics of their holdings. This article presents two units of analysis. First, it focuses on the universe of properties known or suspected to be active in the Cleveland rental market in 2018. The process for identifying rental properties relied on clues found in administrative records and is discussed in the following sections. After identifying these rental properties in Cleveland, this article turns to their owners as a second unit of analysis.

This research uses a variety of public records to identify the Cleveland rental properties and describe their physical and market conditions. To optimize the coverage and accuracy of this research, the research team undertook an elaborate process of acquiring public records, learning about their content, linking them together by property and owner identifiers, and evaluating data quality. Some records came from Cleveland municipal agencies, and others came from the county government or specialized district agencies. For example, the team relied on rental registration records from the City of Cleveland and property characteristics data from the Cuyahoga County Fiscal Office to identify the full universe of rental properties in Cleveland.

The main data source was the Cuyahoga County Fiscal Office Master Annual Property File, which includes detailed information about every parcel in Cleveland, including property characteristics, conditions, values, ownership, foreclosures, and tax delinquency. In addition, the Cleveland Department of Building and Housing supplied rental registration and code violation information, the Cuyahoga Metropolitan Housing Authority supplied Housing Choice Voucher (HCV) program data, the U.S. Postal Service supplied Postal Vacancy records, the Cuyahoga County Land Bank supplied records on demolition used to refine the rental universe, and the Cleveland Housing Court supplied eviction filings data. The following sections discuss the integration of these varied pieces of information to create two files that contain the variables needed for the analysis. Exhibit 1 provides a visual representation of the processes for reshaping this integrated dataset into two analysis datasets: one organized at the property level for investigating the rental market and the other organized at the landlord level to analyze ownership patterns.

Exhibit 1

Steps in Building Rental Property and Linked Landlord Data Sets



Sources: Cuyahoga County Fiscal Office; City of Cleveland Department of Building and Housing

Rental Property File

The research team began by building a file that included every residential parcel in the city of Cleveland during the 3-year period from 2016 to 2018. The team standardized the addresses associated with the parcels so they could be linked to other data sources based on the address and geocoded for mapping or other geographic linkages. For each parcel, the team imported owner names, tax mailing addresses, owner occupancy tax credit status, property characteristics, assessed market values, recent sales prices, foreclosure and tax histories, housing code violations, building permits, building condition ratings, rental registry dates (if any), postal vacancy spells, eviction filings, and whether HCVs had been used. Because most of this information was time-dependent, the team specified whether the record referred to a particular time point or reflected an event occurring within a given period (for example, a code violation in the previous 3 years).

To identify properties that were likely to be rentals, the team used multiple criteria. A residential property was considered for inclusion in the rental universe if (1) it was in the City of Cleveland's rental registry in 2018,[1] (2) the property owner did not claim the owner occupancy tax credit in 2018, or (3) property records indicated that the parcel contained more than one housing unit. The exception to this rule was that if the owner of a two- or three-unit property claimed an owner occupancy tax credit, it was assumed that the property owner occupied one of the units. From this pool of potential rentals, the team then excluded properties for the following disqualifying reasons: a demolition in 2018, Cuyahoga Land Bank ownership, state forfeiture, or long-term vacancy reported in postal records. The team further restricted the file to properties built before 1978 (the year lead paint was outlawed) and excluded public housing authority-owned properties, arriving at a final rental universe of 54,786 properties corresponding to 103,386 rental units.

Landlord File

The focus here was on creating a database with information on the private owners of rental properties. Creating such a database is not a straightforward process, given that local administrative data sources with relevant information about landlords are typically organized around parcels or addresses rather than landlords. Creating a database of landlords necessitated reshaping the parcel-level property file into a landlord-level file consisting of a single record for every unique landlord in Cleveland, with information about the numbers, locations, and characteristics of their properties. The two pieces of information available for constructing the landlord-level database were the names and tax mailing addresses for the owners of all likely rental properties recorded in the Cuyahoga County Fiscal Office tax billing file. Before carrying out this process, it was necessary to clean and standardize the names and addresses.[2]

Once cleaned, the research team undertook a multistep process to uncover the patterns of property ownership, even if the owner names varied somewhat. The first step in the process was to compare

[1] Although Cleveland had a rental registry in 2018, many landlords did not register. Thus, this analysis used additional indicators to identify rental properties.

[2] Examples of the cleaning and standardization procedures include making all text uppercase, removing inconsistently used words (for example, street suffixes, which may be spelled out as "STREET," abbreviated as "ST," or absent) and punctuations (for example, removing the period in "JR."), and applying consistent naming conventions (for example, changing "WEST 25TH" to "W 25" or "TRS" to "TRUSTEES").

the owner's name and tax mailing address of every parcel in the rental universe with the names and addresses of every other rental parcel. The team used string distance metrics from the *R* package *stringdist* to flag cases with at least a superficial similarity between the comparison of owner names or addresses of the two parcels, generating an initial pool of potentially matching parcel pairs to investigate more closely to determine common ownership (van der Loo, 2014).

The team then created many additional helper variables to better assess the degree and nature of the similarities between each pair of owners. Some examples of these helper variables include the number of words appearing in both owner names, the uniqueness of any shared words, the number of initials in common, the string distance between each name after removing any small one- to two-character words from the name strings, and the string distance after first sorting the words in each name string alphabetically (for example, "SMITH, JOHN JR" would be alphabetized as "JOHN JR SMITH").

An iterative process to classify each pair of parcels as a match (that is, owned by the same person or entity) or a nonmatch followed this step. First, the team sorted the pool of potential matches by one or more helper variables and then examined the names and addresses at the top and bottom of the sorted data. If the researchers found the pairs of owner names at either end of the sorted data to be consistently matching or nonmatching, they continued to scan up or down until arriving at a point where the pattern broke down, and matches, nonmatches, and unclear cases began intermixing. At this breakpoint, the team used the values of the helper variables on which the data were sorted to define a new condition in the code for classifying pairs of owners as matches or nonmatches. The researchers completed each iteration by filtering the pool of potential matches by the new condition, leaving behind only the still-unclassified pairs of parcels. They then began a new iteration on the remaining cases, sorting on a new combination of helper variables and repeating the process until all pairs of parcels were classified. After classifying the entire pool of potential matches, the researchers retained the pairs of parcels determined to share a common owner and attached an owner identification to all their properties.

This process allowed the research team to compute variables reflecting landlord characteristics by grouping all an individual's properties and using the information available in the property file described previously. Owners were classified as "persons" if their property deeds contained names of individuals or as "corporate" if their deeds had company or organizational names. They were also classified as being in the local area (defined as Cuyahoga County) or out of town. Similarly, for each landlord, the team computed the number and types of properties and units owned, the average condition ratings and market values of their properties, markers of financial vulnerability (for example, tax delinquencies and foreclosure sales), and participation in the rental registry or HCV program.

These steps resulted in a landlord data file comprising a deduplicated list of 36,659 owners along with summary measures reflecting their rental properties in 2018. However, it is important to note that the research team probably missed some duplicates. If an individual owned properties under the names of several limited liability companies (LLCs) or used various unrelated owner names and addresses, the team's deduplication algorithms may have missed these matches. However, investigating interlocking ownership of businesses or personal relationships not evident in the titling of the property was beyond the scope of this study.

Data Analysis

Data analysis proceeded in two phases. First, the research team explored the characteristics of rental properties and landlords through descriptive statistics, focusing on characteristics that had implications for designing the lead-safe strategy for Cleveland. These characteristics included markers of distressed housing conditions, low property values, limited landlord capacity and connectedness to systems, and business structure. Counts and percentages were reported for the city as a whole and broken down by neighborhood to facilitate planning.

Second, the team employed latent class analysis (LCA) using the data science software Stata to identify classes of landlords based on the characteristics of their rental properties and the size of their portfolios (Stata Press, 2023). LCA is a statistical technique that classifies cases into a specified number of groups or types such that the similarities within types and differences between types are maximized. It should be noted that not all landlords within a classification group will be identical, nor will all the groups differ on every condition. LCA instead finds the distinctions that best account for the patterns in the descriptive information, which are then used to interpret the meaning of the typology.

Results

The analyses presented in this section focus on three domains: the rental property file of privately owned properties built before 1978 (n = 54,786), the rental units within these properties (n = 103,386), and the unique individuals or companies that owned these rental properties (n = 36,659).

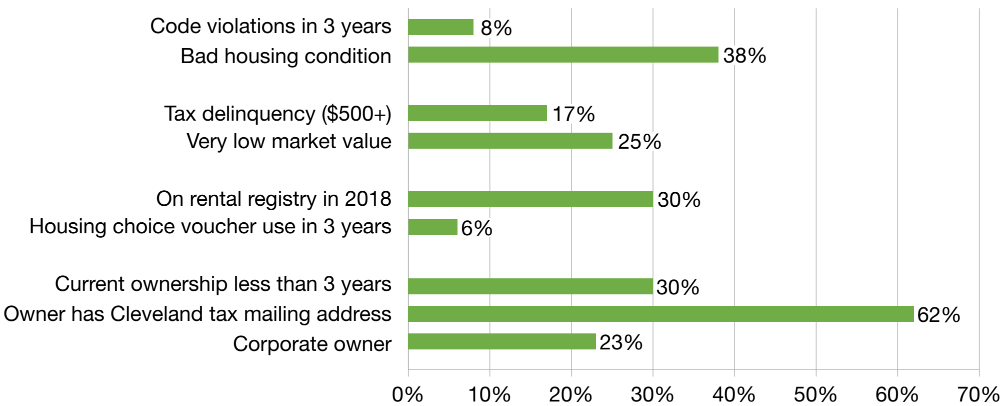
Description of the Rental Universe

This descriptive analysis reports on selected characteristics of rental properties expected to inform the implementation of lead-safe strategies in Cleveland. The analysis first focuses on building type, because detached units would require a different remediation strategy than large apartment buildings. During the study period (2018), single-family structures were prevalent in the rental universe, accounting for 42 percent of all rental units. Another 24 percent of rental units were two-family homes, 21 percent were small buildings (3–20 units), and 12 percent were large buildings (more than 20 units).

Second, many of the Cleveland rental properties showed signs of being distressed due to physical and market conditions. Such properties would likely require costly repairs but have little equity to support financing. As exhibit 2 shows, properties rated as being in bad condition made up 38 percent of the rental property universe, and approximately 8 percent of properties had an open housing code violation. Also, market conditions were unfavorable for many rental properties—roughly one-fourth had a very low assessed market value, defined as less than $25,000 for single and double homes and less than $10,000 per unit for three- or more unit buildings. Furthermore, 17 percent of rental properties were tax-delinquent by at least $500 in 2018, a possible indication of disinvestment.

Exhibit 2

Selected Characteristics of Pre-1978 Rental Properties in Cleveland, 2018 (n = 54,786)



Source: Property file built from multiple administrative data sources as described in exhibit 1

Third, during the study period, relatively few rental properties in Cleveland were connected to government housing programs. As exhibit 2 shows, only 30 percent were included in the rental registry, and 6 percent were taking HCVs. The owners and tenants of the disconnected properties, which make up most of Cleveland rental properties, may have little familiarity with local agencies and regulations, which may make it more challenging to engage them in complying with lead safety inspections and other requirements.

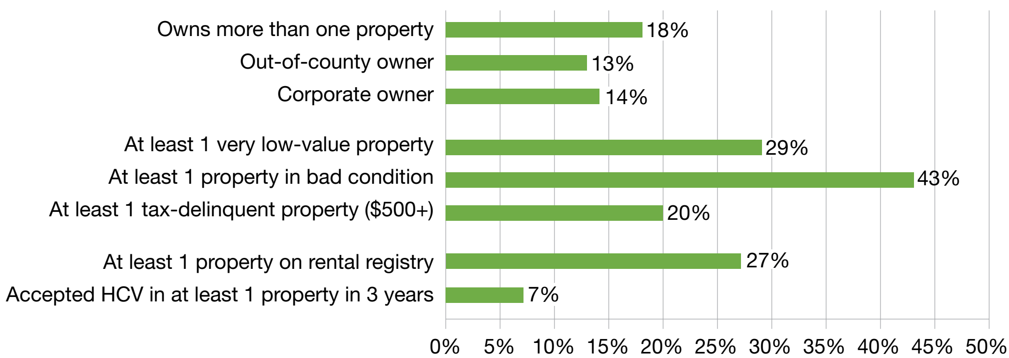
Finally, the location and stability of rental property ownership suggested the need for more than one engagement strategy. As exhibit 2 shows, rental properties turned over with some degree of frequency, with 30 percent having changed ownership within 3 years. However, most properties had the same owner for a longer period. Moreover, entities with addresses in Cleveland or Cuyahoga County suburbs owned 62 percent of rental properties, making personal contact related to lead-safe interventions possible. Most of the remaining owners had addresses outside Cuyahoga County but within Ohio, requiring other forms of contact. Corporate entities (that is, LLCs, limited partnerships, and other organizations or businesses) owned 23 percent of Cleveland rental properties, and 77 percent were titled in the name of individuals, suggesting the need for differentiated approaches to communicate with these individual owners not incorporated as businesses.

Selected Characteristics of Landlords

This section presents another way of looking at the Cleveland rental landscape by focusing on the property owners (exhibit 3). Most landlords owned only a single property in Cleveland, with only 18 percent owning two or more. It should be noted that, because the research team focused specifically on rental properties in Cleveland, landlord ownership of any rental properties outside the city is unknown and beyond the scope of this study. Most owners of Cleveland rental properties had a presence in the city or the surrounding Cuyahoga County area based on the location of their tax mailing address. Corporate entities accounted for 14 percent of these owners, and the other 86 percent of owners were classified as persons. A notable portion of landlords owned properties that had markers of distress or disinvestment, including having at least one property that was in bad condition (43 percent), had very low assessed market value (29 percent), or was tax-delinquent (20 percent). Only a minority of owners had properties in the rental registry (27 percent) or rented to households with HCVs (7 percent).

Exhibit 3

Selected Characteristics of Study Landlords in 2018 (n = 36,659)



HCV = housing choice voucher.
Source: Landlord file built from multiple administrative data sources as described in exhibit 1

To segment the landlord population in a way that could inform local strategy, the research team undertook an LCA of landlord characteristics using Stata's *gsem* command, latent class function, specifying three possible classes based on landlord characteristics (Stata Press, 2023). The number of landlord-held properties entered the model as an ordinal variable with an ordered logit specification. Other variables related to corporate status, share of properties in bad condition, of low value, and with violations and tax delinquency entered the model with a logit specification. Model fit indices—Akaike Information Criteria (AIC) and Bayesian Information Criteria (BIC) — favored the three-class model (AIC: 263,752; BIC: 263,999) over two-class (AIC: 267,588; BIC: 267,750) or one-class models (AIC: 283,142; BIC: 283,219). Estimates for a four-class model did not converge.

The results of the LCA suggest that landlords can be divided into three classes (exhibits 4 and 5). *Class 1* comprised 60 percent of landlords. However, these landlords accounted for only 44 percent of the properties and 45 percent of the rental units. Nearly all *Class 1* landlords (92 percent) owned only a single property in Cleveland, and slightly more than one-half of their properties were single-family homes. Their properties were generally rated as being in above-average or good condition, with few code violations or instances of tax delinquency. Their properties were seldom classified as being of very low value. Only 24 percent of these landlords were in the rental registry, and very few had tenants with HCVs.

Exhibit 4

Landlord Latent Classes: Prevalence and Counts of Properties and Units

	Class 1		Class 2		Class 3		Total	
	n	%	n	%	n	%	n	%
Landlords	21,788	60	11,083	30	3,788	10	36,659	100
Properties	24,328	44	12,197	22	18,261	33	54,786	100
Units	46,902	45	25,705	25	30,779	30	103,386	100

Note: Unique landlords used for latent class analysis, and they might have multiple properties and units.
Source: Landlord file built from multiple administrative data sources as described in exhibit 1

Exhibit 5

Descriptive Characteristics of Landlords by Latent Classes (%)

	Class 1 (n = 21,788)	Class 2 (n = 11,083)	Class 3 (n = 3,788)
At least one property:	%	%	%
in bad condition	7	96	93
with very low market value	8	58	66
with code violations in 3 years	5	13	34
with delinquent tax balance > $500	9	37	34
on rental registry	24	19	70
that accepted HCVs in 3 years	5	4	26
Number of properties owned	%	%	%
1	92	91	0
2	6	7	34
3 or 4	2	1	35
5 or more	0	0	30
All properties owned 3 years or less	27	23	22
Owns only single-family homes	54	26	25
Owner based outside Cuyahoga County	13	11	20
Corporate owner	12	11	36

HCVs = housing choice vouchers.
Source: Landlord file built from multiple administrative data sources as described in exhibit 1

Class 2 accounted for 30 percent of landlords, controlling 22 percent of the rental properties and 25 percent of the rental units. Nearly all (92 percent) had only one property in Cleveland. They tended to own fewer single-family homes than *Class 1* landlords and, instead, owned mostly two-family homes or buildings with three to four units. Nearly all the properties of *Class 2* landlords were in bad condition, more than one-half had very low market values, and more than one-third were tax-delinquent. Few corporate owners were in this class, and owner participation in the rental registry or HCV program was low.

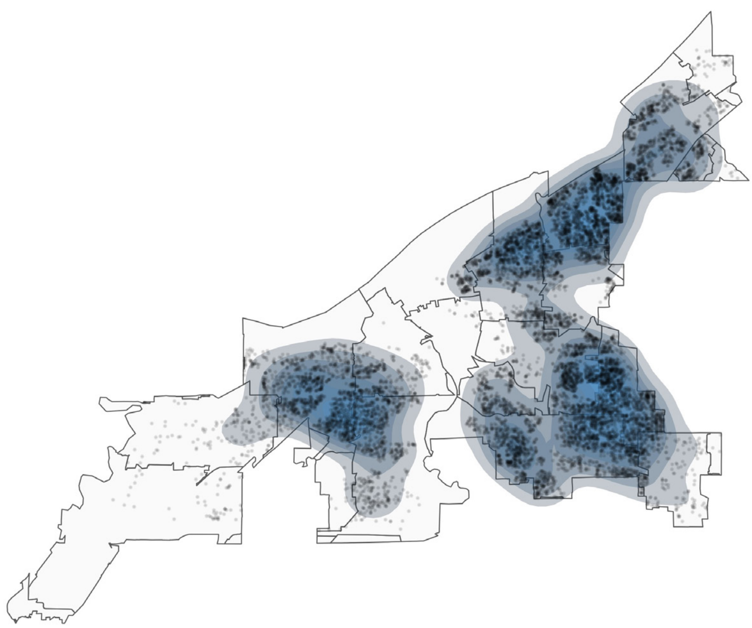
Class 3 landlords comprised 10 percent of rental property owners but represented 33 percent of the rental properties and 30 percent of the rental units in Cleveland. These landlords were more likely to be corporate owners compared with the other classes and tended to own numerous properties and included larger buildings in their portfolios. Most of these landlords (93 percent) had at least one property rated as being in bad condition, and 34 percent had at least one property with code violations or tax delinquency. The owners in this class were the most likely to include corporate entities and be based outside the county. Compared with the other classes, they had high levels of participation in the rental registry (70 percent) and HCV program (26 percent).

A Neighborhood Perspective

Thus far, the focus of this analysis has been on properties and landlords in the entire Cleveland rental market. However, as the lead-safe initiative is being rolled out geographically, it was strategically important to anticipate how the rental property conditions and mix of landlords differed by neighborhood. To illustrate this neighborhood variation, the exhibit 6 map shows the concentration of *Class 2* landlords (black dots) as revealed through LCA compared with the concentration of distressed rental properties (density curves) and other properties. These types of properties and landlords are likely to require more attention and assistance to achieve compliance with the lead-safe ordinance. Thus, the areas with the highest concentrations of points and at the center of the density curves are areas that will require additional resources and effort to bring the rental housing up to lead-safe standards. The concentration of high need overlaps considerably with areas that were historically redlined and hard hit by subprime lending and foreclosure during the previous decade (Perzynski et al., 2022). This pattern points to the ongoing effect of systemic racism in the housing market and its pernicious ongoing effects on the health of African-American children.

Exhibit 6

Spatial Concentration of Selected Property and Landlords Characteristics



Notes: Density curves are rental properties in bad condition or less than $25,000 assessed market value, or both. Points are Class 2 landlord-owned properties.
Source: Landlord file and property file built from multiple administrative data sources as described in exhibit 1

Discussion

The three classes of landlords identified in this study are likely to respond differently to the requirements of Cleveland's lead-safe ordinance or need customized resources to assist them in coming into compliance. Access to capital, likely return on investment in repairs, scale of operation, and managerial capacity are relevant variations among the landlord types that can inform the planning for lead-safe implementation. For example, *Class 2* landlords face particular challenges because most of their properties are in poor condition and need repair, but low market values limit access to conventional home improvement loans. Also, as small owners, they are unlikely to be tapped into networks of suppliers and contractors or be able to achieve economies of scale for their property renovations. They will need access to various types of subsidized capital, such as grants and low-interest loans. Moreover, because *Class 2* landlords typically own only one property in Cleveland and have low participation in the rental registry, outreach will be required to engage them in the process and provide access to information and resources.

Class 3 landlords have the scale and size to overcome some of the limitations facing *Class 2* landlords. They also have more connections to regulatory processes and government programs, as evidenced by their higher participation in the rental registry and HCV program. Although they often have at least one property in poor condition or at low market value, having multiple properties suggests greater capitalization and likely connections to construction services and financing.

Class 1 landlords, the most prevalent type in Cleveland, generally hold single-family properties in good condition and with solid market value. This fact suggests that many of these properties will meet lead-safe standards or that the ratio of repair costs to property values will be favorable for conventional financing. However, because their current level of participation in the rental registry is low and their scale of operation is small, *Class 1* landlords will likely benefit from receiving information about the lead-safe ordinance, the rental registration process, and access to qualified contractors or other resources if repairs are required.

An important implication for lead safety is that a significant portion of the landlords holding distressed properties are individuals with small-scale operations rather than real estate companies or professional entities. Most of their properties are not yet in the rental registry, nor are these owners participating in public programs, such as HCV. The low property values and deteriorated conditions of many of their properties suggest these landlords will find it difficult to get conventional financing in the amount required to upgrade their properties. These small operators may require outreach, information, technical assistance, and other support to bring their properties into the rental registry, complete lead inspections, and upgrade them to lead safety standards. Moreover, they may need to avail themselves of subsidized capital, including grants and low-interest loans, to bring their properties up to lead safety standards.

The research team notes that this article has several limitations. Cleveland properties and landlords probably differ along several other dimensions that could not be ascertained from the administrative records used for this typology. Also, because compliance with the rental registry was low, the team relied on other indicators of the property being a rental, such as the owner-occupancy tax credit, number of units, and so forth. Some properties may have been misclassified.

In addition, the reliance on public records and owner names to link landlords to their properties probably overestimates the unique number of individual entities involved. As demonstrated in several studies, deduplication is especially problematic when owners of record operate under more than one company name (An et al., 2022; Hangen and O'Brien, 2022). Moreover, the classification of landlords is based on a limited set of characteristics derived from administrative records. For example, it was impossible to investigate other important aspects of their businesses, such as the size of their portfolios or professionalization. As the lead-safe initiative unfolds, it will be important to gather landlords' perceptions to gain a deeper understanding of the classification and its implications for lead-safe strategies across all areas of the city.

Conclusion

This analysis demonstrates how multiple administrative data sources—along with record linkage, spatial analysis, and statistical methods—can inform community planning and action on important concerns, such as lead safety in housing. It also illustrates that such data sources can be organized for two different units of analysis—rental properties and property owners—to yield insight into both. Furthermore, it shows how such an analysis can inform policy direction and be used to evaluate strategies for specific problems, constituencies, or locations.

The study identified the total of pre-1978 rental properties in Cleveland's housing market, approximately 100,000 units in 55,000 properties (owned by 36,000 landlords). Importantly, this total includes both registered rentals (those on the city's rental registry) and rentals operating outside the registry. From a policy perspective, lead risk must be reduced in both groups of properties to have an eventual effect on child lead exposures, especially given that non-registered rentals account for approximately 70 percent of the pre-1978 rental housing stock. This finding informed the scope of the Cleveland Lead Safe Initiative and its deployment by city area and has guided an understanding of compliance rates by neighborhood.

In addition, the study assessed both the conditions and owner characteristics of these rental properties. These data were essential in understanding the relative anticipated cost of bringing different types of properties up to lead-safe standards. The analysis identified the likely capacities of property owners to undertake property improvements based on the number and quality of their property holdings. This information was instructive in fashioning outreach strategies to owners and developing financial and other supports to facilitate the achievement of lead safety in their properties. Identifying a population of owners with limited holdings of higher risk properties ensured the strategy offered deeper supports for these owners. As the initiative has implemented its loan and grant offerings, adjustments have been made to the application process and maximum award amounts based on learning from direct experience with properties.

Beyond the trends and patterns this analysis illuminated, the rental property and landlord data continue to inform ongoing program decisions and monitoring related to lead safety. The data sources are updated quarterly and used to populate a public-facing dashboard that allows community partners to monitor the initiative and surface challenges as they emerge. The monitoring data show that compliance rates are much higher among known rental properties compared with probable rentals, suggesting the challenges in outreach to owners who have not previously complied with the city's rental registration requirement. Greater compliance is also evident among properties with larger numbers of units. Low compliance among properties that are singles and doubles has been highlighted as a specific challenge because these properties account for most units in Cleveland's pre-1978 rental housing stock.

Single- and two-family homes owned by individuals based in the region—whether within the city or in the surrounding Cuyahoga County suburbs—dominate the rental housing stock in Cleveland. This preponderance of detached and individually owned rental housing units suggests that the lead-safe initiative will require the ability to customize inspections and repairs to this mix of properties. Although most of Cleveland's rental properties are maintained in above-average or good condition and have solid market value, many properties are unlikely to meet lead safety standards. Monitoring the repair of single- and two-family homes scattered on a case-by-case basis takes a different kind of capacity in city government than tackling code compliance in larger rental properties.

Because the lead-safe initiative is still in its early phase, the focus continues to be on monitoring compliance with rental registry and lead-safe certification requirements, especially in the neighborhoods with a concentration of properties and landlords that present the greatest risk due to poor housing conditions and limited capacity to invest in repairs. The continued use of administrative data—such as housing values, conditions, ownership patterns, and so forth—is key to tracking changes in the rental housing stock and illuminating unanticipated consequences or patterns that have strategic implications for Cleveland's lead-safe initiative. Such analysis will allow not only an examination of compliance with lead safety requirements but also an assessment of how the rental landscape may change over time.

Acknowledgments

The authors acknowledge former director Ayonna Blue Donald, Michael Banks, and Dr. Timothy Kobie from the City of Cleveland Department of Building and Housing for their contributions to this work. This research was made possible through funding from the Mt. Sinai Healthcare Foundation, Saint Luke's Foundation, George Gund Foundation, and Eva L. and Joseph M. Bruening Foundation.

Authors

Claudia Coulton is professor emeritus, Michael Henderson is a senior research associate, Francisca García-Cobián Richter is a research associate professor, Jeesoo Jeon is a Ph.D. candidate, Michael Schramm is a research associate, and Robert L. Fischer is a professor in the Center on Poverty and Community Development at Jack, Joseph and Morton Mandel School of Applied Social Sciences at Case Western Reserve University. April Urban is director of research and impact at Signal Cleveland.

References

An, Brian, Andrew Jakabovics, Anothony W. Orlando, and Seva Rodnyansky. 2022. *Who Owns Urban America? A Methodology for Identifying Real Estate Owners*. Atlanta, GA: Georgia Tech. https://repository.gatech.edu/entities/publication/472788f9-a5e6-4d9b-8238-422d20333bcb.

City of Cleveland. 2019. "CHAPTER 365 – RENTAL REGISTRATION AND LEAD-SAFE CERTIFICATION." https://codelibrary.amlegal.com/codes/cleveland/latest/cleveland_oh/0-0-0-16247#JD_Chapter365.

Coulton, Claudia, Francisca García-Cobián Richter, Youngmin Cho, Jiho Park, Jeesoo Jeon, and Robert L. Fischer. 2023. "Making the Case for Lead Safe Housing: Downstream Effects of Lead Exposure on Outcomes for Children and Youth," *Health & Place* 84. https://doi.org/10.1016/j.healthplace.2023.103118.

Fischer, Rob, Stephen Steh, and Tsui Chan. 2018. *Early Childhood Lead Exposure Among Cleveland Kindergarteners by Neighborhood and School Enrollment*. Cleveland, OH: Case Western Reserve University, Mandel School of Applied Social Sciences. https://case.edu/socialwork/povertycenter/sites/case.edu.povertycenter/files/2019-01/Lead%20Report%20CMSDFinal_0119.pdf.

Garboden, Philip M.E., and Sandra Newman. 2012. "Is Preserving Small, Low-End Rental Housing Feasible?" *Housing Policy Debate* 22 (4): 507–526. https://doi.org/10.1080/10511482.2012.697909.

Gomory, Henry. 2022. "The Social and Institutional Contexts Underlying Landlords' Eviction Practices," *Social Forces* 100 (4): 1774–1805.

Greif, Meredith. 2018. "Regulating Landlords: Unintended Consequences for Poor Tenants," *City & Community* 17 (3): 658–674. https://doi.org/10.1111/cico.12321.

Hangen, Forrest, and Daniel T. O'Brien. 2022. "Linking Landlords to Uncover Ownership Obscurity," *SocArXi*. https://osf.io/anvke/download.

Kingsley, G. Thomas. 2017. *Trends in Housing Problems and Federal Housing Assistance*. Washington, DC: Urban Institute. https://www.urban.org/sites/default/files/publication/94146/trends-in-housing-problems-and-federal-housing-assistance.pdf.

Lanphear, Bruce P., Richard Hornung, and Mona Ho. 2005. "Screening Housing to Prevent Lead Toxicity in Children," *Public Health Reports* 120 (3): 305–310. https://doi.org/10.1177/003335490512000315.

Lead Safe Cleveland Coalition. 2023. "2023 Action Plan." https://leadsafecle.org/sites/default/files/2023-06/2023%20Action%20Plan%20Final%20Draft%20.pdf.

Messamore, Andrew. 2023. "The Institutionalization of Landlording: Assessing Transformations in Property Ownership Since the Great Recession." http://dx.doi.org/10.2139/ssrn.4480068.

Perzynski, Adam, Kristen A. Berg, Charles Thomas, Anupama Cemballi, Tristan Smith, Sarah Shick, Douglas Gunzler, and Ashwini R. Sehgal. 2022. "Racial Discrimination and Economic Factors in Redlining of Ohio Neighborhoods," *Du Bois Review*: 1–17. https://doi.org/10.1017/S1742058X22000236.

Rothstein, Richard. 2017. *The Color of Law: A Forgotten History of How Our Government Segregated America*. New York: Liveright Publishing.

Sampson, Robert J., and Alix S. Winter. 2016. "The Racial Ecology of Lead Poisoning: Toxic Inequality in Chicago Neighborhoods, 1995–2013," *Du Bois Review* 13 (2): 261–283. https://doi.org/10.1017/S1742058X16000151.

Shaw, Mary. 2004. "Housing and Public Health," *Annual Review of Public Health* 25: 397–418. https://doi.org/10.1146/annurev.publhealth.25.101802.123036.

Shiffer–Sebba, Doron. 2020. "Understanding the Divergent Logics of Landlords: Circumstantial Versus Deliberate Pathways," *City & Community* 19 (4): 1011–1037. https://doi.org/10.1111/cico.12490.

Stata Press. 2023. *Structural Equation Modeling Reference Manual*. College Station, TX. ISBN 978-1-59718-397-0.

U.S. Census Bureau. 2021a. "Poverty Status in the Past 12 Months," American Community Survey 1-Year Estimates. https://data.census.gov/table/ACSST1Y2021.S1701?q=poverty&g=010XX00US$0400000.

———. 2021b. "Years Structure Built," American Community Survey 1-Year Estimates. https://data.census.gov/table/ACSDT1Y2021.B25034?q=year+structures+built&g=010XX00US$0400000.

van der Loo, Mark. 2014. "The *Stringdist* Package for Approximate String Matching," *The R Journal* (6): 111–122. https://CRAN.R-project.org/package=stringdist.

Merging Federal Flooding and Housing Data to Gain Insight into Flood Impacts on Federally Assisted Households: A Case Study in Kansas City, Missouri

Mariya Shcheglovitova
U.S. Department of Housing and Urban Development

Gina Lee
U.S. Geological Survey, Maryland-Delaware-District of Colombia Water Science Center

The views expressed in this article are those of the authors and do not represent the official positions or policies of HUD's Office of Policy Development and Research and the U.S. Department of Housing and Urban Development.

Abstract

Previous research on the differential impacts of floods has found that race, ethnicity, income, gender, age, and housing tenure and type influence people's ability to prepare for and respond to flood events. However, studying the impacts of flooding is often challenging due to data limitations, especially for storm- and snowmelt-related flooding in noncoastal areas. This article draws on Federal Emergency Management Agency (FEMA) flood risk maps, the U.S. Geological Survey (USGS) Flood Inundation Mapping (FIM) Program, and U.S. Department of Housing and Urban Development (HUD) administrative data to provide a methodological example of integrating federal flooding and housing data to gain local housing insights. The USGS FIM Program provides a unique opportunity to help communities visualize potential areas at risk for flooding near local streams and rivers. Unlike FEMA flood risk maps, USGS FIM maps allow researchers to investigate local flooding processes, such as the predicted extent and depth of housing-unit flood exposure. To demonstrate the utility of USGS data for housing research, this article presents a case study investigating the impact of flooding on housing units where households receiving federal rental assistance live in Kansas City, Missouri. The presented analysis contrasts local housing unit trends in flood exposure to census-tract-level trends of flood risk derived from FEMA riverine flood maps. This case study demonstrates how USGS and FEMA data can inform housing analyses at different scales for researchers and practitioners interested in flood impacts on local communities and vulnerable populations.

Introduction

Climate change has increased the frequency and severity of extreme weather events, including floods, hurricanes, and severe storms (IPCC, 2022). Extreme weather events pose a threat to the built environment and housing (Hallegatte and Przyluski, 2010) and further exacerbate social inequalities by disproportionately affecting low-socioeconomic-status communities (Howell and Elliott, 2019) and communities of color living in areas of infrastructural disinvestment (Hendricks and Van Zandt, 2021).

Municipalities are taking increasing steps to address disproportionate risks from natural disasters by building back more resilient infrastructure or investing in infrastructural resilience before a disaster (HUD, n.d.a.). However, planning in ways that preempt natural disasters depends on the availability of data used to identify areas and structures most at risk of environmental hazards. This article explores using U.S. Geological Survey (USGS) Flood Inundation Mapping (FIM) Program, Federal Emergency Management Agency (FEMA) flood risk, and U.S. Department of Housing and Urban Development (HUD) administrative data to provide a methodological example of integrating federal flooding and housing data to gain local housing insights.

Flooding is the most prevalent natural disaster in the United States and is expected to become more frequent and severe due to climate and land use change (Brody et al., 2007). Social vulnerability research on flood risk has drawn attention to the impacts of race, ethnicity, income, gender, and age (Cutter, Boruff, and Shirley, 2003; Rufat et al., 2015; Walker, 2012) and housing tenure and type (Lee and Van Zandt, 2018; Mehta, Brennan, and Steil, 2020) on flood exposure and capacities to prepare for and recover from flood events. Research investigating the impacts of Hurricane Harvey found that the areal extent of flooding was significantly greater in neighborhoods with higher percentages of non-Hispanic Black and low-socioeconomic-status residents (Chakraborty, Collins, and Grineski, 2019) and people with disabilities (Chakraborty, Grineski, and Collins, 2019). In addition, the areal extent of Hurricane Harvey flooding within 100 meters of residents' homes was significantly greater for racial/ethnic-minority and low-socioeconomic-status households (Collins et al., 2019). The social impacts of flood exposure cannot be disentangled from structural constraints on housing markets. Many low-income and racial/ethnic-minority households live in flood-prone areas because of the lower housing costs (Levine, Esnard, and Sapat, 2007). Those stark geographies are pronounced particularly in the U.S. South, where Carrera and Coleman Flowers (2018) documented how White landowners and cotton production aggregated in higher elevations with better drainage, whereas Black residents were limited to living in lowland areas that were more prone to flooding.

A growing number of hazard vulnerability studies aim to link social vulnerability and housing by focusing specifically on the location of federally subsidized households and their exposure to major flooding events (Chakraborty et al., 2021; Davlasheridze and Miao, 2021; Hamideh and Rongerude, 2018; Hernández et al. 2018). Chakraborty et al. (2021) investigated the impacts of Hurricane Harvey on tenants receiving federal rental assistance and found that they were more likely to live in areas with greater flood extent. Davlasheridze and Miao (2021) found that not only do floods reduce the number of available housing units for federally subsidized tenants, but they can also increase the average time on waitlists to determine eligibility for housing programs and the share of rent paid by tenants in those programs. Studying the impacts of Hurricane Sandy on federally subsidized households, Hernández et al. (2018) found that many residents in New York City Housing Authority units were unlikely to evacuate or relocate despite experiencing power outages, flooded streets, and damaged building infrastructure because they feared displacement if city inspectors condemned their apartment. The impact of flooding on federally assisted households is not isolated to extreme storm events; a 2017 report by the New York University (NYU) Furman Center found that approximately 9 percent of government-subsidized households live in a 1-percent annual exceedance probability area (commonly referred to as a 100-year floodplain), an area with a 1-percent probability of flooding each year (Rosoff and Yager, 2017). Together, the research points to a need to identify areas where federally assisted households live in housing at risk of flood exposure and to develop strategies to support economically disadvantaged communities likely to experience more frequent and severe storm events.

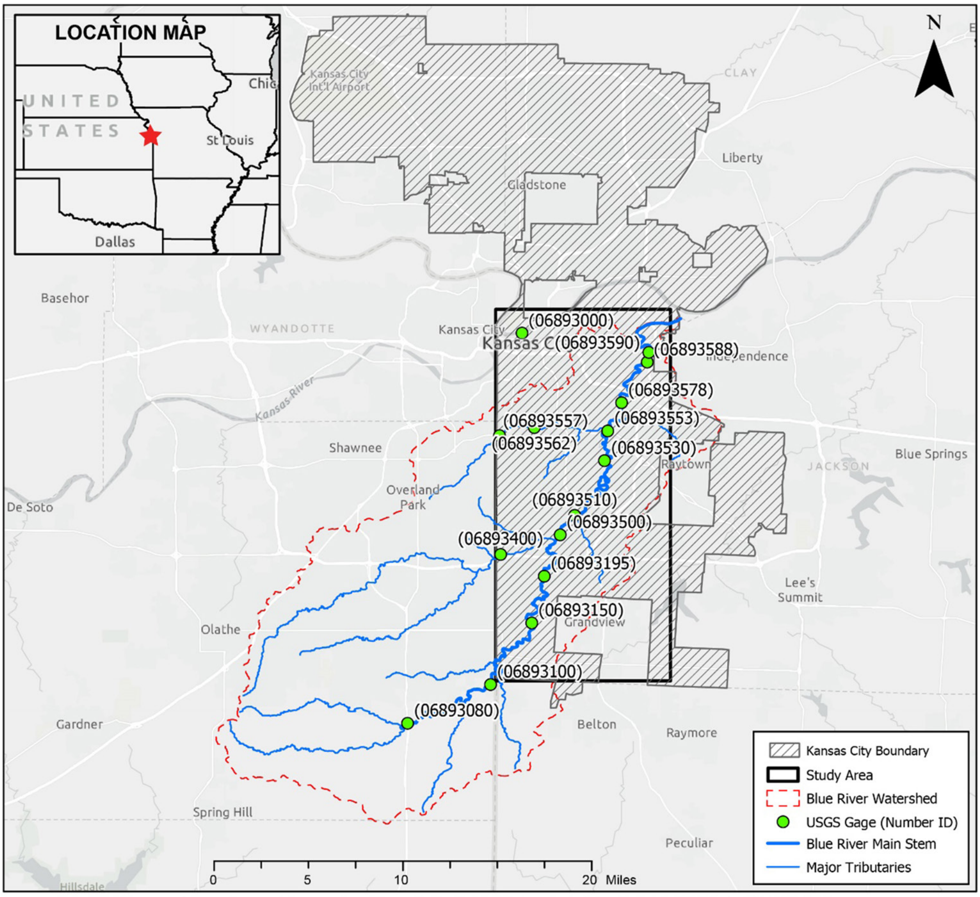
This article contributes to the research on flood exposure for federally subsidized households and housing units. Rather than investigating flooding following a natural disaster, the authors examine the possibility of using USGS Flood Inundation Mapper data (USGS, 2019) to identify flood exposure under set stream conditions for housing units where federally assisted households live. The authors also explore the use of FEMA flood risk maps (FEMA, n.d.) to ascertain relative flood risk for census tracts where federally assisted households are located. The study focuses specifically on riverine flooding, which can be more unpredictable than coastal flooding and can occur more frequently, with annual flash flooding due to climate change-driven increases in storm magnitude and frequency (Vanucchi, 2021) and increased runoff from impervious surfaces due to urbanization (Shuster et al., 2005). A focus on riverine flooding can also support U.S.-based studies that aim to investigate links between racial and economic segregation and vulnerability to climate hazards. Previous studies found that areas of concentrated social disadvantage in the United States were associated with greater inland flooding, whereas areas with more socially advantaged populations were associated with greater coastal flooding (Chakraborty et al., 2014; Qiang, 2019; Ueland and Warf, 2006). The sections that follow present a case study for Kansas City, Missouri (KCMO), integrating USGS riverine flood models and HUD administrative data, and compare USGS flood models with FEMA flood risk maps to provide a blueprint for data sources and analysis methods that can inform future research on flood risk across different scales of analysis.

Study Area: Blue River Kansas City, Missouri

The USGS FIM program works with local communities to produce flood models for stream sections identified by USGS and local stakeholders. Although the FIM map library contains flood inundation maps for 27 U.S. states, the maps are limited to local streams and rivers. In KCMO, the FIM maps encompass the Lower Blue River and its tributaries (exhibit 1). Thus, the study area in this analysis does not encompass all KCMO but is limited to a bounding box containing the USGS FIM mapped area.

Exhibit 1

Blue River Watershed with Locations of USGS Gages Used to Develop Flood Inundation Models



Basemap: Missouri Dept. of Conservation, Missouri DNR, Esri, HERE, Garmin, SafeGraph, FAO, METI/NASA, USGS, EPA, NPS

USGS = U.S. Geological Survey.

Notes: The study area containing the Blue River within KCMO is outlined. Eight-digit numbers along the Blue River main stem and major tributaries represent the numeric IDs of USGS stream gage sites.

Sources: 2021 118th Congressional District TIGER/Line Place State-based Shapefile; National Hydrography Dataset (NHD); USGS StreamStats website (https://streamstats.usgs.gov)

Demographics

Kansas City, Missouri, is a city in the U.S. Midwest that covers approximately 313 square miles (810.7 square kilometers) and is home to more than half a million (508,090) residents (U.S. Census Bureau, 2023). The population of KCMO is racially and ethnically diverse; 11 percent of the population identifies as Hispanic or Latino, 26 percent of the population identifies as Black or African-American, and 56 percent of the population identifies as White.[1] The median household

[1] Data accessed from American Community Survey table B03002: HISPANIC OR LATINO ORIGIN.

income in KCMO is $63,396, and approximately 13.4 percent of the population lives in poverty (U.S. Census Bureau, 2023).

Like many other U.S. metropolitan areas, KCMO has been taking steps to address gaps in affordable housing. In May 2021, KCMO created the Housing and Community Development Department. One aim of the department is to assist renters with finding and maintaining safe and affordable housing (City of Kansas City, Missouri, 2021). More than 27,000 people in KCMO live in HUD-assisted households, the majority of whom are extremely low-income (82 percent; HUD, n.d.b). More than 70 percent of households receiving federal rental assistance in KCMO are female-headed (72 percent), and most household heads are non-Hispanic Black (76 percent; HUD, n.d.b). Approximately 90 percent of all households receiving federal rental assistance in KCMO live in the study area containing the Lower Blue River.

Hydrology

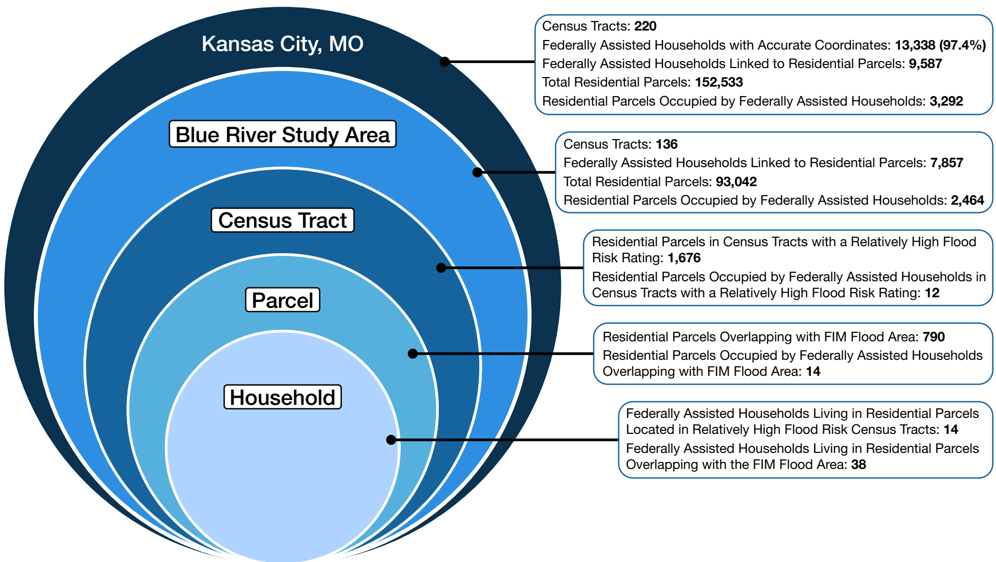
Flooding is generally the most common and costliest type of disaster Missouri experiences (Missouri Department of Public Safety, n.d.). Much of the historic flooding in KCMO has occurred along the Blue River and several tributaries, most notably in 1951, 1961, 1977, 1984, 1990, 1998, 2010, and 2017 (Heimann et al., 2014; USGS, 2023a). The Blue River is an approximately 39.8-mile (64.1-kilometer)-long tributary of the Missouri River, with a drainage area of approximately 270.5 square miles (700.5 square kilometers; USGS, 2023b). Along the lower reaches of the mainstem floodplain and along its major tributaries (Brush Creek, Indian Creek, and Tomahawk Creek), the river is moderately to highly developed, with a mix of residential and commercial properties. The Blue River flows northward through most of the southern half of the Kansas City metropolitan area within Johnson and Wyandotte Counties in Kansas and Jackson and Cass Counties in Missouri. The headwaters of the Blue River outside the city limits consist of grass- or forestland (exhibit 1; Wilkison et al., 2006); however, the watershed is still under development, with continued construction of residential properties in the headwaters (Heimann et al., 2014). The river is prone to annual flooding due to urbanization of the floodplain and surrounding watershed, leading to increased runoff and changes in the natural basin hydrology (Driever and Vaughn, 1988; Wilkison, Armstrong, and Blevins, 2002) that will be exacerbated further by projected increases in precipitation due to climate change (Byun, Chiu, and Hamlet, 2019). Levees have been built near the confluence of the Blue River and Missouri River and near the confluence of Indian Creek and Blue River to mitigate flooding (Heimann et al., 2014), and the channel has been straightened and armored to minimize inundation and maximize stormwater conveyance (Wilkison et al., 2006).

Data Sources

Many spatial studies of flooding investigate impacts at the census-tract or block-group level because data are commonly available at this spatial scale. Household- and housing-unit-level studies are less common because they often require that researchers collect household survey data rather than relying on publicly available data sources (Collins et al., 2019) or have flood models accurate enough to make inferences about flood depth and extent at the housing unit level. HUD administrative data are a unique dataset that allows researchers to study housing-unit-level environmental impacts across a subset of the population. When HUD administrative data are paired with USGS FIM models and local parcel data, researchers can investigate the depth and extent of flooding that will affect the housing units where federally assisted households live (exhibit 2).

Exhibit 2

Datasets Used in the Analysis Within the Lower Blue River Study Area



Notes: The text boxes summarize information about the variables from each contributing dataset. The nested diagram represents the relationship between spatial scales (e.g., a household is contained within a parcel) and is not meant to be a representation of physical scale.

Sources: Authors' analysis of census tract flood risk ratings from Federal Emergency Management Agency Riverine Flood Risk maps; U.S. Geological Survey Flood Inundation Mapping Program flood models for the Lower Blue River; Kansas City, MO, city parcel data; HUD administrative data

Housing: HUD Administrative Data and Kansas City Parcel Data

Household-level data for participants in federal housing rental assistance programs were extracted from HUD's internal database containing information collected via HUD form 50058.[2] This dataset contains entries describing demographic, economic, and program variables for all tenants receiving federal rental assistance in tenant- and project-based rental assistance programs. The Housing Choice Voucher (HCV) program is the primary tenant-based rental assistance program administered by HUD. In this program, tenants can use their housing subsidy to rent a housing unit on the private market that meets housing condition, health, and safety requirements. In project-based rental assistance programs, private property owners enter a contract with HUD to provide affordable rental units to tenants participating in HUD programs. Unlike tenant-based programs, tenants participating in project-based rental assistance programs cannot take subsidies with them when they move.

[2] HUD makes this household-level data available to external researchers via data licenses (https://www.huduser.gov/portal/research/pdr_data-license.html). Extracts of these data are publicly available at the census-tract level (https://www.huduser.gov/portal/datasets/assthsg.html).

Data current as of December 2022 for all households in Kansas City, Missouri, participating in HUD rental assistance programs were extracted from HUD's internal database. To observe whether tenant- or project-based program participants live in areas of more or less flood exposure, the authors stratified households by program type. The downstream analysis relied on having accurate locations of households participating in HUD programs. Although all records in HUD's database contain addresses and are geocoded, the level of geocode quality varies; therefore, the authors removed records whose addresses could not be verified with latitude and longitude coordinates accurate to the dwelling rooftop level.

Households receiving federal rental assistance may reside in single and multifamily homes. To identify flood impacts on housing units where households receiving federal rental assistance live, household-level HUD administrative data were linked to KCMO parcel data (Bender, 2023; exhibit 2). The presented analysis considered only residential parcels identified by KCMO land use codes for single-family, mobile home, townhouse, duplex, condominium, and multifamily development (Bender, 2021). Identifying both the location and footprint of housing units affected by flooding allowed the authors to estimate the relative flood risk using FEMA flood maps and the predicted depth and extent of flooding for affected parcels using USGS FIM maps.

Flooding: FEMA Riverine Flood Risk and USGS Flood Inundation Mapper

FEMA's National Risk Index (NRI; Zuzak et al., 2023) was used to investigate census-tract-level riverine flood risk trends. Housing parcels occupied by federally assisted households located in census tracts likely to be affected by riverine flooding were identified using FEMA's NRI Riverine Flooding map (FEMA, n.d.). FEMA calculates risk ratings at the census-tract level on the basis of floodplain boundaries and historic storm and flood events, representing a relative risk where communities are grouped in percentiles based on national ratings (Zuzak et al., 2023). Rather than relying on a single source of flooding data, the presented analysis is informed by combining FEMA and USGS data sources. This approach can help guide decisionmaking under increasingly uncertain climate scenarios, for which flood maps can become quickly outdated (Smiley, 2020).

The purpose of the FIM Program is to provide information to communities regarding local flood risks and planning tools for cost-effective mitigation. Unlike the FEMA flood maps, USGS FIM maps do not indicate the risk of inundation but only a detailed model of the extent and depth of inundation for a given flood stage. The FIM Program has two main functions. The first is to partner with communities to create and validate a library of maps displaying potential areas of flooding over a range of water levels for local streams and rivers. USGS has standardized the procedures for creating flood inundation maps for flood-prone communities using scientifically sound methods, including hydraulic and topographic modeling based on real data (USGS, 2023a). The second goal of the FIM Program is to provide these inundation maps online along with additional data, including real-time streamflow data, flood forecasts, and potential loss estimates.

USGS began creating the flood inundation library for the Blue River and selected tributaries in cooperation with the city of Kansas City in 2012. The library consists of 345 estimated flood inundation maps along a 39.7-mile stretch of the Blue River, subdivided into 15 reaches based on USGS stream gage locations, to its confluence with the Missouri River in KCMO (exhibit 1).

The inundation maps depict the areal extent of modeled flooding at various flood stages and the depth of water at each flood stage. The library of flood inundation maps was developed using a variety of data sources, including streamgage data and existing hydraulic models from the U.S. Army Corps of Engineers (USACE) and the city of Kansas City. Additional model parameters, including topographic and bathymetric data, were collected along several cross sections along the study reach; geometry data of bridges and structures crossing the channel were collected to model backwater effects, and appropriate roughness coefficients were refined by model calibration. These data were then used to compute water-surface profiles using the USACE Hydrologic Engineering Center's River Analysis System (HEC-RAS) Depth-Averaged Flow and Sediment Transport Model (FST2DH) software programs. These flood stages were created at 1-foot intervals referenced to the streamgage datum and ranging from the National Weather Service Action stage (the approximate top of bank flow—i.e., the amount of flow that a channel can carry without overflowing its bank, thus 0 feet of inundation) to that which exceeds the stage of the estimated 0.2-percent annual exceedance probability. The simulated water-surface profiles were then combined with a digital elevation model of the study area to delineate estimated flood inundation areas as shapefile polygons and depth grids for each water-surface profile in a geographic information system (GIS) software program.[3] The study area for this analysis is defined as the region in KCMO containing the Blue River and tributary flood inundation maps produced by USGS.

Methods

All analyses were performed using R Statistical Software (v4.2.2; The R Foundation, 2022). To facilitate the replication of this analysis, the R code is available in the appendix of this article.

Data Linkage

HUD administrative household data was linked to housing-unit parcels and census tracts. First HUD administrative data was subset to only households whose address coordinates were accurate to the dwelling rooftop level. Then, the spatial intersection function in R's *sf* package (Pebesma, 2018) was used to identify residential parcels occupied by federally assisted households on the basis of georeferenced coordinates for addresses. Parcels were considered to be residential if they were identified in the KCMO parcel file as single-family homes, mobile homes, townhouses, duplexes, multifamily homes, or condominiums. Census tracts containing residential parcels occupied by federally assisted households were similarly identified to quantify the number of housing units in census tracts with low, moderate, and high FEMA riverine flood risk ratings.

Flood Measures

Using both FEMA and USGS data facilitated the measurement of several indicators of flood risk and exposure. FEMA risk ratings are calculated at the census-tract level on the basis of floodplain boundaries and historic storm and flood events. These ratings represent a relative flood risk based on national ratings (Zuzak et al., 2023). FEMA flood risk ratings are grouped into census tracts

[3] For more information regarding the methodology of the inundation map models, please refer to USGS Scientific Investigations Report 2014-5180 (Heimann et al., 2014) and the USGS Flood Inundation Mapping Science website (https://www.usgs.gov/mission-areas/water-resources/science/flood-inundation-mapping-science).

of Very High (80th to 100th percentile), Relatively High (60th to 80th percentile), Relatively Moderate (40th to 60th percentile), Relatively Low (20th to 40th percentile), and Very Low (0 to 20th percentile) risk (Zuzak et al., 2023). Census tracts with no infrastructural, population, or agricultural annual loss associated with riverine flooding expected are classified as No Rating (Zuzak et al., 2023).

Unlike FEMA maps, the USGS FIM maps do not indicate the risk of inundation but only a detailed model of the extent and depth of inundation for a given flood stage. To calculate the expected flood depth and extent for affected parcels, a GIS-based methodology comprising of several steps was used. First, the area covered by all pixels with a predicted flood depth greater than zero from the FIM model within each parcel was summed and the maximum flood depth in the parcel area was identified. FEMA's guide to retrofitting homes to mitigate flooding identifies flood depth as affecting structures when floodwaters exert increased pressure as flood depth increases. Two to 6 feet of flooding can push on exterior walls and up on floors; if a structure is not designed to resist that pressure, it can cause structural damage, possibly leading to the structure collapsing (FEMA, 2014). Thus, flood inundation depths were categorized into intervals of 0–2 feet, 2–6 feet, 6–15 feet, and 15 feet or greater. Finally, the expected flood extent was calculated by dividing the flooded area by the area of the parcel (in square feet) to derive the proportion of the parcel expected to experience flooding. The R package *stars* (Pebesma and Bivand, 2023) was used to analyze flood raster grids.

Results and Discussion

This analysis draws on two federal data sources predicting local impacts of flooding and HUD administrative data on households receiving federal rental assistance linked to city residential parcel data to investigate the impacts of flooding on structures in which federally assisted households live. The following sections present results summarizing anticipated flood impacts when using FEMA and USGS data and briefly discuss exposure to flood hazards across different federal rental assistance programs. The discussion concludes by contrasting FEMA and USGS flood data and describing the potential applications and limitations of those data sources.

FEMA Flood Risk Maps

More than one-half of residential parcels occupied by federally assisted households in the study area (51.1 percent) are in census tracts with no FEMA riverine flood risk rating, indicating no expected annual loss due to flooding. A relatively small percentage are in areas on the very low (5.3 percent) and relatively high extremes of riverine flood risk (0.5 percent). The majority are in census tracts with a relatively low (29.3 percent) or relatively moderate (13.9 percent) flood risk rating. Trends for housing units where federally assisted households live are relatively consistent with all residential housing units in the study area (exhibit 3).

Exhibit 3

Share of All Residential Parcels and Residential Parcels Occupied by Federally Assisted Households in the Lower Blue River Study Area Located in Census Tracts Classified by FEMA Riverine Flood Risk Ratings

FEMA Riverine Flood Risk Rating	Residential Parcels Occupied by Households Receiving Federal Rental Assistance (%)	All Residential Parcels (%)
No Rating	51.1	52.1
Very Low	5.3	6.9
Relatively Low	29.3	27.6
Relatively Moderate	13.9	11.5
Relatively High	0.5	1.8

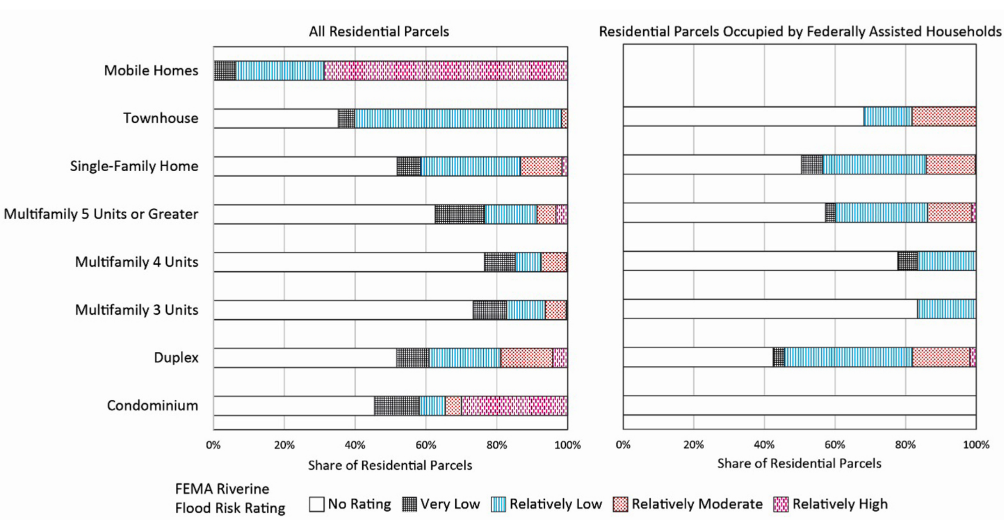
FEMA = Federal Emergency Management Agency.

Sources: Authors' analysis of FEMA Riverine Flood Risk maps; Kansas City, MO, city parcel data; HUD administrative data

A total of 1,676 residential parcels in the study area are located in census tracts with a relatively high riverine flood risk rating, including 12 occupied by federally assisted households. The 12 affected residential parcels are occupied by 14 households receiving federal rental assistance. For all residential parcels, housing in census tracts with a relatively high riverine flood risk rating includes single-family, multifamily, and mobile homes (exhibit 4). Notably, more than one-half of all mobile homes (68.8 percent) in the study area are in census tracts with a relatively high riverine flood risk rating.

Exhibit 4

All Housing Units and Housing Units Occupied by Households Receiving Federal Rental Assistance, by Census Tract Flood Risk Rating and Parcel Structure Type



FEMA = Federal Emergency Management Agency.

Sources: Authors' analysis of FEMA Riverine Flood Risk maps; Kansas City, MO, city parcel data; HUD administrative data

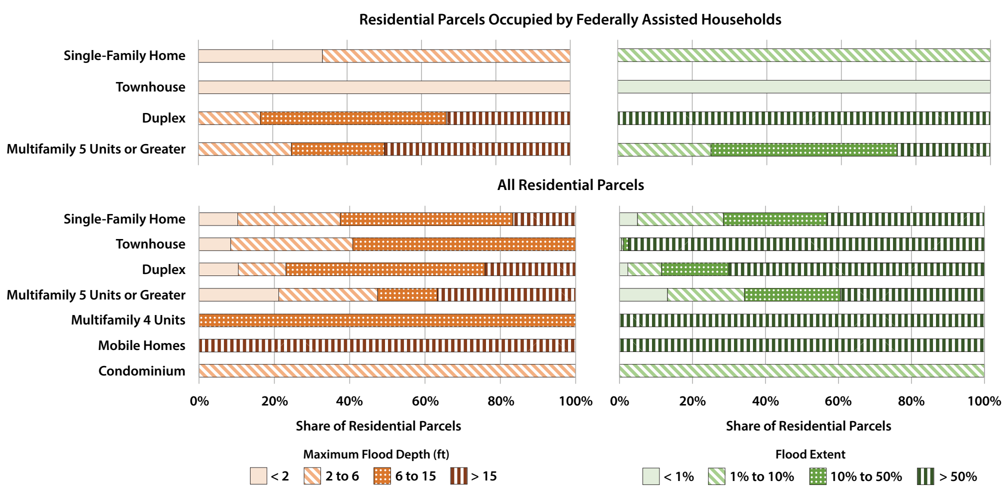
USGS FIM Maps

The study area contains 61 percent of the residential parcels in KCMO, and 8.4 percent of residential parcels in the study area are occupied by federally assisted households. Approximately 0.9 percent of the residential parcels in the study area have some overlap with USGS flood inundation maps for the Blue River. This share is lower for parcels occupied by federally assisted households (0.2 percent).

Federally assisted households in 14 residential parcels are predicted to be affected by flooding from the Lower Blue River. Those parcels are occupied by 38 households receiving federal rental assistance. As with the presented analysis of FEMA flood risk data, using FIM flood maps reveals flood impacts across single-family, multifamily, and mobile homes. In addition, FIM data facilitated an estimation of the expected flood depth and extent for affected parcels (exhibit 5). For all residential parcels, including those occupied by households receiving federal rental assistance, the predicted maximum flood depth is lower for single-family homes and townhomes than duplexes, multifamily homes, and mobile homes (exhibit 5). Single-family homes and townhomes occupied by federally assisted households within the predicted flood area of the Lower Blue River may be less affected by flooding than those not occupied by federally assisted households because the former are in regions with lower flood depths and have a smaller degree of overlap with expected flood areas (exhibit 5).

Exhibit 5

Expected Maximum Flood Depth and Flood Extent Affecting All Housing Units and Housing Units Occupied by Households Receiving Federal Rental Assistance that Overlap with U.S. Geological Survey Flood Inundation Mapping Maps, by Parcel Structure Type



Note: Lighter colors represent lower predicted flood depth and extent.

Sources: Authors' analysis of U.S. Geological Survey Flood Inundation Mapping flood models for the Lower Blue River; Kansas City, MO, city parcel data; HUD administrative data

Federal Rental Assistance Programs and Flood Exposure

Approximately 90 percent of the housing units in KCMO occupied by households receiving federal rental assistance are occupied by tenants participating in the tenant-based HCV program. Tenants participating in public housing (5.9 percent), project-based Section 8 (2.8 percent), project-based HCV (0.96 percent), and other multifamily programs (0.52 percent) occupy the remaining housing units linked to households receiving federal rental assistance.

Investigating flood risk relative to housing assistance program type revealed that only housing units occupied by tenants participating in the tenant-based HCV program were in census tracts with a relatively high flood risk rating (exhibit 6). However, housing units in census tracts with a relatively high flood risk rating represent a small share (0.5 percent) of overall housing units occupied by households participating in the tenant-based HCV program. Housing units occupied by households participating in the project-based HCV program had the largest share in census tracts with relatively moderate flood risk. Similar to the analysis of FEMA riverine flood risk ratings, only participants in the tenant-based HCV program were observed living in areas that are predicted to be affected by flooding from the Blue River (exhibit 7).

Exhibit 6

Share of Housing Units Occupied by Federally Assisted Households in the Lower Blue River Study Area, by Census Tract FEMA Riverine Flood Risk Rating and Federal Rental Assistance Program Type

FEMA Riverine Flood Risk Rating	Residential Parcels Occupied by Federally Assisted Households (%)	All Housing Choice Vouchers (%)	Tenant-Based Vouchers (%)	Project-Based Vouchers (%)	Project-Based Section 8 (%)	Public Housing (%)	Other Multifamily (%)
No Rating	51.1	51.8	51.9	37.5	61.4	34.7	53.8
Very Low	5.3	5.4	5.4	0.0	0.0	6.1	7.7
Relatively Low	29.3	28.5	28.4	37.5	15.7	49.0	23.1
Relatively Moderate	13.9	13.8	13.7	25.0	22.9	10.2	15.4
Relatively High	0.5	0.5	0.5	0.0	0.0	0.0	0.0

FEMA = Federal Emergency Management Agency.

Sources: Authors' analysis of FEMA Riverine Flood Risk maps; Kansas City, MO, city parcel data; HUD administrative data

Exhibit 7

Share of Housing Units Occupied by Federally Assisted Households in the Lower Blue River Study Area that Fall Within and Outside the Flood Area Predicted in U.S. Geological Survey Flood Inundation Mapping Maps, by Rental Assistance Program Type

USGS Flood Inundation Map Area	Residential Parcels Occupied by Federally Assisted Households (%)	All Housing Choice Vouchers (%)	Tenant-Based Vouchers (%)	Project-Based Vouchers (%)	Project-Based Section 8 (%)	Public Housing (%)	Other Multifamily (%)
In Flood Area	0.6	0.6	0.6	0.0	0.0	0.0	0.0
Outside Flood Area	99.4	99.4	99.4	100.0	100.0	100.0	100.0

USGS = U.S. Geological Survey.

Sources: Authors' analysis of USGS Flood Inundation Mapping flood models for the Lower Blue River; Kansas City, MO, city parcel data; HUD administrative data

Overall, the authors' analysis found a higher share of housing units occupied by tenants participating in the tenant-based HCV (85.8 percent) program located in census tracts with Relatively Low, Very Low, and No Rating classifications for flood risk than both project-based HCV (75.0 percent) and project-based Section 8 (77.1 percent) program participants. This finding suggests that in KCMO, the tenant-based voucher program may provide opportunities for many recipients to live in areas of lower environmental risk and that project-based programs may concentrate recipients in areas of greater risk. However, the small share of housing units occupied by tenant-based HCV program participants living in census tracts identified as having a relatively high flood risk and areas overlapping with the predicted flood zone of the Lower Blue River suggests that opportunities exist to provide housing counseling to tenant-based voucher recipients during their search for housing to prevent them from renting housing units in high-risk flood areas, where they may be displaced.

Comparison of FEMA and USGS Flood Data

Twelve residential parcels were identified as occupied by households receiving federal rental assistance located in census tracts with a relatively high flood risk rating and 14 were identified in areas predicted to be affected by flooding from the Lower Blue River in USGS FIM maps. The authors expected a high degree of overlap between residential parcels in census tracts with a relatively high FEMA flood risk and areas that are predicted to be affected by flooding in USGS FIM models; however, only one parcel occupied by a household receiving federal rental assistance was found in a census tract with relatively high flood risk that overlapped a USGS predicted flood area. Of the remaining residential parcels, seven were in census tracts with relatively moderate flood risk and six with relatively low flood risk.

The lack of overlap between flood data sources is likely due to the different methods used to produce FEMA and USGS flood maps. FEMA risk ratings are calculated at the census-tract level on the basis of floodplain boundaries and historic storm and flood events. FEMA ratings represent a relative flood risk based on national ratings (Zuzak et al., 2023). Unlike the FEMA flood maps, the USGS FIM maps do not indicate the relative risk of inundation but only a detailed model of the extent and depth of inundation for a given flood stage. USGS flood inundation maps are developed using a variety of local data sources, including streamgage data, one-dimensional HEC-RAS hydraulic models, and topographic and bathymetric data collected along cross sections in the study reach. Thus, FEMA maps represent a relative measure of flood risk while USGS maps provide a model of expected local flood depth and extent. A combined analysis of these data sources identified 25 residential parcels occupied by federally assisted households within the study area at high risk of flooding. Although predicted risk does not guarantee flood exposure, the expected flood impacts on those housing units can be further explored through outreach and ground truthing.

The study area in this analysis was limited to the region in KCMO with available USGS FIM data containing the Lower Blue River and its tributaries (exhibit 1). The study area contains the majority of all residential parcels (60 percent) and the majority of residential parcels occupied by federally assisted households (75 percent) in KCMO. However, residential parcels located outside the study area may also be at risk of flooding: 2,077 additional residential parcels and 16 additional residential parcels occupied by households receiving federal rental assistance are located in census tracts with a Relatively High FEMA Riverine Flood Risk Rating outside the study area.

Conclusions

The USGS FIM Program provides a unique opportunity to help communities visualize potential areas at risk for flooding near local streams and rivers. Unlike FEMA flood risk maps, USGS FIM maps allow researchers to investigate local flooding processes, such as the predicted extent and depth of housing-unit flood exposure. To demonstrate the utility of USGS data for housing research, this paper presented a case study using these data to investigate the impact of flooding on housing units in KCMO, where households receiving federal rental assistance live. Integrating HUD administrative data with FEMA and USGS maps facilitated the identification of 25 housing units occupied by federally assisted households at high risk of flood exposure. This case study demonstrates how USGS and FEMA data can inform housing analyses at different scales for researchers and practitioners interested in flood impacts on local communities and vulnerable populations.

Both USGS and FEMA flooding data sources have limitations. USGS FIM maps are confined to certain stream reaches with USGS streamgages, making a national analysis impossible using these data. However, at the local level, USGS FIM maps allow researchers to identify the expected flood extent and depth for affected households. USGS FIM map libraries are available for 155 sites in 27 states. The presented analysis for KCMO can serve as a blueprint for studies replicated at other sites with available USGS FIM data. Although FEMA riverine flood risk maps are national in scope, the census tract flood risk measures are relative to national indicators of flood risk. Using multiple flooding data sources in a local analysis can help overcome gaps in data and provide information for affected households at multiple scales. For instance, multiscale flood vulnerability studies have identified that finer spatial scales (e.g., the census block versus the census tract) allow researchers to identify vulnerable communities overlooked at larger spatial scales (Remo, Pinter, and Mahgoub, 2016; Tanir et al., 2021). The presented analysis demonstrates this assertion at the parcel level. Six housing units occupied by federally assisted households at high risk of flooding from the Blue

River were identified on the basis of USGS FIM maps but located in census tracts with a relatively low FEMA flood risk.

This case study demonstrates the potential of using flood maps and parcel-level analysis to identify structures where households receiving federal rental assistance live in regions of high flood inundation and risk. Unlike census-tract-level analysis based on FEMA flood risk maps, combining parcel data and USGS FIM flood maps can give a more detailed picture of where the greatest damage from flooding may occur. However, both approaches can underestimate the impact of flooding in urban areas, where impervious surfaces and the capacity of stormwater systems can lead to flooding beyond floodplains. Recent calls point to limitations of existing flood maps and advocate for new analyses and maps to incorporate urban components that influence flooding (National Academies of Science, Engineering, and Medicine, 2019). Future research could draw on data sources and flood models that incorporate infrastructural elements to better describe the movement of water in urban landscapes. Future studies can also draw on the demographic variables present in HUD administrative data to understand who will be affected by flooding in addition to the impacts of flooding on structures where federally assisted households live.

Appendix: R Code for Analysis

```
#2023 Analysis for KC HUD - USGS Case
#for data cleaning
library(tidyverse)
#for working with spatial data
library(sf)
library(mapview) #interactive map viewer
library(tigris) #Census Tiger line shape files
#packages for raster data analysis
library(raster)
library(stars)
library(nngeo)
sf_use_s2(FALSE) # setting to prevent invalid loop error for st_join

#Data Setup
#Data Files
#HUD household - level data - internal to HUD
MOHUD_2022_new <- read_csv("Path to Household level lat lon data")
KCHUD_2022_sf <- MOHUD_2022_new %>%
        filter(!is.na(UNIT_LAT_DEG_MSRE)) %>%
        filter(UNIT_CITY_NAME == "Kansas City") %>%
        st_as_sf(., coords=c("UNIT_LGT_DEG_MSRE","UNIT_LAT_DEG_MSRE"), crs=4326) %>%
        filter(UNIT_LVL_CD == "R")

#FEMA National Risk Index for KC
NRI_KC <- st_read("NRI_KC.shp") #National census tract level file downloaded from
https://hazards.fema.gov/nri/data-resources#shpDownload
```

```
#Kansas Parcels
Parcels <- st_read("geo_export_a877e9f6-455d-48c1-b296-db1ee89e1444.shp") #Downloaded
from https://data.kcmo.org/dataset/Parcels/vuy6-s5is
LandCodes <- read.csv("Land_Use_Codes.csv") #Downloaded from https://data.kcmo.org/
Construction/Land-Use-Codes/83fx-3sa2

Parcels <- Parcels %>%
        left_join(., LandCodes, by =c("landusecod" = "Code"))

#Clip NRI data to KC and write out data file
NRI <- st_read("ShapeFiles/NRI_Shapefile_CensusTracts/NRI_Shapefile_CensusTracts.shp")
KC <- st_read("ShapeFiles/KC_Boundary/KC_Boundary.shp")
NRI <- st_transform(NRI, st_crs(KC))
NRI_KC <- st_intersection(st_make_valid(NRI), KC)
st_write(NRI_KC, "ShapeFiles/NRI_KC.shp")

#Reproject to USGS Flood Grid CRS
grids <- c("BlueR_12thStreet(06893590)_depth_grids/27", #1
    "BlueR_17thStreet(06893588)_depth_grids/26.flt", #2
    "BlueR_63rdStreet(06893530)_depth_grids/63rd_19.flt", #3
    "BlueR_BlueRidge(06893150)_depth_grids/26.flt", #4
    "BlueR_COave(06893553)_depth_grids/25.flt", #5
    "BlueR_Highway71(06893510)_depth_grids/hwy71mo_23.flt", #6
    "BlueR_KansasCity(BannisterRd_06893500)_depth_grids/28.flt", #7
    "BlueR_KennethRd(06893100)_depth_grids/moken_26.flt", #8
    "BlueR_RedBridge(06893195)_depth_grids/29.flt", #9
    "BlueR_StadiumDr(06893578)_depth_grids/mostad_28.flt", #10
    "BlueR_Stanley(06893080)_depth_grids/mostan_21.flt", #11
    "BrushCk_RockhillRd(06893562)_depth_grids/rockmo_22.flt", #12
    "BrushCk_Wardpkwy(06893557)_depth_grids/14/", #13
    "IndianCk_103rdStreet(06893400)_depth_grids/mo103rd_16.flt", #14
    "MORiver_Backwater(06893000)_depth_grids/morvrback_22.flt") #15

gridx <- read_stars(grids[15])
projection <- st_crs(gridx)
NRI_KC_Flood <- st_transform(NRI_KC, projection)

Parcels <- st_transform(Parcels, projection)
KCHUD_2022_sf <- st_transform(KCHUD_2022_sf, projection)

#Link HUD household data to Parcels (objectid) and NRI census track riverine flood risk index
(RFLD_RISKR)

#Point in Polygon for KCMO parcel data
HUD_Parcels <- st_join(st_make_valid(KCHUD_2022_sf), Parcels, join = st_intersects)
```

```
HUD_Parcels_count <- count(as_tibble(HUD_Parcels), objectid)
Parcels_HUD_sf <- left_join(Parcels, HUD_Parcels_count, by = "objectid") %>%
        rename(HUD_HH = n)

Parcels_HUD_sf <- HUD_Parcels %>%
left_join(Parcels_HUD_sf, ., by = "objectid")

#Point in Polygon for NRI flood data
Parcels_NRI <- st_join(st_centroid(Parcels_HUD_sf), NRI_KC_Flood, join = st_within)

#Filter residential land use codes
Parcels_NRI <- Parcels_NRI %>%
        st_drop_geometry() %>%
        filter(landusecod %in% c(1111, 1112, 1121, 1122, 1123, 1124, 1125, 1126))

#Define Study Area
studyArea <- data.frame(lon = c(-94.60825698480055, -94.4743909204287), lat =
c(38.84500509860179, 39.129928)) %>%
        st_as_sf(coords = c("lon", "lat"), crs = 4326) %>%
        st_bbox() %>%
        st_as_sfc()

studyArea <- st_transform(studyArea, projection)

#Clip Parcels to Study Area
myParcels <- Parcels[st_make_valid(studyArea), ] %>% dplyr::select(objectid, geometry)
Parcels_SA <- Parcels_NRI[st_make_valid(studyArea), ]

Parcels_SA_sf <- Parcels_SA %>%
        st_drop_geometry() %>%
        left_join(., myParcels, by = "objectid") %>%
        st_as_sf()

# Link Parcel data to USGS flood grids
##Area is in feet2 in both the parcel and flood files
#Function to calculate Flood area and depth
FloodStats <- function(grid, shape) {
        gridx <- read_stars(grid) %>%
                st_as_sf()
        mycol <- names(gridx)[1]
        shape <- st_transform(shape, st_crs(gridx))
        Parcelsx <- shape %>%
                st_join(., gridx) %>%
                drop_na(objectid) %>%
                st_drop_geometry() %>%
                group_by(objectid) %>%
```

```
                    summarise(MaxDeapth = max(!!as.name(mycol)), Count = n()) %>%
                    mutate(Area = ifelse(!is.na(MaxDeapth), Count * st_area(gridx[1,]), NA))

}

#Read in raster grids and calculate flood area and depth for parcels
SummaryStats1 <- FloodStats(grids[1], Parcels_SA_sf)
SummaryStats2 <- FloodStats(grids[2], Parcels_SA_sf)
SummaryStats3 <- FloodStats(grids[3], Parcels_SA_sf)
SummaryStats4 <- FloodStats(grids[4], Parcels_SA_sf)
SummaryStats5 <- FloodStats(grids[5], Parcels_SA_sf)
SummaryStats6 <- FloodStats(grids[6], Parcels_SA_sf)
SummaryStats7 <- FloodStats(grids[7], Parcels_SA_sf)
SummaryStats8 <- FloodStats(grids[8], Parcels_SA_sf)
SummaryStats9 <- FloodStats(grids[9], Parcels_SA_sf)
SummaryStats10 <- FloodStats(grids[10], Parcels_SA_sf)
SummaryStats11 <- FloodStats(grids[11], Parcels_SA_sf)
SummaryStats12 <- FloodStats(grids[12], Parcels_SA_sf)
SummaryStats13 <- FloodStats(grids[13], Parcels_SA_sf)
SummaryStats14 <- FloodStats(grids[14], Parcels_SA_sf)
SummaryStats15 <- FloodStats(grids[15], Parcels_SA_sf)

#Join Flood Parcel data frames
KCParcelFloodStats <- rbind(SummaryStats1[!is.na(SummaryStats1$MaxDeapth),],
        SummaryStats2[!is.na(SummaryStats2$MaxDeapth),],
        SummaryStats3[!is.na(SummaryStats3$MaxDeapth),],
        SummaryStats4[!is.na(SummaryStats4$MaxDeapth),],
        SummaryStats5[!is.na(SummaryStats5$MaxDeapth),],
        SummaryStats6[!is.na(SummaryStats6$MaxDeapth),],
        SummaryStats7[!is.na(SummaryStats7$MaxDeapth),],
        SummaryStats8[!is.na(SummaryStats8$MaxDeapth),],
        SummaryStats9[!is.na(SummaryStats9$MaxDeapth),],
        SummaryStats10[!is.na(SummaryStats10$MaxDeapth),],
        SummaryStats11[!is.na(SummaryStats11$MaxDeapth),],
        SummaryStats12[!is.na(SummaryStats12$MaxDeapth),],
        SummaryStats13[!is.na(SummaryStats13$MaxDeapth),],
        SummaryStats14[!is.na(SummaryStats14$MaxDeapth),],
        SummaryStats15[!is.na(SummaryStats15$MaxDeapth),])

#Only keep max flood depth for each parcel
KCParcelFloodStats_noDUP <- KCParcelFloodStats %>%
        group_by(objectid) %>%
        slice(which.max(MaxDeapth))

#Join HUD parcels with flood data
```

```
HUD_Parcels_Flood <- Parcels_SA_sf %>%
        left_join(., st_drop_geometry(KCParcelFloodStats_noDUP), by = "objectid") %>%
        filter(landusecod %in% c(1111, 1112, 1121, 1122, 1123, 1124, 1125, 1126))

#Assign Area and depth categories
#0-2 ft, 2-6 ft, 6-15 ft, 15+ ft
HUD_Parcels_Flood <- HUD_Parcels_Flood %>%
        mutate(DepthGroup = case_when(MaxDeapth < 2 ~ "< 2",
                MaxDeapth >= 2 & MaxDeapth < 6 ~ "2 - 6",
                        MaxDeapth >= 6 & MaxDeapth < 15 ~ "6 - 15",
                        MaxDeapth >= 15 ~ ">15"),
                AreaGroup = case_when(100*area/ShapeArea < 1 ~ "< 1",
                        100*area/ShapeArea >= 1 & 100*area/ShapeArea < 10 ~ "1 - 10",
                        100*area/ShapeArea >= 10 & 100*area/ShapeArea < 50 ~ "10 - 50",
                        100*area/ShapeArea >= 50 ~ "> 50"))
```

Acknowledgments

We would like to thank the editors of this symposium and the anonymous reviewers who provided comments on this paper. We would also like to thank U.S. Department of Housing and Urban Development and U.S. Geological Survey colleagues for providing internal reviews. Any use of trade, firm, or product names is for descriptive purposes only and does not imply endorsement by the U.S. Government.

Authors

Mariya Shcheglovitova is a social science analyst at the U.S. Department of Housing and Urban Development, Program Monitoring and Research Division. Gina Lee is a hydrologist at the U.S. Geological Survey, Maryland-Delaware-District of Colombia Water Science Center.

References

Bender, Kate. 2023. Parcels. https://data.kcmo.org/dataset/Parcels/vuy6-s5is. City of Kansas City, MO.

———. 2021. Land Use Codes. https://data.kcmo.org/Construction/Land-Use-Codes/83fx-3sa2. City of Kansas City, MO.

Brody, Samuel D., Sammy Zahran, Praveen Maghelal, Himanshu Grover, and Wesley E. Highfield. 2007. "The Rising Costs of Floods: Examining the Impact of Planning and Development Decisions on Property Damage in Florida," *Journal of the American Planning Association* 73 (3): 330–45.

Byun, Kyuhyun, Chun-Mei Chiu, and Alan F. Hamlet. 2019. "Effects of 21st Century Climate Change on Seasonal Flow Regimes and Hydrologic Extremes Over the Midwest and Great Lakes Region of the US," *Science of the Total Environment* 650: 1261–1277.

Carrera, Jennifer S., and Catherine Coleman Flowers. 2018. "Sanitation Inequity and the Cumulative Effects of Racism in Colorblind Public Health Policies," *American Journal of Economics and Sociology* 77 (3–4): 941–66.

Chakraborty Jayajit, Timothy W. Collins, and Sara E. Grineski. 2019. "Exploring the Environmental Justice Implications of Hurricane Harvey Flooding in Greater Houston, Texas," *American Journal of Public Health* 109 (2): 244–50.

Chakraborty, Jayajit, Timothy W. Collins, Marilyn C. Montgomery, and Sara E. Grineski. 2014. "Social and Spatial Inequities in Exposure to Flood Risk in Miami, Florida," *Natural Hazards Review* 15 (3): 0401400.

Chakraborty Jayajit, Sara E. Grineski, and Timothy W. Collins. 2019. "Hurricane Harvey and People with Disabilities: Disproportionate Exposure to Flooding in Houston, Texas," *Social Science and Medicine* 226: 176–81.

Chakraborty, Jayajit, Ashley A. McAfee, Timothy W. Collins, and Sara E. Grineski. 2021. "Exposure to Hurricane Harvey Flooding for Subsidized Housing Residents of Harris County, Texas," *Natural Hazards* 106 (3): 2185–205.

City of Kansas City, Missouri. 2021. "New Housing Department to Launch." https://www.kcmo.gov/Home/Components/News/News/1606/16.

Collins, Timothy W., Sarah E. Grineski, Jayajit Chakraborty, and Aaron B. Flores. 2019. "Environmental Injustice and Hurricane Harvey: A Household-Level Study of Socially Disparate Flood Exposures in Greater Houston, Texas, USA," *Environmental Research* 179 (Part A): 108772.

Cutter, Susan L., Bryan J. Boruff, and W. Lynn Shirley. 2003. "Social Vulnerability to Environmental Hazards," *Social Science Quarterly* 84 (2): 242–61.

Davlasheridze, Meri, and Qing Miao. 2021. "Natural Disasters, Public Housing, and the Role of Disaster Aid," *Journal of Regional Science* 61 (5): 1113–35.

Driever, Steven L., and Danny M. Vaughn. 1988. "Flood Hazard in Kansas City Since 1880," *Geographical Review* 78 (1): 1–19.

Federal Emergency Management Agency (FEMA). n.d. *Riverine Flooding*. https://hazards.fema.gov/nri/riverine-flooding.

———. 2014. *Homeowner's Guide to Retrofitting: Six Ways to Protect Your Home from Flooding*, 3rd ed. Washington, DC: FEMA. https://www.fema.gov/sites/default/files/2020-08/FEMA_P-312.pdf.

Hallegatte, Stéphane, and Valentin Przyluski. 2010. The Economics of Natural Disasters: Concepts and Methods. Policy research working paper 5507. Washington, DC: The World Bank.

Hamideh, Sara, and Jane Rongerude. 2018. "Social Vulnerability and Participation in Disaster Recovery Decisions: Public Housing in Galveston After Hurricane Ike," *Natural Hazards* 93 (3): 1629–48.

Heimann, David C., T.E. Weilert, B.P. Kelly, and S.E. Studley. 2014. *Flood-Inundation Maps and Wetland Restoration Suitability Index for the Blue River and Selected Tributaries, Kansas City, Missouri, and Vicinity, 2012* (ver. 1.1, April 2015). U.S. Geological Survey Scientific Investigations Report 2014–5180. Reston, VA: USGS. http://dx.doi.org/10.313/sir20145180.

Hendricks, Marccus D., and Shannon Van Zandt. 2021. "Unequal Protection Revisited: Planning for Environmental Justice, Hazard Vulnerability, and Critical Infrastructure in Communities of Color," *Environmental Justice* 14 (2): 87–97.

Hernández, Diana, David Chang, Carole Hutchinson, Evanah Hill, Amenda Almonte, Rachel Burns, Peggy Shepard, Ingrid Gonzalez, Nora Reissig, and David Evans. 2018. "Public Housing on the Periphery: Vulnerable Residents and Depleted Resilience Reserves Post-Hurricane Sandy," *Journal of Urban Health* 95 (5): 703–15.

Howell, Junia, and James R. Elliott. 2019. "Damages Done: The Longitudinal Impacts of Natural Hazards on Wealth Inequality in the United States," *Social Problems* 66 (3): 448–67.

Intergovernmental Panel on Climate Change (IPCC). 2022. *Climate Change 2022: Impacts, Adaptation, and Vulnerability. Working Group II Contribution to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change* [H.O. Pörtner, D.C. Roberts, M. Tignor, E.S. Poloczanska, K. Mintenbeck, A. Alegría, M. Craig, S. Langsdorf, S. Löschke, V. Möller, A. Okem, B. Rama (eds.)]. Cambridge University Press. Cambridge, UK: Cambridge University Press. https://doi.org/10.1017/9781009325844.

Lee, Jee Young, and Shannon Van Zandt. 2018. "Housing Tenure and Social Vulnerability to Disasters: A Review of the Evidence," *Journal of Planning Literature* 34 (2): 156–70.

Levine, Joyce N., Ann-Margaret Esnard, and Alka Sapat. 2007. "Population Displacement and Housing Dilemmas Due to Catastrophic Disasters," *Journal of Planning Literature* 22 (1): 3–15.

Mehta, Aditi, Mark Brennan, and Justin Steil. 2020. "Affordable Housing, Disasters, and Social Equity: LIHTC as a Tool for Preparedness and Recovery," *Journal of the American Planning Association* 86 (1): 75–88.

Missouri Department of Public Safety. n.d. Missouri Floodplain Management/Floodplain Insurance Programs. https://sema.dps.mo.gov/programs/floodplain/.

National Academies of Science, Engineering, and Medicine. 2019. "New Report Calls for Different Approaches to Predict and Understand Urban Flooding." News release. Washington, DC: National Academies of Sciences. https://www.nationalacademies.org/news/2019/03/new-report-calls-for-different-approaches-to-predict-and-understand-urban-flooding#:~:text=The%20report%20calls%20for%20multi,population%20of%20the%20United%20States.

Pebesma, Edzer. 2018. "Simple Features for R: Standardized Support for Spatial Vector Data," *The R Journal* 10 (1): 439–46.

Pebesma, Edzer, and Roger Bivand. 2023. *Spatial Data Science: With Applications in R*. London: Chapman and Hall/CRC. https://doi.org/10.1201/9780429459016.

Qiang, Yi. 2019. "Disparities of Population Exposed to Flood Hazards in the United States," *Journal of Environmental Management* 232: 295–304.

The R Foundation. 2022. The R Project for Statistical Computing. https://www.R-project.org/.

Remo, Jonathan W. F., Nicholas Pinter, and Moe Mahgoub. 2016. "Assessing Illinois's Flood Vulnerability Using Hazus-MH," *Natural Hazards* 81 (1): 265–87.

Rosoff, Stephanie, and Jessica Yager. 2017. "Housing in the U.S. Floodplains." Data brief. New York University Furman Center for Real Estate and Urban Policy.

Rufat, Samuel, Eric Tate, Christopher G. Burton, and Abu Maroof. 2015. "Social Vulnerability to Floods: Review of Case Studies and Implications for Measurement," *International Journal of Disaster Risk Reduction* 14: 470–86.

Shuster, William D., James Bonta, Hale W. Thurston, Elizabeth Warnemuende, and D. R. Smith. 2005. "Impacts of Impervious Surface on Watershed Hydrology: A Review," *Urban Water Journal* 2 (4): 263–75.

Smiley, Kevin T. 2020. "Outdated and Inaccurate, FEMA Flood Maps Fail to Fully Capture Risk," *Urban Edge: Insights*, September 30. https://kinder.rice.edu/urbanedge/outdated-and-inaccurate-fema-flood-maps-fail-fully-capture-risk.

Tanir, Tugkan, Selina J. Sumi, Andre de Souza de Lima, Gustavo de A. Coelho, Sukru Uzun, Felicio Cassalho, and Celso M. Ferreira. 2021. "Multi-Scale Comparison of Urban Socio-Economic Vulnerability in the Washington, DC Metropolitan Region Resulting from Compound Flooding," *International Journal of Disaster Risk Reduction* 61: 102362.

Ueland, Jeff, and Barney Warf. 2006. "Racialized Topographies: Altitude and Race in Southern Cities," *Geographical Review* 96 (1): 50–78.

U.S. Census Bureau. 2023. Kansas City City, Missouri Profile. https://data.census.gov/profile/Kansas_City_city,_Missouri?g=160XX00US2938000. Accessed July 7, 2023.

U.S. Department of Housing and Urban Development (HUD). n.d.a. "Climate Action for Resilient, Livable, and Equitable Communities," *PD&R Edge*. https://www.huduser.gov/portal/pdredge/pdr_edge_featd_article_091415.html.

———. n.d.b. Picture of Subsidized Households. Washington, DC: HUD Office of Policy Development and Research. https://www.huduser.gov/portal/datasets/assthsg.html.

U.S. Geological Survey (USGS). 2019. Flood Inundation Mapper. https://www.usgs.gov/tools/flood-inundation-mapper. Accessed June 10, 2023.

———. 2023a. USGS Surface-Water Data for the Nation. http://waterdata.usgs.gov/nwis/sw. Accessed July 2023.

———. 2023b. USGS Watershed Boundary Dataset (WBD) for 2-digit Hydrologic Unit – 07, May 25. https://www.usgs.gov/national-hydrography/watershed-boundary-dataset.

Vanucchi, Jamie L. 2021. "Adapting Inland Floodplain Housing to a Changing Climate: Disturbance, Risk, and Uncertainty as Drivers for Design." In *Climate Adaptation and Resilience Across Scales: From Buildings to Cities*, edited by Nicholas B. Rajkovich and Seth H. Holmes. New York: Routledge: 172–88.

Walker, Gordon. 2012. "Flood Vulnerability: Uneven Risk and Injustice of Disaster." Chapter 6 in *Environmental Justice: Concepts, Evidence, and Politics*. London: Routledge: 156–178.

Wilkison, Donald H., Daniel J. Armstrong, and Dale W. Blevins. 2002. *Effects of Wastewater and Combined Sewer Overflows on Water Quality in the Blue River Basin, Kansas City, Missouri and Kansas, July 1998–October 2000*. U.S. Geological Survey Water-Resources Investigations Report 02–4107. Reston, VA: USGS. https://pubs.usgs.gov/wri/wri02-4107/pdf/wri02-4107.pdf.

Wilkison, Donald H., Daniel J. Armstrong, Richard D. Norman, Barry C. Poulton, Edward T. Furlong, and Steven D. Zaugg. 2006. *Water Quality in the Blue River Basin, Kansas City Metropolitan Area, Missouri and Kansas, July 1988 to October 2004*. U.S. Geological Survey Scientific Investigations Report 2006–5147. Reston, VA: USGS. https://pubs.usgs.gov/sir/2006/5147/pdf/sir2006-5147.pdf.

Zuzak, Casey, Anne Sheehan, Emily Goodenough, Alice McDougall, Carly Stanton, Patrick McGuire, Matthew Mowrer, Benjamin Roberts, and Jesse Rozelle. 2023. *National Risk Index: Technical Documentation*. Washington, DC: Federal Emergency Management Agency.

The Health Status of Women with Children Living in Public and Assisted Housing: Linkage of the National Health Interview Survey to U.S. Department of Housing and Urban Development Administrative Data

Veronica Helms Garrison
Jacqueline V. Bachand
U.S. Department of Housing and Urban Development, Office of Policy Development and Research

Cindy Zhang
Christine Cox
Cordell Golden
Kimberly A. Lochner
National Center for Health Statistics, Centers for Disease Control and Prevention

Disclaimer: The findings and conclusions in this manuscript are those of the authors and do not necessarily represent the official position of the U.S. Department of Housing and Urban Development or the Centers for Disease Control and Prevention.

Abstract

For more than a decade, the U.S. Department of Housing and Urban Development (HUD) and the National Center for Health Statistics (NCHS) have partnered to link NCHS national health survey data with HUD administrative records on persons participating in federal public and assisted housing programs.

This study used 2015–18 National Health Interview Survey (NHIS)-HUD linked data to examine women 18–44 years old with children and renting their home who were receiving HUD assistance (n=852) and a comparison population of women of the same age with children, who were low-income renters but did not link to HUD records at the time of their NHIS interview (n=894).

Abstract (continued)

The population of HUD-assisted women differed from the comparison group on key sociodemographic characteristics and health indicators. HUD-assisted women were more likely to report their health as fair or poor and to being a current smoker. HUD-assisted women also were less likely to be uninsured and more likely to have a regular source of care.

The findings in this article are exploratory but demonstrate how the NCHS-HUD-linked data can be a resource for researchers and policymakers in further examining housing status as an important social determinant of health.

Introduction

Social determinants of health (SDOH) are the nonmedical conditions that influence health, such as birthplace, living situation, work, and age (Braveman and Gottlieb, 2014; World Health Organization, n.d.). Within that context, access to safe, affordable, and stable housing is increasingly identified as an important SDOH (Krieger and Higgins, 2002; Swope and Hernández, 2019). A large share of American households experience "worst case housing needs,"[1] a long-standing U.S. Department of Housing and Urban Development (HUD) measure that seeks to quantify the national extent of unmet housing needs for affordable and quality rental housing. A recent HUD report to the U.S. Congress estimated that approximately 8 million renter households faced substantial worst case housing issues in 2019, including 2.2 million households with children, a number that translates to approximately 40 percent of U.S. families with children (Alvarez and Steffen, 2021).

HUD federal housing subsidy programs aim to reduce unmet housing needs by subsidizing rental costs for safe and affordable units and support HUD's mission to "create strong, sustainable, inclusive communities and quality affordable homes for all" (HUD, n.d.b.). Although public rental assistance programs seek to alleviate housing affordability issues for households with lower incomes, the need for housing assistance greatly outweighs the demand. According to the 2021 American Housing Survey, only approximately one-quarter of renter households eligible for housing assistance receive it (U.S. Census Bureau, 2021).

Women and children represent more than 75 percent of HUD-assisted people. HUD provides housing rental assistance to approximately 9 million people annually, including approximately 4.2 million women and 3.2 million children, including foster children, from birth to age 17 (HUD, 2022). HUD is increasingly committed to better understanding the health and social needs of women and children living in public and assisted housing. The current HUD Strategic Plan for Fiscal Years (FY) 2022–2026 states that "housing is the foundation on which we live, grow, and thrive" and includes a milestone to "improve maternal and child health outcomes" (HUD, n.d.a.).

[1] Worst case housing needs for renter households are defined as having household incomes at or below 50 percent of the Area Median Income, not receiving government housing assistance, and paying more than one-half of their income for rent, living in severely inadequate conditions, or both (https://www.huduser.gov/portal/sites/default/files/pdf/Worst-Case-Housing-Needs-2021.pdf).

Further, to support the *White House Blueprint for Addressing the Maternal Health Crisis* report, HUD has committed to conducting interagency research to better understand the unique health and social needs of HUD-assisted women and children (The White House, 2022).

A growing body of research underscores a link between housing stability attributable to federal housing assistance programs and health outcomes. Housing assistance has been shown to be associated with reduced odds of self-reported fair or poor health and psychological distress for participants in the public housing program, but results were not consistent across HUD program types (Fenelon et al., 2017). Another study found that receipt of housing assistance was associated with lower uninsured rates and lower unmet healthcare needs (Simon et al., 2017). A HUD report highlighting the characteristics of HUD-assisted adults found that approximately 74 percent of adults identified as female reported high rates of health conditions, medical diagnoses, and healthcare utilization (Helms, Sperling, and Steffen, 2017).

Previous randomized control trials also showed a strong link between housing status and maternal and child health outcomes. The Family Options Study, a large-scale experiment for homeless families with young children, showed that families who received housing vouchers (the most sustainable and long-term intervention option) reported fewer child separations, decreased maternal psychological distress, decreased economic stress, fewer child behavior problems, and less household food insecurity (Gubits et al., 2016, Shinn et al., 2016). Still, limited data exist that allow for the examination of health outcomes for women with children receiving HUD assistance.

For more than a decade, the HUD Office of Policy Development and Research (PD&R) and the National Center for Health Statistics (NCHS), Centers for Disease Control and Prevention have partnered to link NCHS national health survey data with HUD administrative records[2] on people participating in federal public and assisted housing programs. The availability of those linked data sources provides a unique opportunity to examine housing status as an important social determinant of health for lower-income women with children.

Using the 2015–18 National Health Interview Survey (NHIS) data linked to HUD housing assistance data, this study examined women 18–44 years with children, who were renting their home and receiving HUD assistance, and a comparison population of women of the same age with children, who were low-income renters but did not link to HUD records at the time of their NHIS interview. Leveraging this linked data source may allow for a better understanding of the health status and healthcare access of women with children receiving HUD housing assistance compared with a group likely eligible for housing assistance but not receiving it.

Methods

Data Sources

The discussed study used the National Health Interview Survey linked with HUD administrative data on public and assisted housing programs. The following section provides more information about each data source.

[2] NCHS Data Linked to HUD Housing Assistance Program Files: https://www.cdc.gov/nchs/data-linkage/hud.htm.

National Health Interview Survey. NHIS is a nationally representative cross-sectional household interview survey conducted continuously throughout the year by NCHS. NHIS is designed to monitor the health of the civilian noninstitutionalized U.S. population by collecting data on a broad range of health topics. The sample probability design permits a representative sampling of households and noninstitutional group quarters. The core questionnaire contained four major components for the 2015–18 NHIS. The household composition component collected basic demographic and relationship information for all persons living in the housing unit. The family component collected basic demographic, health insurance, and health information about all family members from a single family member. The sample adult core and sample child core components included one randomly selected adult (aged 18 or older) and one randomly selected child (aged 17 or younger) if the family included children. Detailed information regarding the design, content, use of NHIS, annual sample sizes, and NHIS response rates are available in the annual NHIS Survey Description documents (National Center for Health Statistics, 2016, 2017, 2018, and 2019). For this analysis, information collected from the family, person, and sample adult core components was used. Information collected in the sample adult core component is self-reported unless the person was physically or mentally unable to provide it, in which case a knowledgeable proxy could answer.

HUD Administrative Data on Public and Assisted Housing Programs. HUD is the primary federal agency overseeing domestic housing programs and policies. HUD programs can be lumped broadly into two categories: project-based and tenant-based housing. In project-based housing, the subsidy is tied to a physical unit. HUD project-based programs include the public housing (PH) program and various multifamily housing (MF) programs, such as Project-Based Section 8, Section 811, and Section 202. The subsidy is tied to the household in HUD's sole tenant-based housing program, the Housing Choice Voucher (HCV) program. The HCV program is thus unique from other HUD programs in that HCV households can enter the private housing market and have a greater choice regarding the unit they rent.

HUD collects detailed administrative data on families participating in its programs through administrative forms.[3] For PH and HCV programs, data are collected via housing agencies at the local or state level. Data for MF program types are collected through owners of private buildings (NCHS, 2019b). This analysis focused on persons in HUD subsidized housing in HUD's three largest housing assistance programs: HCV, PH, and MF. All three program categories use three factors to determine program eligibility: (1) U.S. citizenship or eligible immigration status, (2) family size, and (3) gross annual income. Some HUD program categories (HCV and PH) can establish waitlist preferences for special populations, such as older adults, unsheltered families, or persons living with a disability, but the presence of those populations in a household does not affect overall program eligibility.

[3] Administrative Form HUD-50058, the "Family Report," is used to collect data on the people who participate in the HCV and public housing programs (https://www.hud.gov/sites/dfiles/OCHCO/documents/50058.PDF). PHAs participating in the Moving to Work (MTW) demonstration have fewer data requirements; Administrative Form HUD-50058 MTW ("MTW Family Report") is used only by PHAs participating in MTW (http://portal.hud.gov/hudportal/documents/huddoc?id=DOC_10236.pdf). Administrative Form HUD-50059, "Owners Certification of Compliance with HUD's Tenant Eligibility and Rent Procedures," is used to collect data on the people who participate in multifamily programs (http://portal.hud.gov/hudportal/documents/huddoc?id=50059.pdf).

NHIS-HUD Linked Data. The study used 2015–18 NHIS-HUD linked data. NHIS participants were considered linkage-eligible during the 2015–18 survey years if they provided the last four digits of their Social Security number (SSN4) or an affirmative response to the followup question to allow linkage without SSN4 and sufficient information for linkage, such as date of birth, sex, first name, and last name. Linkage eligibility is distinct from HUD program eligibility, which defines whether a person meets the eligibility criteria for HUD housing assistance. The linkage was conducted using both deterministic and probabilistic approaches. For the probabilistic linkage process, scoring was conducted according to the Fellegi-Sunter methodology (Fellegi and Sunter, 1969). Linkage methods and evaluation of NHIS-HUD linked data, including validation and quality control processes, are described in more detail elsewhere (National Center for Health Statistics, 2022). Due to confidentiality requirements, the restricted-use NCHS-HUD data are accessible only through the NCHS and Federal Statistical Research Data Centers. The linkage of NHIS-HUD data was conducted through a memorandum of understanding between NCHS and HUD. The data linkage work was performed at NCHS with approval provided by NCHS's Research Ethics Review Board.

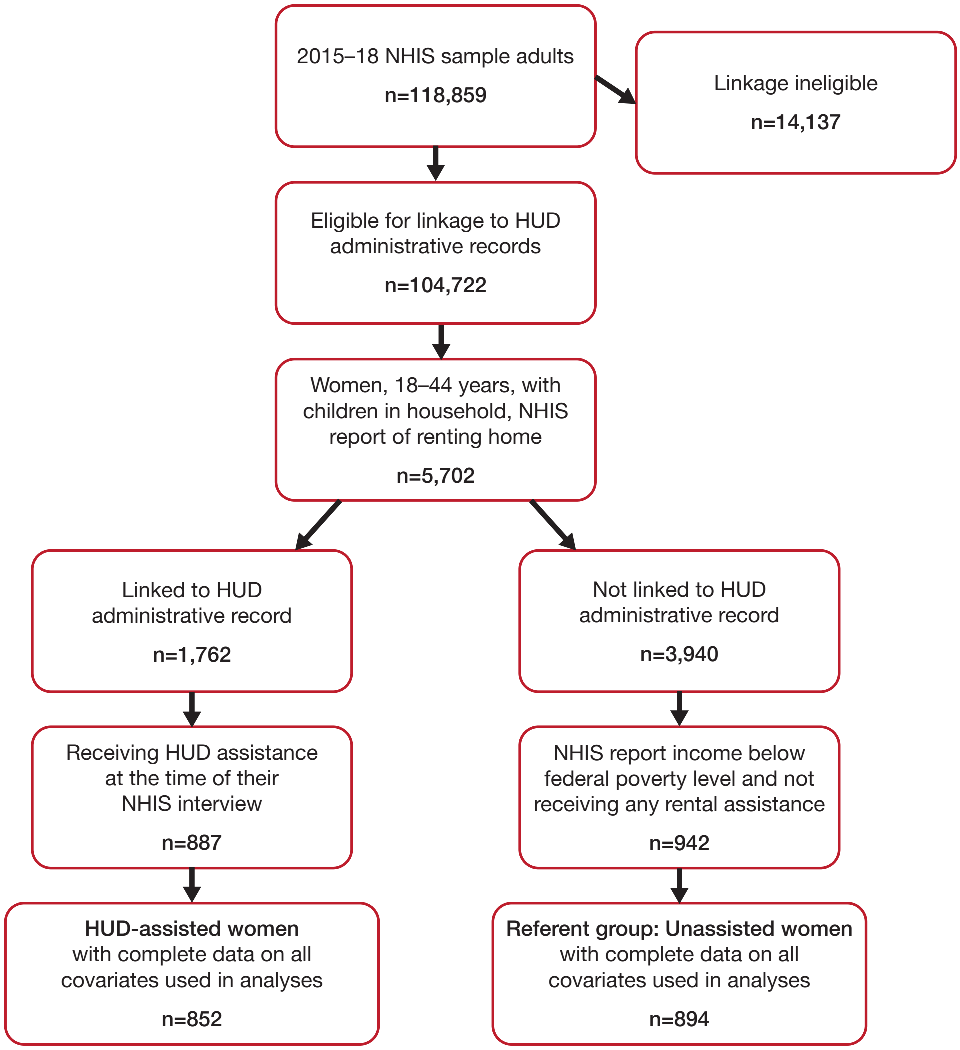
Study Population

For the pooled NHIS years 2015–18, 104,722 (88 percent) sample adults were eligible for linkage to HUD administrative records. The analytic sample was restricted to female NHIS sample adult participants who were eligible for linkage, 18–44 years of age, the parent of at least one child (0–17 years of age) residing in the family household at the time of interview, and renting their home (n=5,702). Among this group, 1,762 were linked to a 2015–18 HUD administrative data record, with 887 determined to be receiving HUD assistance at the time of their NHIS interview, on the basis of episode-level data. An *episode* is defined as a single continuous period of enrollment in a HUD program on the basis of dates of HUD transactions. The group of 3,940 who did not link was further restricted to having a family income below the federal poverty threshold defined by the Census Bureau (U.S. Census Bureau, n.d.) and did not report receiving any type of rental assistance at the time of the NHIS interview (n=942). For the group that did not link to HUD administrative data concurrent with their interview, restricting them to having a family income below the federal poverty level was used to approximate the income requirement for HUD program eligibility. Gross annual income is a significant program eligibility factor, and most HUD-assisted households have annual incomes below the federal poverty threshold.

The final analytic sample was restricted to those with complete information on all variables used in the analyses, resulting in 852 receiving HUD assistance at the time of the NHIS interview, referred to as "HUD-assisted women," and 894 who did not link to a HUD record, referred to as "unassisted women." Limiting results to complete information on all variables reduced the sample by approximately 4 percent for the HUD-assisted group and 5 percent for the comparison group. Exhibit 1 provides a complete description of the creation of the analytic sample.

Exhibit 1

Analytic Sample: 2015–18 National Health Interview Survey Sample Adult Participants Linked to U.S. Department of Housing and Urban Development Program Data



HUD = Housing and Urban Development. NHIS = National Health Interview Survey.

Notes: HUD-assisted *refers to NHIS sample adults eligible for linkage—women 18–44 years with children and linked to a HUD administrative record concurrent with their NHIS interview.* Unassisted renters *refers to NHIS sample adults eligible for linkage—women 18–44 years with children, not linked to a HUD administrative record, reported on the NHIS incomes below the federal poverty threshold, and renting their home.*

Source: 2015–18 National Health Interview Survey, linked to U.S. Department of Housing and Urban Development (HUD) program data (2015–18)

Measures

All variables examined in this analysis were based on self-reported information at the time of NHIS interview.

Sociodemographic Characteristics. Sociodemographic characteristics of the NHIS adult women with children examined in this article include age at interview (18–24, 25–34, and 35–44), race and Hispanic origin (Hispanic or Latino; Black or African-American non-Hispanic—hereafter, Black; Other non-Hispanic—hereafter, Other; and White non-Hispanic—hereafter, White), educational attainment (less than high school; high school diploma or GED; and some college or higher), marital status (married or living with partner; widowed, divorced, or separated; and never married), region (Northeast, Midwest, South, and West), number of adults in the household (one, two or more), and work status in past 12 months (Yes or No).

Health Characteristics. Five health characteristics were examined, spanning health status, health behavior, and healthcare access. Health status was measured as self-rated health, based on response to the following question: "Would you say your health in general is excellent, very good, good, fair, or poor?" Health status was recoded into fair or poor, good, and very good or excellent. Serious psychological distress was measured using the Kessler 6 index, which consists of six questions asking about various feelings of distress: "During the past 30 days, how often did you feel (1) so sad that nothing could cheer you up, (2) nervous, (3) restless or fidgety, (4) hopeless, (5) that everything was an effort, and (6) worthless?" Response categories included "all," "most," "some," "little," or "none of the time," with total values representing a scale of 0–24 points. Consistent with other studies, this variable was recoded with a score of 13 or more, indicating the presence of serious psychological distress (Kessler et al., 2002). Current cigarette smokers were defined as adults who had smoked 100 cigarettes during their entire life and answered "every day" or "some days" to the question, "Do you now smoke cigarettes every day, some days, or not at all?" NHIS-reported health insurance status was recoded into three categories: "public or other health insurance," "private health insurance," or "uninsured." Women were considered to have public health insurance if they reported coverage through Medicare, Medicaid, a State Children's Health Insurance Program or other state-sponsored health plan, or other government or military health plans. NHIS participants were asked if they had a usual source of health care via the following question: "Is there a place that you usually go to when you are sick or need advice about your health?"

Analytic Approach

Data analysis was performed using SAS software, version 9.4 (published by SAS Institute, Cary NC), and SAS-callable SUDAAN, version 11.03 (published by RTI, Research Triangle Park, NC), a software package that accounts for the complex sample design of NHIS. Estimates in this article were calculated using the NHIS sample adult sampling weights, which account for nonresponse and unequal probabilities of selection and were adjusted further to account for linkage eligibility (National Center for Health Statistics, 2022). Standard errors were estimated, accounting for the NHIS complex survey design. Differences were evaluated with two-sided Wald tests at the 0.05 significance level. All estimates presented in this article met NCHS standards of reliability (Parker et al., 2017).

Results

Characteristics of the Study Population

The study population consisted of 852 HUD-assisted women and 894 unassisted women. Both groups were restricted to women between the ages of 18 and 44, with children aged 0 to 17, who reported renting their homes. Exhibit 2 displays the weighted percentages of the selected sociodemographic characteristics for both groups. Distributions of HUD-assisted and unassisted women were significantly different on all sociodemographic characteristics examined.

Exhibit 2

Sociodemographic Characteristics of Women 18–44 Years with Lower Incomes, Who Have Children in the Household and Rent Their Housing, by Federal Housing Assistance Status, 2015–18

Characteristic	HUD-assisted Renters (n=852)			Unassisted Renters (n=894)		
	n	%	SE	n	%	SE
Age (years)[a]						
18–24	110	15.4	1.9	128	16.3	1.7
25–34	455	51.9	2.2	416	42.2	2.0
35–44	287	32.7	2.2	350	41.5	2.1
Race and Ethnicity[c]						
Black, non-Hispanic	411	51.4	2.7	119	14.1	1.5
Other, non-Hispanic	42	4.6	1.0	47	8.3	1.5
White, non-Hispanic	215	23.6	2.2	278	30.2	2.1
Hispanic	184	20.4	2.1	450	47.4	2.2
Marital Status[c]						
Married/Living with Partner	123	20.5	1.9	437	61.4	2.0
Widowed/Divorced/Separated	163	15.8	1.5	205	16.2	1.5
Never Married	566	63.8	2.2	252	22.4	1.6
Number of Adults in Household[c]						
1	677	66.9	2.1	409	29.4	1.7
2+	175	33.1	2.1	485	70.6	1.7
Region[c]						
Northeast	174	22.3	2.7	121	16.7	1.8
Midwest	164	19.9	2.2	129	14.4	1.6
South	383	43.7	3.0	363	40.1	2.3
West	131	14.1	2.1	281	28.8	2.1
Educational Attainment[b]						
Less than High School	218	24.7	1.8	331	34.6	2.1
High School Diploma or GED	262	31.9	2.0	265	32.0	2.1
Some College or Higher	372	43.4	2.1	298	33.4	2.0
Work Status: Past 12 Months[c]						
Yes	579	67.1	2.0	546	54.8	2.1
No	273	32.9	2.0	348	45.2	2.1

SE = standard error. HUD = U.S. Department of Housing and Urban Development.

[a] Wald test comparing HUD-assisted renters to unassisted renters, $p<0.01$.

[b] Wald test comparing HUD-assisted renters to unassisted renters, $p<0.001$.

[c] Wald test comparing HUD-assisted renters to unassisted renters, $p<0.0001$.

Source: 2015–18 National Health Interview Survey, linked to U.S. Department of Housing and Urban Development (HUD) program data (2015–18)

HUD-assisted women tended to be between the ages of 25 and 34 (51.9 percent), Black (51.4 percent), never married (63.8 percent), and the only adult in their household (66.9 percent). Most unassisted women were Hispanic (47.4 percent) or White (30.2 percent), married or living with a partner (61.4 percent), and had one or more other adults also living in their household (70.6 percent).

When examining the socioeconomic and educational characteristics of the two groups, most HUD-assisted women reported working during the previous year (67.1 percent), and a little more than one-half (54.8 percent) of unassisted women reported working during the previous year. Approximately 24.7 percent of HUD-assisted women with children reported less than a high school diploma, and 43.4 percent reported some college education. Among unassisted women, 34.6 percent reported less than a high school diploma, 32.0 percent reported a high school diploma or GED, and 33.4 percent reported some level of college.

Health Characteristics of the Study Population

Exhibit 3 shows weighted percentages of selected health characteristics. The distributions of self-reported health and current smoking status were significantly different across the two groups. Among HUD-assisted women, 21.0 percent reported their health as fair or poor, and 46.0 percent reported their health status as very good or excellent, whereas 11.3 percent of unassisted women reported their health status as fair or poor, and 54.1 percent reported very good or excellent. The percentage of HUD-assisted women who were current smokers was 1.7 times higher than unassisted women (30.1 percent vs. 17.4 percent); however, reports of serious psychological distress did not differ between the two groups.

Exhibit 3

Health Status and Healthcare Access of Women 18–44 years with Lower Incomes, Who Have Children in the Household and Rent Their Housing, by Federal Housing Assistance Status, 2015–18 (1 of 2)

Characteristic	HUD-assisted Renters (n=852)			Unassisted Renters (n=894)		
	n	%	SE	n	%	SE
Self-reported General Health[a]						
Fair/Poor	181	21.0	1.7	117	11.3	1.2
Good	294	33.0	2.0	300	34.6	2.2
Very Good/Excellent	377	46.0	2.2	477	54.1	2.2
Serious Psychological Distress						
Yes	70	8.1	1.3	59	7.1	1.2
No	782	91.9	1.3	835	92.9	1.2
Current Cigarette Smoker[a]						
Yes	266	30.1	2.1	170	17.4	1.5
No	586	69.9	2.1	724	82.6	1.5
Health Insurance Status[a]						
Public/Other	585	69.0	2.4	471	55.3	2.2
Private	135	15.1	1.7	120	12.2	1.3
Uninsured	132	15.9	1.9	303	32.5	2.0

Exhibit 3

Health Status and Healthcare Access of Women 18–44 years with Lower Incomes, Who Have Children in the Household and Rent Their Housing, by Federal Housing Assistance Status, 2015–18 (2 of 2)

Characteristic	HUD-assisted Renters (n=852)			Unassisted Renters (n=894)		
	n	%	SE	n	%	SE
Usual Source of Care[a]						
Yes	709	82.6	1.8	666	71.9	2.0
No	143	17.4	1.8	228	28.1	2.0

SE = standard error. HUD = U.S. Department of Housing and Urban Development.
[a]Wald test comparing HUD-assisted to unassisted renters, p<0.0001.
Source: 2015–18 National Health Interview Survey, linked to U.S. Department of Housing and Urban Development program data (2015–18)

When examining measures related to access to care, health insurance coverage and having a usual source of care varied between the two groups. Most HUD-assisted women reported having public health insurance (69.0 percent), and 15.9 percent reported being uninsured, whereas unassisted women reported higher levels of being uninsured (32.5 percent) and having no usual source of care (28.1 percent).

Discussion

Using the linked 2015–18 NHIS-HUD data, this article presents a descriptive summary of health characteristics and healthcare access for women with children receiving HUD assistance. The findings in this article build upon two recent HUD "health pictures" of adults and children receiving HUD assistance (Helms, Sperling, and Steffen, 2017; Helms et al., 2018) and further support initiatives to better understand the health needs of HUD-assisted women and children (HUD, n.d.a.; The White House, 2022). The study examined women 18–44 years old with children, who were renting their home and receiving HUD assistance, and a comparison population of women of the same age with children, who were low-income renters but did not link to HUD records at the time of their NHIS interview. Because income is a significant factor for HUD program eligibility and only approximately one in four households eligible for housing assistance receive it, the comparison population attempts to reflect women with children who are eligible for housing assistance but are not receiving it. The population of HUD-assisted women differed from the comparison group on key sociodemographic characteristics and health indicators.

Although HUD-assisted and unassisted women had similar percentages of women who were 18–24 years old, approximately 50 percent and 30 percent of HUD-assisted women were 25–34 and 35–44 years of age, respectively, compared with approximately 40 percent of unassisted women in each of those age groups. Most HUD-assisted women were Black non-Hispanic, although nearly one-half of unassisted women were Hispanic. Only approximately one in five HUD-assisted women reported being married or living with a partner, and two-thirds reported being the only adult in their household. Among unassisted women, most (61 percent) reported being married or living with a partner. Compared with unassisted women, a larger proportion of HUD-assisted women reported working in the past year and receiving some college education.

HUD-assisted women reported higher levels of fair or poor health status, with approximately one-fifth self-reporting this outcome (21 percent), compared with approximately 11 percent among unassisted women, although reported levels of serious psychological distress were similar between the two groups. Consistent with previous literature, nearly one-third of HUD-assisted women were current smokers at the time of their health interview (Helms, King, and Ashley, 2017), whereas less than 20 percent of unassisted women were current smokers. Demographic factors may account for differences in current smoking. For example, unassisted women were more likely to be Hispanic, and Hispanics have lower prevalence estimates of current smoking (Arrazola et al., 2023). HUD-assisted women had higher reported levels of public health insurance and lower levels of uninsured status compared with unassisted women. Fewer than 20 percent of HUD-assisted women and 28 percent of unassisted women reported no usual source of care, suggesting an area for future research between having health insurance and access to health care among this population.

The objective of this study has been to demonstrate how the NCHS-HUD linked data can serve as a resource for researchers and policymakers in further examining housing status as an important social determinant of health. The findings are exploratory and raise additional questions about the health of HUD-assisted women that could be addressed with additional research examining education level, work status, and receipt of housing assistance among lower-income, female-led households with children. For example, do HUD-assisted women have higher levels of public health insurance and lower levels of uninsured status because they can more successfully navigate complex social safety net systems due to their education? Does the stability of federal housing assistance programs allow women to pursue additional education? Do lower-income women with higher levels of education possess the skills and knowledge needed to obtain housing assistance?

Limitations

NHIS participants eligible for linkage to HUD records are a self-selected subset of the initial survey participants and may differ from those not eligible for linkage. In this study, 88 percent of sample adults in the 2015–18 NHIS were eligible for linkage. The NHIS sample weights were adjusted using variables that influence the probability of linkage eligibility while maintaining population totals to mitigate linkage-eligibility bias. This weight adjustment is similar to nonresponse adjustment (Aram et al., 2021).

This analytic strategy cannot fully account for unobserved characteristics that distinguish those who apply for and receive housing assistance versus those who do not. Although previous studies using the NHIS-HUD linked data have employed the use of a “pseudo-waitlist” group (Simon et al., 2017) or took into account those who had applied but entered assistance at a later date (Fenelon et al., 2017), focusing on women 18–44 years of age with children prohibited such approaches due to smaller sample sizes in this study. Using that information would have further restricted the sample because waitlist information is unavailable for multifamily housing programs in HUD administrative records. The comparison group that did not link to HUD administrative records likely represents a population eligible for housing assistance but not receiving it because gross annual income is the most significant federal housing assistance eligibility factor, and demand for housing assistance outweighs supply.

Finally, examinations by HUD program category were not possible due to limited sample sizes. Future planned linkages between NHIS and HUD administrative data may allow aggregation of more NHIS years, increasing sample size and allowing for more examinations of selection bias and differences across HUD program categories.

Conclusion

The findings in this article are exploratory. The objective has been to present preliminary estimates and to demonstrate how the NCHS-HUD linked data can serve as a resource for researchers and policymakers in further examining housing status as an important social determinant of health.

Authors

Veronica Helms Garrison is a social science analyst (team lead, data and research) at the U.S. Department of Housing and Urban Development, Office of Policy Development and Research. Jacqueline V. Bachand is a social science analyst at the U.S. Department of Housing and Urban Development, Office of Policy Development and Research. Cindy Zhang is an information technology specialist, Division of Analysis and Epidemiology, National Center for Health Statistics, Centers for Disease Control and Prevention. Christine Cox is a statistical consultant, Division of Analysis and Epidemiology, National Center for Health Statistics, Centers for Disease Control and Prevention. Cordell Golden is the Data Linkage Methodology and Analysis Branch chief, Division of Analysis and Epidemiology, National Center for Health Statistics, Centers for Disease Control and Prevention. Kimberly A. Lochner is the associate director for science, Division of Analysis and Epidemiology, National Center for Health Statistics, Centers for Disease Control and Prevention.

References

Alvarez, Thyria, and Barry L. Steffen. 2021. *Worst Case Housing Needs: 2021 Report to Congress*. Prepared for the U.S. Department of Housing and Urban Development, Office of Policy Development and Research. Washington, DC: Government Printing Office. https://www.huduser.gov/portal/sites/default/files/pdf/Worst-Case-Housing-Needs-2021.pdf.

Aram, Jonathan, Cindy Zhang, Cordell Golden, Carla E. Zelaya, Christine S. Cox, Yeats Ye, and Lisa B. Mirel. 2021. "Assessing Linkage Eligibility Bias in the National Health Interview Survey," *Vital and Health Statistics* 2 (186).

Arrazola, René A., Todd Griffin, Natasha B. Lunsford, Deirdre Kittner, Philip Bammeke, Elizabeth A. Courtney-Long, and Brian S. Armour. 2023. "US Cigarette Smoking Disparities by Race and Ethnicity—Keep Going and Going!" *Preventing Chronic Disease* 20 (220375).

Braveman, Paula, and Laura Gottlieb. 2014. "The Social Determinants of Health: It's Time to Consider the Causes of the Causes," *Public Health Reports* 129 (1, Supplementary Issue 2): 19–31.

Fellegi, Ivan P., and Alan B. Sunter. 1969. "A Theory for Record Linkage," *Journal of the American Statistical Association* 64 (328): 1183–210.

Fenelon, Andrew, Patrick Mayne, Alan E. Simon, Lauren M. Rosen, Veronica Helms, Patricia Lloyd, Jon Sperling, and Barry L. Steffen. 2017. "Housing Assistance Programs and Adult Health in the United States," *American Journal of Public Health* 107 (4): 571–78. https://doi.org/10.2105/AJPH.2016.303649.

Gubits, Daniel, Marybeth Sinn, Michelle Wood, Stephen Bell, Samuel Dastrup, Claudia D. Solari, Scott R. Brown, Debi McInnis, Tom McCall, and Utsav Kattel. 2016. *Family Options Study: 3-Year Impacts of Housing and Services Interventions for Homeless Families*. Prepared for the U.S. Department of Housing and Urban Development, Office of Policy Development and Research. Washington, DC: Government Printing Office. https://www.huduser.gov/portal/sites/default/files/pdf/Family-Options-Study-Full-Report.pdf.

Helms, Veronica E., Brian A. King, and Peter J. Ashley. 2017. "Cigarette Smoking and Adverse Health Outcomes Among Adults Receiving Federal Housing Assistance," *Preventive Medicine* 99: 171–177.

Helms, Veronica E., Jon Sperling, and Barry L. Steffen. 2017. *A Health Picture of HUD-Assisted Adults, 2006–2012: HUD Administrative Data Linked with the National Health Interview Survey*. Prepared for the U.S. Department of Housing and Urban Development, Office of Policy Development and Research. Washington, DC: Government Printing Office. https://housingis.org/sites/default/files/Health-Picture-of-HUD.pdf.

Helms, Veronica E., Barry L. Steffen, Elizabeth C. Rudd, and Jon Sperling. 2018. *A Health Picture of HUD-Assisted Children, 2006–2012: HUD Administrative Data Linked with the National Health Interview Survey*. Prepared for the U.S. Department of Housing and Urban Development, Office of Policy Development and Research. Washington, DC: Government Printing Office. https://www.huduser.gov/portal//portal/sites/default/files/pdf/Health-Picture-of-HUD-Assisted-Children.pdf.

Kessler, Ronald C., Gavin Andrews, Lisa J. Colpe, E. Hiripi, Daniel K. Mroczek, Sharon Lise Normand, Ellen E. Walters, and Alan M. Zaslavsky. 2002. "Short Screening Scales to Monitor Population Prevalences and Trends in Non-specific Psychological Distress," *Psychological Medicine* 32 (6): 959–76.

Krieger, James, and Donna L. Higgins. 2002. "Housing and Health: Time Again for Public Health Action," *American Journal of Public Health* 92 (5): 758–68. https://doi.org/10.2105/AJPH.92.5.758.

National Center for Health Statistics. 2022. *The Linkage of the National Center for Health Statistics (NCHS) Survey Data to U.S. Department of Housing and Urban Development (HUD) Administrative Data: Linkage Methodology and Analytic Considerations*. Prepared for the Centers for Disease Control and Prevention, National Center for Health Statistics, Division of Analysis and Epidemiology. Hyattsville, MD: Government Publishing Office. https://www.cdc.gov/nchs/data/datalinkage/NCHS-HUD-Linked-Data-Methodology-and-Analytic-Considerations.pdf.

———. 2019a. *2018 National Health Interview Survey*. https://www.cdc.gov/nchs/nhis/data-questionnaires-documentation.htm.

———. 2019b. *A Primer on HUD Programs and Associated Administrative Data*. https://www.cdc.gov/nchs/data/datalinkage/primer-on-hud-programs.pdf.

———. 2018. *2017 National Health Interview Survey*. https://www.cdc.gov/nchs/nhis/data-questionnaires-documentation.htm.

———. 2017. *2016 National Health Interview Survey*. https://www.cdc.gov/nchs/nhis/data-questionnaires-documentation.htm.

———. 2016. *2015 National Health Interview Survey*. https://www.cdc.gov/nchs/nhis/data-questionnaires-documentation.htm.

Parker, Jennifer D., Makram Talih, Donald J. Malec, Vladislav Beresovsky, Margaret Carroll, Joe F. Gonzalez, Jr., Brady E. Hamilton, Deborah D. Ingram, Kenneth Kochanek, Frances McCarty, Chris Moriarity, Iris Shimizu, Alexander Strashny, and Brian W. Ward. 2017. "National Center for Health Statistics Data Presentation Standards for Proportions: Data Evaluation and Methods Research," *Vital Health Statistics* 2 (175).

Shinn, Marybeth, Michelle Wood, Stephen Bell, Samuel Dastrup, Claudia D. Solari, Scott R. Brown, Debi McInnis, Tom McCall, and Utsav Kattel. 2016. *Family Options Study: 3-Year Impacts of Housing and Services Interventions for Homeless Families*. Washington, DC: U.S. Department of Housing and Urban Development, Office of Policy Development and Research. https://www.huduser.gov/portal/sites/default/files/pdf/Family-Options-Study-Full-Report.pdf.

Simon, Alan E., Andrew Fenelon, Veronica Helms, Patricia C. Lloyd, and Lauren M. Rossen. 2017. "HUD Housing Assistance Associated with Lower Uninsurance Rates and Unmet Medical Need," *Health Affairs* 36 (6): 1016–23.

Swope, Carolyn B., and Diana Hernández. 2019. "Housing as a Determinant of Health Equity: A Conceptual Model," *Social Science & Medicine* 243 (112571). https://doi.org/10.1016/j.socscimed.2019.112571.

U.S. Census Bureau. n.d. *Poverty Thresholds*. https://www.census.gov/data/tables/time-series/demo/income-poverty/historical-poverty-thresholds.html.

———. 2021. *American Housing Survey (AHS)*. https://www.census.gov/programs-surveys/ahs/data/interactive/ahstablecreator.html?s_areas=00000&s_year=2021&s_tablename=TABLE1&s_bygroup1=1&s_bygroup2=1&s_filtergroup1=1&s_filtergroup2=1.

U.S. Department of Housing and Urban Development (HUD). n.d.a. *FY 2022–2026 HUD Strategic Plan*. https://www.hud.gov/HUD-FY22-26-Strategic-Plan-Focus-Areas.

———. n.d.b. *Mission*. https://www.hud.gov/about/mission.

———. 2022. *2022 December Extract Standardized Across the Public and Indian Housing Information Center and the Tenant Rental Assistance Certification System*. Unpublished dataset, April 2023.

The White House. 2022. *White House Blueprint for Addressing the Maternal Health Crisis*. https://www.whitehouse.gov/wp-content/uploads/2022/06/Maternal-Health-Blueprint.pdf.

World Health Organization. n.d. *Social Determinants of Health*. https://www.who.int/health-topics/social-determinants-of-health#tab=tab_1.

Building a Transformational Data Resource to Support Housing Research: The Wisconsin Experience

Marah Curtis
Kurt Paulsen
Hilary Shager
University of Wisconsin

Abstract

This article describes recent efforts by University of Wisconsin-Madison (UW) researchers to connect federal and local housing program data with the Wisconsin Administrative Data Core (WADC). These connections create an innovative and transformational approach to support housing research that informs policy and program design. Developed and maintained by UW's Institute for Research on Poverty (IRP) in collaboration with Wisconsin state agency partners, WADC links large volumes of standardized, longitudinal administrative data from nearly all Wisconsin social welfare programs (for example, Medicaid, Supplemental Nutrition Assistance Program, Temporary Assistance for Needy Families, child welfare, child support, childcare subsidies, unemployment insurance, and homelessness services), information on incarceration from the Department of Corrections, and children's educational outcomes from the Wisconsin Department of Public Instruction. The data system relies on a file known as the Multi-Sample Person File (MSPF), which contains one observation per individual, with no individual appearing twice. The MSPF can be linked with program participation data files, allowing researchers to group individuals by case or family, supporting integrated analysis of multiple program participation and individual and family outcomes over time. Leveraging recent funding opportunities, and via a data-sharing agreement with the U.S. Department of Housing and Urban Development (HUD), the authors connect federal housing program participation data with WADC. IRP is conducting two proof-of-concept studies analyzing the effects of these programs on adult health and child educational outcomes. IRP also recently incorporated the state's Homeless Management Information System into WADC and is pursuing opportunities to incorporate localized data from the Emergency Rental Assistance program established during the COVID-19 pandemic, as well as publicly available eviction data. Linking such data opens an expansive new research agenda to include the study of multiple public program participation and policy interactions; explore a wide berth of individual, family, and community outcomes; and inform actionable policy and practice recommendations. This article shares insights from UW's experience developing and maintaining agency partnerships, and this valuable data resource, which might be applied in other states, discusses the potential of linked administrative data to advance future interdisciplinary, applied housing research and evidence-based policymaking.

Introduction

Housing is one of the most fundamental needs and is critical to family and individual well-being, yet many people struggle to maintain decent, affordable, and stable housing. Low-income individuals and families face particular challenges as rising housing costs and stagnant incomes make housing unaffordable (JCHS, 2023). These families often seek aid from federal housing assistance programs. However, because housing assistance is not an entitlement, many families who need help securing or maintaining adequate, stable housing do not receive it, and others spend years on waiting lists (Acosta and Gartland, 2021; Ellen, 2020). For those who receive housing benefits, the support can be significant, typically amounting to several thousands of dollars per year, which both stabilizes housing and increases economic resources for other household needs.

Despite great interest in evaluation research to examine the effects of housing assistance on family and individual well-being, high costs and data unavailability have hampered high-quality studies. To identify the effects of housing assistance on well-being, careful empirical evaluation research needs to account for household selection into housing assistance, interaction with other income-qualified social programs, and neighborhood or housing market effects. This research can be undertaken as experiments with random selection, such as in the Family Options Study and Moving to Opportunity (MTO) experiments. Both the Family Options Study and MTO experiments involved random assignment of households to different treatment groups in several U.S. cities. Both demonstrated that housing assistance has significant and positive benefits for family and child well-being (Gubits et al., 2016; Sanbonmatsu et al., 2011).

However, because random assignment experiments in social policy evaluations are both expensive and limited, researchers have more recently focused attention on using quasi-experimental research designs with large administrative datasets. These methods expand the reach of research and evaluation to a larger number of places and policies. A number of recent studies have used U.S. Department of Housing and Urban Development (HUD) administrative data on housing program participants for program evaluation, but they are limited to the data and variables in the HUD data (Ellen, 2020; Ellen et al., 2016, 2021; Schwartz et al., 2019). Ellen et al. (2016) use the geographic locations from housing program administrative data to evaluate the school-attendance-area quality location decisions of voucher holders. Fenelon et al. (2017) combine HUD administrative data with survey data to examine adult health. Schwartz et al. (2019) merge HUD administrative data with New York City public schools' administrative data and test scores. However, as far as the authors know, no current studies integrate administrative data from housing programs with other means-tested social programs.

In an attempt to remedy this dearth of actionable evidence, researchers at the Institute for Research on Poverty (IRP) at the University of Wisconsin-Madison (UW) sought to increase the amount and quality of housing-related data available in the Wisconsin Administrative Data Core (WADC), one of the richest collections of linked administrative data on program participation in the country. This article describes how these researchers facilitated partnerships between academia, government, and practitioners and developed technical solutions to create an innovative and transformational data resource and approach to support housing research that informs policy and program design.

The Wisconsin Administrative Data Core

In 2008, stemming from a series of previous evaluation partnerships between IRP and government entities dating back to the 1980s—and the realization that ongoing integration of state administrative data could facilitate rigorous and actionable research—IRP partnered with the newly formed Wisconsin Department of Children and Families (DCF) to propose a set of integrated data development, analysis, and evaluation activities. The initial project, titled Building an Integrated Data System to Support the Management and Evaluation of Integrated Services for Temporary Assistance for Needy Families (TANF)-Eligible Families, was funded under the Office of Planning, Research, and Evaluation, Administration for Children and Families, U.S. Department of Health and Human Services' funding opportunity—Federal-State Partnerships to Build Capacity in the Use of TANF and Related Administrative Data. The project's ultimate goal was "… to create a data resource to support the integrated analysis of the earnings, income, and multiple program participation trajectories of Wisconsin families participating in TANF and other income and work support programs" (Brown et al., 2020: 2). The expectation was that such a resource could facilitate important contributions to program evaluation and administration, as well as basic research. As IRP's technical report on lessons learned in the development of WADC noted, a number of existing strengths facilitated the successful execution of the project:

> (1) A new administrative structure that brought TANF, child welfare, childcare, and child support administration within a single department, DCF, as of July 2008; (2) substantial prior experience using administrative data for research, program monitoring, and management improvement and high-level commitment to expanding these efforts; and (3) a long-term collaborative relationship between Wisconsin State agencies and researchers at IRP. (Brown et al., 2020: 2)

During the past 15 years, with substantial investment from researchers working through IRP, additional funding opportunities, and expanded partnerships with state and local agencies, these initial efforts have grown into the WADC of today.

The current WADC links a large array of cleaned and standardized administrative data covering decades of program history, including data from nearly all Wisconsin social welfare programs (for example, Medicaid, Supplemental Nutrition Assistance Program (SNAP), TANF, child welfare, child support, childcare subsidies, unemployment insurance (UI), homelessness services), as well as information on incarceration from the Department of Corrections and children's educational outcomes from the Department of Public Instruction (exhibit 1).[1]

[1] Exhibit 1 reflects data available in the most recent version of WADC. WADC is rebuilt annually. The next completed version is expected to be released in the fall of 2023 and will include 2022 data for most sources, plus an expansion of two additional data sources: circuit court criminal records (for example, felony, misdemeanor, and traffic cases) and the universe of UI wages (versus wage records only for those in WADC attached to other programs). The latter expansion will likely significantly increase the overall *N* in the WADC sample to approximately 13 million individuals.

Exhibit 1

What Data Are Available in the Wisconsin Administrative Data Core?

Agency or Program	Data [Years Complete Data Available]
Wisconsin Department of Health Services	• FoodShare (Supplemental Nutrition Assistance Program or food stamps) [1989–2021]. • BadgerCare (Medicaid/State Children's Health Insurance Program) [1990–2021]. • Medicaid Claims and Encounters [2009–21]. • Caretaker Supplement [1998–2021].
Wisconsin Department of Children and Families	• W-2 Wisconsin Works (Temporary Assistance for Needy Families) [1997–2021]. • Wisconsin Shares (childcare subsidies) [1997–2021]. • Aid to Families with Dependent Children [1989–98]. • Child Support [1996–2021]. • Child Welfare and Child Protective Services [2005–21].
Wisconsin Department of Workforce Development	• Quarterly wage records [1988–2021]. • Unemployment insurance benefits [2007–21].
Wisconsin Department of Corrections	• Incarceration in state prison facilities [1990–2021].
Milwaukee County	• Incarceration in Milwaukee County jails [1994–2019].
Circuit Court Records	• Public court records [1998–2021]. • 21-county sample of divorce and paternity cases [1980–2019].
Wisconsin Department of Public Instruction	• 4K–12 public school enrollment and school records [2006–21].
Homeless Management Information System	• Homelessness services [2010–21].

4K–12 = 4-year-old kindergarten through 12th grade.
Source: authors

The data system relies on a file known as the Multi-Sample Person File (MSPF), which contains one observation per individual, with no individual appearing twice. The MSPF can be linked with program participation data files and complementary files, allowing researchers to group individuals by case or various definitions of family and follow them over time. The 2021 MSPF includes information for more than 8.3 million individuals.

The fundamental tasks of creating the MSPF include cleaning and standardizing variables used for linking, then match-merging individuals from all data systems with one another, un-duplicating and linking observations so that only one observation per individual remains in the final version. IRP's data science team (currently composed of eight individuals) creates standardized versions of certain data fields, such as name and place of birth, and eliminates unusable observations. Standardization may include eliminating illegal characters from data fields, changing mixed-case character data to uppercase, changing character data to numeric data whenever possible, resolving inconsistent or conflicting information, parsing text into separate fields, and identifying or collapsing missing data codes. Unusable data to be eliminated might include observations with no identifying information, observations clearly used as test cases for training purposes, and extraneous case members (Brown et al., 2020).

Via close collaboration with data providers and learning over time, data scientists also navigate variation in quality of elements across systems. For example, coding social security numbers (SSNs) in UI data is highly accurate because the program relies on records of employment by SSN to distribute benefits. However, this data element may be less accurately coded in programs that do not rely on it for benefit distribution. Data scientists also consider whether data fields are electronically loaded or manually entered, administrative procedural changes over time, and differences in data entry patterns by county. Open and regular communication with data providers and attention to detail on the part of data scientists are required because agencies rarely clearly and accessibly document such changes and procedures (Brown et al., 2020).

Once data are cleaned and standardized, IRP primarily uses individual identifying characteristics and demographics—preferably those that are commonly recorded, relatively uniquely identifying, and unchanging—to accomplish the match-merge. Examples of such identifiers include SSNs and SSN verification codes; personal identification numbers (PIN) cross-loaded from one data system to another; and names, sex, dates of birth and death, place of birth, and parent identifiers (first name, date of birth, and SSN of both mother and father). Race and ethnicity may be used indirectly to refine name standardization processes.

Some linkages are more challenging than others. For example, young children may be more difficult to match simply because their data have not been in WADC for long, and their information will appear in fewer data systems with fewer cross-checking opportunities. Special considerations are also considered for linking data from Hispanic, Hmong, North African, and Middle Eastern populations due to commonality of full names and common approximations of birth dates recorded in U.S. data systems. In addition, not all systems include the same linking variables.[2] Again, IRP data scientists rely on their experience working with the data and observing patterns over time to help improve the ability to adjust the matching algorithm to such challenges and anomalies.

IRP rebuilds the data source annually, thus allowing for continual improvements in IRP's ability to match individuals across systems, drawing on any new data collected. On construction of the MSPF, the data science team is then able to develop additional research files, including (1) a reduced set of demographic variables (removing uniquely identifying personal information for purposes of individual anonymity), with the addition of indicators for which data source provided information, and a constructed (that is, masked) unique PIN; (2) a set of files that allow aggregation of individuals into administrative cases (as programs define) or family units; and (3) participation files that provide information on program participation during a specified period. The constructed unique PIN allows linkages between all aggregation and participation files and the MSPF.

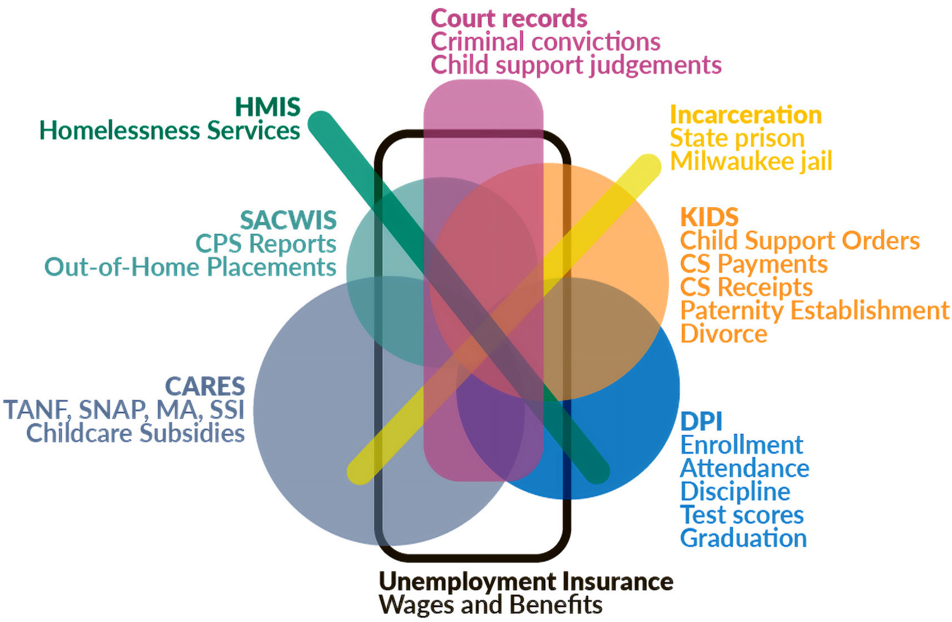
As exhibit 2 illustrates, this data model allows analysis of the full universe of cases or individuals from one source of administrative data, including both those who are included in other systems and those who are not. Researchers can also easily focus on subsets of individuals who participate in constellations of programs or services. Thus, this full merging of multiple administrative data sources, independent of the formulation of specific research questions, significantly broadens the set of questions that can be addressed.

---

[2] The same linking algorithms are applied across all data sources except for UI benefits and wages. UI data do not contain dates of birth, and longer names are truncated. Thus, data scientists are not able to check for a match on date of birth or count on the completeness of names. Several data sources (for example, Department of Public Instruction, Department of Corrections, Milwaukee Jail, and Wisconsin Circuit Court Access Program) do not contain SSNs, and SSNs in other data systems, such as the Statewide Automated Child Welfare Information System and Computer Reporting Network /Income Maintenance Program (CRN/IMP), have quality concerns. The lack of SSNs in these sources decreases certainty in a match, although an exact match on name and date of birth enables a high degree of certainty.

Exhibit 2

Wisconsin Administrative Data Core, Data Model



CARES = Client Assistance for Re-employment and Economic Support. CPS = Child Protective Services. DPI = Department of Public Instruction. HMIS = Homeless Management Information System. KIDS = Kids Information Data System. MA = Medicaid. SACWIS = Statewide Automated Child Welfare Information System. SNAP = Supplemental Nutrition Assistance Program. SSI = Supplemental Security Income. TANF = Temporary Assistance for Needy Families.

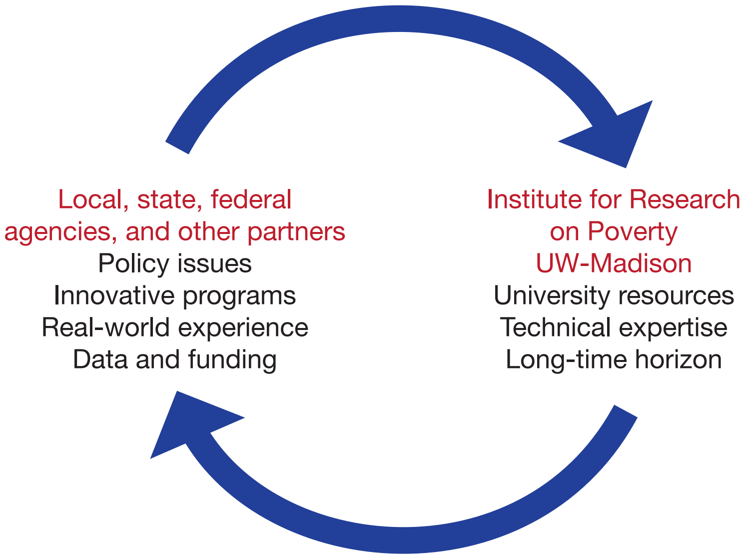
Source: Authors

Putting Administrative Data to Use

Although the technical challenges of developing a data resource like WADC are substantial, perhaps even more important are the development and navigation of the partnerships between data-providing agencies and research organizations. In Wisconsin, it has meant fostering a “logic of collaboration” that supports both policy development and academic research and recognizes that such partnerships require infrastructure and resources to support sustained engagement (exhibit 3). For example, to create the data system, the State of Wisconsin agencies provide data, IRP provides specialized programming and technical support staff and specialized hard- and software, and UW provides funding from federal grants and contracts (often in response to joint UW and agency proposals), state grants and contracts, and foundations. With this engagement, partnerships can develop trust and a shared understanding of useful and interesting questions, appropriate methods, and satisfactory answers. Although the authors found Wisconsin to be somewhat unique in its level of investment in and scope of such partnerships, they believe the principles guiding such work could be applied in other states and with other universities.

Exhibit 3

Institute for Research on Poverty's Logic of Collaboration



Source: authors

Although all projects using this wealth of data must have core policy or practice issues and questions as a basis for research and be approved by contributing state agencies, UW works in partnership with the agencies based on a philosophy of "yours, mine, and ours." Specifically, partners may identify questions and projects that are of primary interest to an agency but require the expertise of IRP researchers to answer (yours); projects that may result in generalizable learning but are of primary interest to IRP researchers (mine); and, optimally, a set of questions and projects that are of high interest to both researchers and agencies, relevant to each other's missions, and responsive to the incentive systems of both academia and public policy (ours). IRP staff work with researchers to ensure the viability and appropriateness of projects before submitting them to agencies for approval and work within each data-providing agency's unique data-governance structure and data-sharing process.

WADC, which provides a unique resource for agencies that cannot otherwise link and analyze across systems, sustains state support of data access and funding, and given its uniqueness as a resource for research that cannot otherwise be completed, sustains commitment and interest of academic researchers. Constraints also shape the partnerships. Importantly, state agencies are not permitted to provide data access for research not relevant to the agency's mission, so researchers need to accept these limitations and understand and explain the utility of their research. From the university perspective, academic freedom demands that research results be made public. Funders do not permit IRP faculty and staff to submit research for "clearance." However, prior to publication or presentation, a 30-day review and opportunity for feedback, which authors may address, is built into all data-sharing agreements. Also, because WADC relies partially on fuzzy matching (that is, approximate matching that may consider transposed numbers, nicknames, and so on) and is a research dataset not meant for operational use (for example, case management or detection of fraud), IRP is not allowed to return matched data to agencies for legal reasons. Therefore, agencies must value and accept independence.

To identify questions of interest and support mutually beneficial research, IRP staff and faculty affiliates engage regularly with agency leadership and staff. For example, IRP supports a state WADC Advisory Council composed of lead agency data stewards and UW faculty and staff; participates in regular briefings and opportunities to check in with individual agencies; has developed sustained research agreements for programs (for example, child support, child welfare, Medicaid); fields contracts for specific programs, projects, pilots, and funding opportunities; participates in "Learning Exchanges" with agency leadership and staff; and provides and receives *ad hoc* technical assistance (some state-funded, some not). Although significant technical, financial, and relationship challenges are involved, investing in researcher-practitioner partnerships and developing linked data systems can lead to a culture of evidence-informed policymaking that is both beneficial to society and rewarding for researchers.

Early Housing-Related Research With WADC and Limitations

Despite being one of the richest collections of linked administrative data in the country for most of its existence, WADC included almost no data about housing. No single state agency collects housing program or homelessness data, and other benefit programs do not require, prioritize, or verify such information. In limited cases, address data for individuals in public assistance programs (Client Assistance for Re-employment and Economic Support [CARES]), child support cases (Kids Information Data System [KIDS]), and child welfare cases (the Statewide Automated Child Welfare Information System [SACWIS]) could be used for neighborhood analysis but without specific housing data. In addition, CARES contains an indicator as to whether someone reported receiving housing assistance, drawn from questionnaires administered to households during initial enrollment into or renewal of SNAP or TANF benefits.

For example, this indicator allowed for some research suggesting small positive effects of a family's initial receipt of housing assistance on students' subsequent academic achievement. The study by Carlson et al. (2019) used WADC to develop two analytic comparisons. First, the authors constructed a "future recipient" comparison group that measured educational outcomes for children living in households that reported receiving housing assistance compared with data for the same children up to 4 pretreatment years. The second analysis compared outcomes for children living in households that reported receiving housing assistance with those of children living in low-income families who did not receive housing assistance but received other means-tested benefits, such as SNAP, TANF, or Medicaid. Results suggested modest positive associations between reported housing assistance receipt and math achievement but not reading achievement (measured using state standardized test scores). Positive findings were concentrated among African-American students and more prominent for students whose families received vouchers versus public housing assistance. Obvious limitations of the study include issues related to reliance on self-reporting of housing assistance, a limited sample (that is, including only information for families receiving SNAP or TANF, or both), and potentially incomplete information about the type of housing assistance received (for example, the data include no record of whether individuals are living in publicly subsidized but privately owned housing units, such as via the Low-Income Housing Tax Credit [LIHTC] program).

Measures of housing assistance that indicate the specific federal program (Housing Choice Voucher, public housing, and so on) and the length of time in a program are imperative to understanding population outcomes and interactions with other benefit programs. Relatedly, tracking down residential moves is a core indicator of housing stability associated with adult and child well-being outcomes. Generally, survey data from longitudinal studies are used to investigate mobility, although survey data may underestimate the most mobile households because they are more likely to be lost due to attrition (Curtis and Warren, 2015). Administrative data are more likely to be able to observe the most vulnerable families if they continue to receive any public benefits. For example, a study using WADC data to examine the regular receipt of child support income on housing outcomes used administrative address data to define mobility and found that more regular child support receipt, holding the overall amount constant, was negatively associated with any moves or multiple moves in a year (Curtis and Warren, 2015). The state keeps track of custodial parents' addresses and updates them when a parent moves or a child support order changes. Researchers could observe ZIP Code changes, although such changes may underestimate short-distance moves.

Despite such limitations, these few studies point to the potential analytic power of improving the amount and quality of housing measures in WADC data for policy-relevant analyses. Specifically, standardized data that allow for the observation of moves, the timing of those moves, participation in public benefits, release from corrections facilities, participation in child protective services, and accurate measures of participation in housing assistance programs open the possibility of a plethora of questions about how housing (stability, housing assistance, location) relates to well-being outcomes. When combined with appropriate measures of the housing context, these administrative data can be used to evaluate neighborhood- or community-level outcomes and to simulate alternative policy scenarios.

Building a Transformational Data Resource to Support Housing Research

To expand the types of applied housing research questions that integrating housing data with other administrative data could address, a group of IRP researchers acquired competitive funding from the UW Office of the Vice Chancellor of Research and Graduate Education and the Wisconsin Alumni Research Foundation. The project, titled "Building a Transformational Data Resource to Support Housing Research," sought to (1) dramatically increase the quality and amount of housing-related data available to researchers via WADC, (2) complete two proof-of-concept research projects, examining links between receipt of housing assistance and health outcomes and investigating links between housing stability and school success, and (3) develop infrastructure (that is, data sharing and expertise) to facilitate pursuit of extramural funding to support further research in these areas. The overarching goal is to generate knowledge that can evaluate and improve public policy affecting the lives of low-income families.

Two proof-of-concept papers are in progress, highlighting the value of merging WADC data with HUD Office of Policy Development and Research's (PD&R) restricted household-level longitudinal tenant data, available through a data license agreement with HUD for these projects. HUD's administrative data cover all Wisconsin participants in public housing, Housing Choice Voucher, and multifamily housing programs. Thus, with WADC, researchers can match households that receive rental assistance under any HUD program with participation in programs such as SNAP or Medicaid. The authors examine households that received any HUD rental assistance from 2003 to 2020, several years before WADC began collecting Department of Public Instruction data used for educational outcomes, through the most recent HUD data available at the time of application. The first paper examines how housing assistance is associated with adult health outcomes in terms of diabetes management and diabetes-related emergency department visits. The sample includes all households that participated in Medicaid linked with those who received housing assistance. Thus, the authors compare those who did not receive housing assistance with those who did among the Medicaid population. The authors hypothesize that housing assistance receipt can improve diabetes management and reduce emergency department use for diabetes-related complications. Participation in HUD's rental assistance programs can be associated with housing stability and increased disposable income (because of reduced rental payments). The authors hypothesize that both the housing stability and income effects should improve health and disease management, leading to better health outcomes for tenants and systems savings through reduced emergency department visits. Stable housing reduces daily living stress, enables medication and disease management, and may also facilitate consistent healthcare use with regular providers. WADC includes Medicaid Claims data for measures of health management and diabetes-related hospital admissions and allows for repeated measures capturing income and program participation in a longitudinal framework, enabling more rigorous analytic approaches to estimate the effect of housing assistance on adult health outcomes with limited bias. Because income qualifications for HUD housing assistance and Medicaid are similar, nearly all HUD-assisted households are Medicaid eligible. However, the health outcomes data are limited to those enrolled in Medicaid and would, therefore, miss HUD-assisted households not enrolled in Medicaid.

The second paper examines how housing stability is associated with children's success in school. Housing instability may operate in several key ways to harm children's learning and school performance. First, frequent moves might necessitate frequent changes in children's schools, which produce disruptions to the continuity of curriculum and content that children are expected to learn. Moves also disrupt attachments to teachers and peers who provide a sense of belonging and foster school success. Second, housing instability is stressful for parents, and the immediate need for housing may divert attention away from engaging in the types of parent-child interactions that support children's learning both in the home and school environments. For example, parents with unstable housing may spend less time reading to and assisting their children with homework at home and be less likely to attend school events or meet with teachers compared with parents with stable housing. As in the first paper, IRP linked HUD housing assistance data to records in WADC and benefits from the ability to construct a well-matched comparison group by selecting children in the same school district using inverse probability treatment weighting based on observed household information and benefit receipt available in WADC data. WADC enables researchers to examine not just test scores but also other important dimensions of children's schooling experiences, such as grade retention, special education placement, attendance, and graduation, which can help answer nuanced mechanism and policy questions. In addition, the ability to investigate questions using a longitudinal framework, observing children's outcomes several years after the start of public housing assistance benefits, allows for more rigorous analytic approaches and the observation of long-term outcomes.

Future Directions and Research Synergies

As the sole federally funded National Research Center on Poverty and Economic Mobility,[3] IRP's research focuses on policies that affect the lives of those households that HUD's rental programs also potentially assist and are most at risk of housing instability and extreme housing cost burdens. Because housing is foundational to a family's health, well-being, and economic mobility, recent interest in housing research has increased substantially. The tremendous federal investments in stable and safe housing and safety net programs during the COVID-19 pandemic recognized the foundational role of housing. Significant increases in rents, cost burdens, and homelessness in 2022 and 2023 brought the housing crisis front and center. Thus, researchers face an unprecedented opportunity to fill the substantial gap in knowledge regarding the effectiveness of housing assistance programs, their interaction with other social welfare programs, and their effects on health and well-being.

As described previously, WADC has benefited from partnerships with state agencies that oversee specific public programs and, therefore, are able to share all program data. The challenge with capturing housing data is that no state-level equivalent to HUD exists, and data about homelessness, housing assistance programs, housing quality, and so on are spread across multiple state agencies, local public housing authorities, and nonprofit organizations—which requires investing in relationship development with multiple agencies. Long-term partnerships are vital. Data acquisition and merging are primary steps. However, researcher engagement and skill with the data and relationships grow over time.

At the interagency level in Wisconsin, bipartisan legislation in 2017 created the Interagency Council on Homelessness, modeled on the successful U.S. Interagency Council on Homelessness within the federal government. In addition to participation from the state's Housing Finance Authority (Wisconsin Housing and Economic Development Authority) and the state's Community Development Block Grant (CDBG) and HOME Investment Partnerships Program allocating agency (Wisconsin Department of Administration [DOA]), participation includes the U.S. Department of Veterans Affairs, DCF, Department of Public Instruction, Department of Health Services, and Department of Corrections in addition to all the Continuums of Care (Dane, Milwaukee, and Racine Counties and Balance of State). IRP staff and faculty have spoken about housing and homelessness data and research issues with the Interagency Council, conducted learning exchanges within its existing state agency relationships, and regularly consulted with agencies on housing-related issues. These researcher relationships existed beforehand, which allowed for a more formal engagement with support from the UW Office of the Vice Chancellor of Research and Graduate Education and the Wisconsin Alumni Research Foundation to build out housing data. The role of institutional support in fostering research collaborations is very important to acknowledge. IRP

[3] IRP is currently engaged in a 5-year (2021–26) cooperative agreement with the Office of the Assistant Secretary for Planning and Evaluation in the U.S. Department of Health and Human Services to serve as the sole federally funded National Research Center on Poverty and Economic Mobility in the United States.

continues to formalize these relationships, and the following outlines two emerging housing data sources the authors are working to use and integrate into WADC.

First, IRP recently integrated the state's Homeless Management Information System (HMIS) data via a data-sharing relationship with the Institute for Community Alliances, a nonprofit organization providing HMIS training and support for multiple states, including Wisconsin. This new data partnership allows for research on the effectiveness of rapid rehousing or permanently supportive housing on client outcomes in health or employment and could also examine the interaction of housing supports with other programs, such as Medicaid or SNAP. These data could be further integrated with federal data to examine whether households participating in HUD rental assistance programs are less likely to use homeless services. These first-order questions offer needed information about how housing policies and programs interact with other systems and affect well-being. Another use of these data includes sets of questions examining trajectories of vulnerable populations over time, the role of state and federal housing, and other social welfare programs on transitions to community from correctional facilities or child welfare.

The second type of data to integrate is Emergency Rental Assistance (ERA) data. Due to the COVID-19 emergency and fears that housing instability and mass evictions would accelerate viral transmission, the federal government imposed an eviction moratorium for nonpayment of rent and distributed $45 billion in ERA in addition to increased investments through the Coronavirus Aid, Relief, and Economic Security Act for CDBG, HOME, Emergency Solutions Grants program, project-based rental assistance, tenant-based rental assistance, and public housing.

In contrast to HUD's Section 8 platform assistance programs, ERA assistance was not administered through public housing authorities or HUD's Office of Multifamily contracts, payment standards and Fair Market Rents were not in effect, and property owners were not required to submit to Real Estate Assessment Center inspections. ERA was designed to keep tenants in their current units, and emergency rental assistance payments did not require tenants to pay 30 percent of income in rent.

The ERA program is likely the most substantial, limited-time, new rental assistance program ever. For these reasons, careful research on how funds were distributed and the effects of ERA programs on housing markets, housing stability, and household outcomes can serve to inform future policy approaches to keep renting families safely and stably housed.

Because so many factors were changing simultaneously, larger research studies with adequate power and variables to control for household and housing market-level variation are necessary to disentangle the direct and indirect effects of ERA programs. The existing WADC institutional structures, relationships, and "know-how" mean such efforts are well positioned to bring in local-level ERA data to understand how these housing funds affected well-being and provide evidence-based research for housing policy development.

In addition to integrating HMIS and ERA data into WADC, future possibilities could include tenant household data from LIHTC units (such as from HUD form 52697), Low Income Home Energy Assistance Program recipient information from Wisconsin DOA, and eviction filings and court data from the Wisconsin Circuit Court Access Program. Each of these data sources has unique challenges and data restrictions, and the authors have not yet negotiated data use agreements with the relevant state agencies. However, the primary reason to acquire housing data and merge it with WADC is because, without accounting for the housing environment, the authors argue that social scientists, policymakers, and community planners have an incomplete understanding of the lives, well-being, and capacities of the communities they serve. Home is the essential living space for life activities and deserves rigorous focus.

Ongoing Challenges and Opportunities

Although WADC has the potential to be a transformational resource for housing research, significant challenges to using it effectively remain. Each new housing data source the authors integrate comes with a significant learning curve as programmers, primary researchers, and students learn how to work with the new information. In addition, significant investments in developing trusting relationships with data-providing agencies remain key to maintaining functional researcher-practitioner partnerships (above and beyond the technical "know how" required to maintain the data resource). It is often a smart strategy to seek mutually agreed on, important policy-relevant questions that can begin the process of using the data. Partnerships are vital because additional funding must be identified and secured for analyses to proceed. The Institute for Research on Poverty provides the mechanism to keep all partners engaged—sometimes during very long periods—as appropriate funding mechanisms, scholars, and questions are negotiated collaboratively.

To access WADC, current data-sharing agreements require collaboration with a University of Wisconsin researcher.[4] The authors envision potential partnerships with researchers from other states or partnering with researchers interested in using WADC data across at least three policy research categories. First, for poverty and social welfare researchers, the authors hope to demonstrate that incorporating housing data and investigating interactions with housing programs is necessary to study and affect human well-being. Second, the medical and healthcare fields show tremendous interest in understanding housing as one of the social determinants of health. The authors hope to demonstrate methods of integrating housing assistance data with program participation data such as Medicaid. Third, the authors imagine that the network of housing researchers will begin to appreciate why data linkages to other programs, such as SNAP or UI, are needed to understand and improve the multifaceted program and policy environment that economically vulnerable families experience. The authors are excited that HUD PD&R has made available restricted tenant household data for integration with other administrative data, subject to data-use agreements and privacy protections. The authors hope to encourage researchers across a range of disciplines and policy domains to consider using HUD's data resources and develop their own data core models with state agencies. The authors are also hopeful that cross-state work can allow productive collaborations to answer pressing questions about housing and the other policies and programs that support our populations, with cities, states, and localities as policy laboratories.

[4] See IRP's WADC webpage for more information about partnering with UW researchers to use WADC data https://www.irp.wisc.edu/wadc/.

Acknowledgments

The authors gratefully acknowledge the financial support for this research that the University of Wisconsin-Madison Office of the Vice Chancellor for Research and Graduate Education provided with funding from the Wisconsin Alumni Research Foundation. The authors also thank Institute for Research on Poverty (IRP) data scientist Katie Thornton for her helpful answers to questions about the construction of WADC and the IRP editors and two referees for their helpful comments.

Authors

Marah A. Curtis is a Vilas Distinguished Achievement Professor of Social Work at the University of Wisconsin-Madison (UW). Kurt Paulsen is a professor of Urban Planning in the Department of Landscape Architecture at UW. Hilary Shager is the associate director for Programs and Management at UW's Institute for Research on Poverty.

References

Acosta, Sonya, and Erik Gartland. 2021. "Families Wait Years for Housing Vouchers Due to Inadequate Funding: Expanding Program Would Reduce Hardship, Improve Equity." Center on Budget and Policy Priorities. https://www.cbpp.org/sites/default/files/7-22-21hous.pdf.

Brown, Patricia R., Katie Thornton, Dan Ross, Jane A. Smith, and Lynn Wimer. 2020. *Technical Report on Lessons Learned in the Development of the Institute for Research on Poverty's Wisconsin Administrative Data Core*. Madison, WI: University of Wisconsin, Institute for Research on Poverty. https://www.irp.wisc.edu/wp/wp-content/uploads/2020/08/TechnicalReport_DataCoreLessons2020.pdf.

Carlson, Deven, Hannah Miller, Robert Haveman, Sohyun Kang, Alex Schmidt, and Barbara Wolfe. 2019. "The Effect of Housing Assistance on Student Achievement: Evidence from Wisconsin," *Journal of Housing Economics* 44: 61–73. https://doi.org/10.1016/j.jhe.2019.01.002.

Curtis, Marah A., and Emily J. Warren. 2015. "Child Support Receipt, Mobility, and Housing Quality," *Housing Studies* 26 (5): 747–765. https://doi.org/10.1080/02673037.2015.1121212.

Ellen, Ingrid Gould. 2020. "What Do We Know About Housing Choice Vouchers?" *Regional Science and Urban Economics* 80: 103380.

Ellen, Ingrid Gould, Keren Mertens Horn, and Amy Ellen Schwartz. 2016. "Why Don't Housing Choice Voucher Recipients Live Near Better Schools? Insights From Big Data," *Journal of Policy Analysis and Management* 35 (4): 884–905.

Ellen, Ingrid Gould, Katherine O'Regan, and Sarah Strochak. 2021. *Using HUD Administrative Data to Estimate Success Rates and Search Durations for New Voucher Recipients*. Washington, DC: U.S. Department of Housing and Urban Development, Office of Policy Development and Research.

Fenelon, Andrew, Patrick Mayne, Alan E. Simon, Laren M. Rosen, Veronica Helms, Patricia Lloyd, Jon Sperling, and Barry L. Steffen. 2017. "Housing Assistance Programs and Adult Health in the United States," *American Journal of Public Health* 107 (4): 571–578.

Gubits, Daniel, Marybeth Shinn, Michelle Wood, Stephen Bell, Samuel Dastrup, Claudia D. Solari, Scott R. Brown, Debi McInnis, Tom McCall, and Utsav Kattel. 2016. *Family Options Study: 3-Year Impacts of Housing and Services Interventions for Homeless Families*. Washington, DC: U.S. Department of Housing and Urban Development, Office of Policy Development and Research. https://www.huduser.gov/portal/sites/default/files/pdf/Family-Options-Study-Full-Report.pdf.

Joint Center for Housing Studies (JCHS). 2023. *The State of the Nation's Housing 2023*. Cambridge, MA: Harvard University, JCHS. https://www.jchs.harvard.edu/sites/default/files/reports/files/Harvard_JCHS_The_State_of_the_Nations_Housing_2023.pdf.

Sanbonmatsu, Lisa, Jens Ludwig, Lawrence F. Katz, Lisa A. Gennetian, Greg J. Duncan, Ronald C. Kessler, Emma Adam, Thomas W. McDade, and Stacy Tessler Lindau. 2011. *Moving to Opportunity for Fair Housing Demonstration Program: Final Impacts Evaluation*. Washington, DC: U.S. Department of Housing and Urban Development, Office of Policy Development and Research. https://www.huduser.gov/publications/pdf/mtofhd_fullreport_v2.pdf.

Schwartz, Amy Ellen, Keren Mertens Horn, Ingrid Gould Ellen, and Sarah. A. Cordes. 2019. "Do Housing Choice Vouchers Improve Academic Performance? Evidence From New York City," *Journal of Policy Analysis and Management* 39 (1): 131–158.

Promoting Affordable Housing in Well-Resourced Neighborhoods: A Regional Approach to Assessing Neighborhood Resources in New York State

Pooya Ghorbani
Courtney Wolf
Ben Wetzler
Simon McDonnell
Bobbetta Davis
Parker Pence
New York State Homes and Community Renewal

Disclaimer: The authors are employees of New York State Homes and Community Renewal (HCR). However, any analyses, conclusions, and views expressed in this article should not be considered the views or positions of HCR and/or the state of New York.

Abstract

Studies have extensively explored the neighborhood effects on low-income families' welfare and the future outcomes of their children. These studies have motivated public policy at various levels to encourage "moving to opportunity"—improving access to affordable housing in neighborhoods with better resources. This study constructs a comprehensive Neighborhood Resource Index score for assessing the resource levels of neighborhoods in New York State. This index is based entirely on publicly available data and accounts for variations in the density and urban nature of different regions across the state. The article uses this index score to explore the placement of affordable housing built with support from Low-Income Housing Tax Credits. It demonstrates the applicability of the index in improving access to affordable housing in well-resourced areas. The findings indicate that although units have been disproportionately built in lower-score neighborhoods, policies and initiatives, such as those employed by New York State, can offer strategies for increasing the amount of affordable housing in neighborhoods with higher resource scores.

Introduction

An extensive body of scholarship documents the advantages of living in better-resourced neighborhoods for low-income families. The significance of neighborhood characteristics has motivated a number of federal initiatives, such as Moving to Opportunity by the U.S. Department of Housing and Urban Development (HUD).[1] This study introduces a comprehensive methodology for quantifying neighborhood resources. It then demonstrates the practical applicability of this methodology by documenting the distribution of newly built units—specifically those that used Low-Income Housing Tax Credits (LIHTC)—in well-resourced neighborhoods across New York State.

The Low-Income Housing Tax Credit is the largest program to promote the development and preservation of affordable housing in the United States. It encourages the development of rental units for low-income families by providing tax credits to private-sector builders. Nationwide, the program has helped produce more than 3.5 million affordable units since its inception in 1986 (HUD, 2023). To analyze the placement of new affordable homes in well-resourced areas, this study takes advantage of administrative LIHTC data made available to the authors through the New York State Homes and Community Renewal (HCR; New York State's affordable housing agency). These data include the universe of properties that received tax credits administered by New York State since its beginning in the late 1980s.[2] To construct an index score of neighborhood resource levels, the study uses data from the U.S. Census Bureau and other sources at the census tract level. This article builds on HCR's prior efforts to identify well-resourced areas based on poverty rates and educational outcomes. It expands that measure by constructing a composite index score for each neighborhood that is an aggregate of multiple neighborhood characteristics. The metrics for neighborhood resource level and the structure of the indexed score presented in this article are generally modeled after the Child Opportunity Index (Acevedo-Garcia et al., 2014). To account for regional variations in urban settings and population densities across the state, the Neighborhood Resource Index is constructed separately for four distinct regions. The study overlays a sample of state-administered LIHTC properties on neighborhoods and analyzes the placement of those properties in high- versus low-resourced neighborhoods.

This article proposes an index of neighborhood resource levels that employs a diverse set of indicators of local economic and other conditions, thereby alleviating potential inaccuracies that may arise from relying merely on official poverty rates in measuring neighborhood resourcefulness. This index score relies entirely on publicly available data and, therefore, can be updated annually. This feature is especially important for devising new policies and updating the existing ones according to the most recent data available.

The article's analysis of LIHTC developments indicates that units have been disproportionately built in a relatively small concentration of neighborhoods that have lower resource scores, although analysis of recent data suggests an expansion of the areas in which LIHTC units are built, including more well-resourced areas. This distribution in LIHTC units is not unique to New York and exists

[1] See https://www.hud.gov/programdescription/mto.

[2] A part of New York State's tax credit allocation authority is granted to the city of New York. Both HCR and New York City's housing agencies finance LIHTC projects within New York City through separate programs. This study only looks at LIHTC projects financed by the State.

across the country. New York has been proactive in creating avenues of access to better-resourced neighborhoods—and equity in access to resources more generally—through the LIHTC program and other complementary initiatives.

The article proceeds as follows: Review of Theoretical and Empirical Literature reviews three areas of scholarship that are related to moving to well-resourced areas; Prior Work Done by New York State summarizes HCR's recent efforts to facilitate more units in higher-resourced areas; Data and Methodology introduces the data sources and methodology for generating neighborhood resource scores; Discussion of Findings presents the analyses and findings; and Conclusions summarizes the findings and concludes with some policy recommendations.

Review of Theoretical and Empirical Literature

This article's exploration of affordable housing in highly resourced areas encompasses three domains of scholarship: (1) advantages of living in well-resourced neighborhoods for children of lower-income families; (2) evaluation of LIHTC as a program; and (3) the role of local zoning codes in potentially limiting multifamily developments. This section provides a brief review of the literature in each of those areas.

Place-Based Attributes and Outcomes for Children

The economics research documenting the impacts of locational attributes on children's future outcomes is immense. Pioneering theories dating back to the 1960s explored the extent to which community factors influence individual socioeconomic success (see Jencks and Mayer, 1990, for a thorough review of early studies).[3] Social scientists have theorized that living in a well-resourced neighborhood results in better future outcomes through such mechanisms as peer and adult influences. Over time, and with the spread of computer software, the empirical work has been extended to test these theories in different settings and uses a diverse array of indicators for measuring both well-resourced neighborhoods and future outcomes.

Neighborhoods that provide better opportunities for children have been characterized by a comprehensive list of economic and social attributes. The most frequently used factors in research studies include crime and poverty rates; access to public transportation (de Souza Briggs, Popkin, and Goering, 2010); access to high-performing schools (McClure and Schwartz, 2021); local employment rates, demographics, and socioeconomic status (Ellen, Horn, and Kuai, 2018; Lens, 2014); and local income diversity (Chetty et al., 2022). Similarly, researchers have used economic and social attributes to measure future outcomes. Economic indicators include, but are not limited to, job earnings, upward mobility (Chetty and Hendren, 2015), and college attendance (Chetty e. al., 2011; 2018). Social factors include family structure (Chetty et al., 2016), teenage pregnancy, and incarceration (Pollakowski et al., 2022).

Two prominent examples of more recent scholarship motivate this article's definition of well-resourced areas in this article: *The Opportunity Atlas* by Raj Chetty and colleagues (2018) and the

[3] Generally, Jencks and Mayer (1990) categorized the studies that address such influences into those that study the advantages, disadvantages, and irrelevance of living near advantaged neighbors.

Child Opportunity Index developed by Acevedo-Garcia and colleagues (2014). Chetty and colleagues follow a nationally representative sample of over 20 million children born between 1978 and 1983, and they provide estimates for earnings and other outcomes based on the characteristics of the census tracts in which the children grew up—characteristics such as household income, poverty rate, racial composition, and population density. Similarly, the *Child Opportunity Index* ranks census tracts based on the "pathways, through which neighborhood environments influence child development" (Acevedo-Garcia et al., 2014; 7). These pathways include indicators such as high school graduation rates, access to green space, commute duration, and public assistance rate. Both studies acknowledge the role of neighborhood quality in child outcomes and assess that quality using a number of place-based attributes, which is generally the approach in this article, too.

Another strand of the scholarship on the role of neighborhoods in future outcomes addresses the barriers to neighborhood choice and explores ways to overcome those barriers. Bergman and colleagues (2019) found that the provision of information, alongside ongoing assistance and counseling services, significantly improves households' likelihood of moving to well-resourced areas. Other studies (e.g., Godinez-Puig, Garriga, and Freemark, 2023) show that a lack of affordable housing options in high-opportunity areas explains the lower rates of low-income households moving to such neighborhoods. This latter literature, to which this article primarily contributes, raises another question: Why is affordable housing scarce in areas with greater resources?

Low-Income Housing Tax Credit and Its Challenges

Numerous programs, tax breaks, and funding sources support the provision of affordable housing at local, state, and federal levels.[4] The largest subsidy for the development or preservation of low-income rental housing is the LIHTC, a provision of the Internal Revenue Code created as part of the Tax Reform Act of 1986. LIHTC has subsidized the development or preservation of over 3.5 million units in 52,000 buildings across the United States through 2021 (HUD, 2023). In New York State, LIHTC subsidies have helped finance the development of at least 304,000 units since inception,[5] which make up roughly 10 percent of all multifamily rental units in the state and surpass the number of public housing units by 56 percent.

Investors in affordable housing development claim LIHTC against their federal income tax liability. The federal government allocates LIHTC funds to state housing agencies, usually a state housing finance agency (HFA), based on state populations.[6] State housing agencies then award credits to individual developments based on a Qualified Allocation Plan (QAP) that sets out the state's eligibility criteria and other priorities. The amount of credits awarded to a project is based primarily on the construction costs (which generally include the cost of development minus land price), location, and the proportion of units set aside for tenants from different income categories, with bonuses given for additional criteria or priorities set forth in the QAP. State QAPs can play

[4] Major subsidies for affordable housing are provided by the federal government and administered by state and local governments in the form of block grants or allocations. Some examples include public housing, rent payment assistance for tenants (Housing Choice Vouchers), or assistance for rural housing.

[5] This figure includes both New York State and the city of New York's LIHTC programs since the inception of the program, including developments that have since left supervision.

[6] Allocations were originally set at \$1.25 per capita, but they were subsequently increased to \$1.75 in 2002 and \$2.20 in 2008, and they were pegged to inflation. As of this writing, state allocations are at \$2.75 per person.

a significant role in shaping the siting patterns of affordable housing (Ellen, Horn, and Kuai, 2018). Whereas much of the eligibility and allocation criteria for LIHTC is set in federal statute, HFAs have some discretion in what types of projects get prioritized. For instance, federal statute encourages developments in areas with high housing costs or large populations below the poverty line by offering additional credits for projects in these areas (defined by statute). State HFAs may choose to offer additional credits or set aside credits to meet specific state housing goals, such as serving special populations or increasing access to affordable housing in well-resourced areas. A smaller version of the credit is available to all properties financed through federally tax-exempt Private Activity Bonds.[7] In the case of New York State, both credit allocations and Private Activity Bond financing are administered by HCR.

Developments must set the rent for at least 20 percent of their units as affordable to low-income tenants to be eligible for LIHTC.[8] Tax credits are awarded over the course of 10 years, and buildings must maintain affordability for at least 15 years after being put in use.[9]

Once the development costs are estimated, real estate investors consider the amount of tax credits, other tax benefits, and possible profits if the property is sold. LIHTC investors are usually large financial institutions that buy tax credits assembled from multiple developments. Assembling the credits is done by for-profit or nonprofit entities called syndicators who charge a portion of the credit for their service fees. Developers who sell the credits to the investors use the proceeds as equity in the development. As the structure of the subsidy implies, a portion of every LIHTC credit goes toward investment returns and syndication fees.

Researchers have identified several shortcomings of LIHTC. Regarding the subsidy's primary goal of expanding affordability, some have argued that LIHTC can be inferior to other forms of subsidy that are tied to the tenant's income, such as housing vouchers or public housing (Eriksen, 2009). Unlike those programs, LIHTC rents are independent of changes in the tenant's income. Therefore, in instances where the Area Median Income grows faster than the tenant's income, rent can become burdensome. Scholars have also argued that the subsidy in its current form does not provide incentives for developers to target the lowest income groups, especially in higher-cost areas (Schwartz, 2010).

LIHTC has also been subjected to criticisms about inefficiency (Clancy, 1990; Stegman, 1991). As briefly mentioned previously, the full amount of every LIHTC dollar given out by the federal government does not reach low-income tenants because investor profits and syndication fees are

[7] The smaller version of the credit associated with Private Activity Bond deals is commonly known as the "4 percent" credit, as opposed to the full "9 percent" credit. These names refer to the proportion of covered project costs that can be claimed each year over the course of a 10-year credit period.

[8] Affordability is defined as rent being at or below 50 percent of the Area Median Income (AMI). Alternatively, 30 percent of units can be set aside for tenants with incomes at 60 percent of AMI.

[9] For new construction and substantial rehabilitation projects, the 10-year credit is calculated based on 70 percent of the present value of a project's qualified costs, which translates to 9 percent annually. Projects that receive other federal subsidies or tax-exempt financing (in addition to LIHTC) are eligible for smaller subsidies based on 30 percent of the present value of their qualified costs, amounting to approximately 4 percent annually. The Revenue Reconciliation Act of 1989 requires properties to maintain their affordability for an additional 15 years after the first one expires.

deducted.[10] As exemplified by the financial crisis of 2008, tax credits may lose their attractiveness for investors in times of economic (or other) shocks (Schwartz, 2010), cutting developers short of sufficient equity for producing affordable housing. Scholars have argued that during such times, LIHTC can be costlier than the units the market would have generated without the subsidies (Eriksen and Lang, 2018). Another risk becomes apparent in high-inflation periods. In such periods, inflation drives up prices of building materials and makes affordable housing development costlier than usual. Because the Federal Reserve usually responds to inflationary risks by increasing interest rates, securing mortgages without other sources of financing becomes more difficult for below-market rental projects (Capps, 2023). This circumstance can burden states with the need for providing additional subsidies or force developers to allocate smaller shares of new units to affordable housing.

Units supported by LIHTC are also found to be challenging to maintain in the long run. Almost any residential building would need physical improvement and upgrading of major systems one or two decades after being put in service, which can be challenging for LIHTC buildings with very low-income tenants. In such cases, the rent income is not large enough to support refinancing the property and paying for the necessary repairs (Schwartz and Melendez, 2008). States have been increasingly providing new tax credits and tax-exempt bonds for these purposes, although this forces them to choose between financing new units and preserving existing ones.

Scholars have also noted an imbalanced spatial distribution of units supported by LIHTC nationally. According to Climaco and colleagues (2009), as of 2006, 35 percent of tax credit units were located in areas with 30 percent or more of their populations below the poverty line, and 44 percent of tax credit units were in tracts with over 50 percent minority populations (those proportions for all rental units pooled together during that same period were 21 and 32 percent, respectively). However, scholars have also argued that there can be an upside to placing LIHTC projects in low-income neighborhoods because LIHTC tenants usually have incomes above the poverty line and can improve average local income levels (Schwartz, 2010).

In summary, the academic literature on LIHTC highlights some of the reasons behind the scarcity of affordable housing in well-resourced areas. One reason may be that the tax credits can be less cost-effective, especially in higher-cost areas or when a larger portion of the credit is deducted from profits and fees. Furthermore, the cost of building new LIHTC units can be influenced by macroeconomic dynamics and other factors that may be beyond states' control. Maintaining existing units may also be difficult without additional subsidies. These factors, however, do not fully explain why low-income units are less likely to be built in areas with greater resources. The following section reviews the literature on another possible explanation: Zoning.

[10] In the subsidy's early years, the price paid by investors for every dollar of LIHTC credit was as low as $0.40. However, as the program became more familiar and was made permanent by Congress in 1993, investors have been willing to pay much more, and the price has even exceeded $1.00 at times, including the period leading to the Great Recession and more recently during 2015–16. Before the COVID-19 pandemic, the equity price of each LIHTC credit was about $0.93, and most recently it is $0.90.

Municipal Land Use Regulations and Low-Income Housing

Zoning regulations in America are defined by shielding residential land use in general, and low-density single-family functions in particular, from other forms of land use (Hirt, 2014).[11] A half-century ago, such protective approaches tended to be justified based on the negative outcomes of metropolitan growth, such as air pollution or traffic congestion (Tolley, 1974). However, many examples of dense urban areas have shown much greater gains from agglomeration efficiencies—including reduced costs of transporting goods and access to service, specialization of human capital, and knowledge spillovers. Ellickson (2022) provides a summary of studies related to agglomeration efficiency. Therefore, the persistence of zoning most likely has to do with more than protecting environmental quality.

After more than a century since their inception, local zoning ordinances still govern what gets built and where it gets built. The endurance of zoning regulations in American cities is primarily due to their function as the protector of private property values, especially of detached single-family homes.[12] Higher property values are favored not only by property owners but also by local governments, for whom property tax revenues are the lifeline in providing basic services and fulfilling political promises (Bassett, 1922; Fischler, 1998). This arrangement was coined by William Fischel (2005) as the "homevoter theory," and it has been studied by various scholars (e.g., Been, Madar, and McDonnell, 2014).

Social science scholars have frequently criticized zoning practices for various reasons. Most prominently, the separation of low-density residential districts has been argued to segregate people by class, race, and gender (Haar and Kayden, 1989; Micklow, 2008) and limit diversity and interaction (Jacobs, 1961). Detached single-family zoning is also shown to cause environmental damage by requiring wasteful infrastructure extensions and imposing higher maintenance costs and longer commute times (Schuetz, 2022).

Also relevant to this article is the economics literature on the effect of zoning ordinances on housing prices. This literature employs classical supply and demand frameworks to explain how zoning restrictions inflate housing prices and make housing unaffordable by putting a cap on supply (Been, Ellen, and O'Regan, 2019; Glaeser and Gyourko, 2002; Quigley and Raphael, 2004; Saiz, 2010). Based on this argument, policymakers have prescribed boosting housing supply in low-density areas as a remedy to housing affordability issues,[13] with the assumption that newly built units will be occupied by higher-income households, and lower-income households would be able to afford older units through the filtering process (Mast, 2023). However, empirical research

[11] In a survey of zoning ordinances of 25 of the largest 50 cities in the U.S., Hirt (2014) finds that 15 cities impose an absolute ban on business functions in residential areas. Across the cities that keep residential and business land uses completely separate, the overwhelming majority of the residential districts are low-density, ranging from 73 percent in Atlanta, Georgia, to 90 percent in Cleveland, Ohio, to nearly 100 percent in El Paso, Texas.

[12] Pioneering arguments for citywide zoning codes often presented the protection of property values as a tool to enhance equity by distributing the gains among the American masses rather than a few (Bassett, 1922). In practice, however, zoning codes have served the benefits of homeowners and businessowners over those of low-income renters. For various discussions of this topic in cities like New York, Chicago, Pittsburg, and others, see Fischel (2004).

[13] California and Massachusetts have implemented laws known as "builder's remedy," which generally allow developers to bypass the local/municipal authority to reject their applications as long as they achieve certain development goals. For California, see Elmendorf (2022). For Massachusetts, see Reid, Galante, and Weinstein-Carnes (2016).

is inconclusive about whether building more housing per se would address affordability issues. Some studies have found that new construction can even lead to higher home values nearby through amenity and aesthetic effects (Zahirovich-Herbert and Gibler, 2014), whereas some others argue that building new units would reduce rents in the surrounding area and limit displacement (Asquith, Mast, and Reed, 2019; Li, 2022). Even when new construction is found to lower rents and improve affordability, outcomes can vary by housing market segment, affecting the higher end of the market more than others (Stacy et al., 2023). The growing significance of affordability in the pro-supply discourse has urged many government entities to specifically encourage affordable housing in jurisdictions with restrictive zoning and isolated low-density land use districts. More than 500 inclusionary zoning programs across the country represent these efforts to promote affordable housing through mandatory or voluntary zoning reforms, although their effectiveness has been subject to criticism (Hickey, Sturtevant, and Thaden, 2014).

Despite all the policy efforts, zoning regulations continue to be a key element in the disparate distribution of low-income and/or multifamily housing development across the country.

Prior Work Done by New York State

Defining Well-Resourced Areas

New York State's 2016 Analysis of Impediments to Fair Housing Choice (2016 A-I) took a proactive, data-driven approach to understanding the spatial distribution of new affordable rental housing construction in the state. In this article, HCR identified 2,062 Census tracts in 59 of the state's 62 counties as potentially likely to provide educational and other opportunities for lower-income families and their children. This effort was informed by findings from HUD's Moving to Opportunity (MTO) research and other related studies (see Chetty, Hendren, and Katz, 2016; Turner, 2011). The 2016 A-I found a lack of affordable housing in well-resourced neighborhoods, and this shortage contributed to a disparity in access to valuable community assets. The agency subsequently began developing a standard methodology for identifying "opportunity census tracts," later renamed Well-Resourced Areas (WRA).

The agency published WRA designations in 2018 with a binary "in-or-out" structure, meaning that tracts were deemed either well-resourced or not. The criteria relied on two measures: the poverty rate as calculated by the U.S. Department of Commerce and a relative measure of school performance using New York State Education Department (NYSED) testing data in reading and math (see more details in the Tract-Level Data and Opportunity Metrics section of this article). The decision to use a relatively simple, binary designation was intended to keep the designation easy to understand and predict for the agency program staff and affordable housing developers.

Since 2018, HCR has updated the list of WRA tracts twice in 2020 and 2023. Exhibit 1 presents the share of WRA tracts in the three cycles.

Exhibit 1

Number of Well-Resourced Area Tracts as Defined by New York State Homes and Community Renewal

2018 Tracts			2020 Tracts			2023 Tracts		
WRA	Total	%	WRA	Total	%	WRA	Total	%
1,447	4,900	30	1,432	4,900	29	1,786	5,411	33

WRA = Well-Resourced Areas.
Sources: American Community Survey 5-year estimates, 2021; New York State Education Department, Annual English Language Arts and Mathematics Data Reporting, Grades 3-8

Incorporation of Well-Resourced Areas into HCR Programs

In recent years, HCR has increasingly incorporated WRA designations in its programs to further encourage developments in well-resourced neighborhoods. The agency amended its Qualified Allocation Plans (QAP)[14] for the federal LIHTC in 2019 and 2021 to define the construction of new multifamily rental housing in WRAs as a "State Housing Goal." HCR set aside a subset of its multifamily housing capital funding budget specifically for developments in these tracts. The agency has also adjusted its 2021 QAP—their section on Project Scoring and Ranking Criteria (Item F)—to assign five points to "Housing Opportunity," or building a project in a high-opportunity area.[15] Furthermore, the State's attention to well-resourced areas is also reflected in its most recent Request for Proposals for LIHTC in 2022. This document specifically mentions that Housing Opportunity Projects (HOP)—developments located in WRAs—may qualify for a 30-percent boost to their base credit amount (New York State Division of Housing and Community Renewal, 2022; 14).

The State's latest Assessment of Fair Housing made reference to HCR's use of federal housing subsidies (including HUD capital funding, vouchers, and LIHTC credits) to support new affordable multifamily rental housing in areas of high opportunity. The assessment considered multiple definitions of "areas of high opportunity," including grouping census tracts by the agency's designated WRAs. The assessment found that HCR's portfolio of newly awarded LIHTC projects from the period 2015–20 was concentrated in non-WRA census tracts, but it also noted a growing number of new units financed in WRAs. Exhibit 2 summarizes the share of LIHTC units in WRA and other neighborhoods (New York State Division of Housing and Community Renewal, 2016).

Exhibit 2

Concentrations of 2015–20 LIHTC Construction Starts in WRA vs. Non-WRA Tracts

	Multifamily Units (2015–20) per 1,000 Households	Multifamily Units, New Construction (2015–20) per 1,000 Households	Multifamily Units, Rehabilitation (2015–20) per 1,000 Households	Total Households
WRA	1.51	1.23	0.29	2,608,767
Non-WRA	10.68	5.32	5.36	4,634,320

LIHTC = Low-Income Housing Tax Credit. WRA = Well-Resourced Areas.
Sources: Administrative data from New York State Division of Housing and Community Renewal (DHCR) and New York State Housing Finance Agency

[14] The Internal Revenue Code (Section 42(m)(1)(B)) requires all housing credit agencies, including New York State, to prepare and adopt Qualified Action Plans for each funding round. This document generally lays out the agency's housing goals and priorities, and the criteria used for giving preference in allocating housing credits to selected project types.

[15] Low-Income Housing Credit Qualified Action Plan, 9 CRR-NY 2040.14 (2021).

Data and Methodology

This article combines two sets of data to develop the Neighborhood Resource Index and explore the placement of low-income multifamily homes in high-resource areas. These data sets are (1) administrative data on all newly constructed multifamily housing in New York State that benefited from state-administered LIHTC and (2) census tract-level data from various sources.

Administrative Data

The pool of LIHTC properties used in this article is a subset of all properties that have ever used state-administered LIHTC credits in New York State since the program's inception. Therefore, any reference to LIHTC properties in this article concerns only properties that received LIHTC through New York State, unless otherwise noted. In order to assess developers' decisions about the siting of low-income housing, this subset includes only new construction projects and excludes rehabilitation or preexisting properties. The study sample includes properties financed only by LIHTC credits and those that combined credits with other sources of capital financing from the state. The study sample excludes any LIHTC properties financed solely by entities other than HCR. The data also provide the number of units within each property, which can be a combination of subsidized and market-rate units.

One of the challenges of working with this dataset is determining a time identifier (e.g., year) for each property. The development process for residential properties usually spans multiple years and can be even longer for developments that use public financing and tax credits, especially when low-income developments face local opposition. The state's data management system stores several time identifiers for each project, including construction start, issuance of the certificate of occupancy, and a "placed in service" date. However, the data do not provide specific information about when siting decisions were made. Among the date variables available, the construction start date is closest to the beginning of the development's process, so that variable becomes the basis for assigning "year" values to the buildings.

The LIHTC sample include 4,180 newly constructed buildings since 1988 that were financed by tax credits administered by the state. When analyzed on their own, LIHTC properties are considered in their entire history, but in most cases, when analyzed in relation to their hosting census tracts, the sample is limited to 2015–23. The reason is twofold: First, as mentioned previously, New York State became more proactive in 2016 in its focus on well-resourced neighborhoods; second, from a methodological standpoint, the metrics of neighborhood resources rely mostly on 5-year average data that cover the period of 2016 through 2021, so they may not reliably represent the conditions of the neighborhoods where projects were built many years earlier. Given the article's approach to assigning year values to properties, it is possible that siting decisions for developments in the earlier part of the study period were made before 2015, but in the absence of a more accurate date identifier, the analysis uses 2015 as the cut-off point. This trimmed sample includes projects built after 2015 and has 1,425 observations. Exhibit 3 presents summary statistics of the state-administered LIHTC sample.

Exhibit 3

LITHC Properties in New York State

	Number of Buildings	Number of Units	Number (Share) of Income-Restricted Units	Share of Income-Restricted Units in Average Building (%)
1988–2014	2,735	57,743	52,973 (91.7)	94.9
2015–23	1,425	48,015	39,219 (81.7)	84.9

Source: Homes and Community Renewal's (HCR) administrative Low-Income Housing Tax Credits (LIHTC) data

Tract-Level Data and Opportunity Metrics

As briefly mentioned in previous sections, HCR's earlier definition of Well-Resourced Areas was based on two indicators: poverty rates and school test scores, motivated by work done by Chetty and colleagues (Chetty, Hendren, and Katz, 2016). According to that methodology, a census tract would be designated as WRA if (1) its poverty rate is lower than 10 percent and (2) its share of students rated proficient on mathematics and English Language Arts (ELA) tests exceeds the state's median. More specifically, the latter indicator captured the proportion of students in third through eighth grades scoring in the top two tiers of proficiency in each tract over the last 5 years, compared with the median proportion of proficient students across the entire state. This method was also based on a binary designation, with which census tracts were either designated as WRA or not. In its most recent iteration, this approach resulted in 1,606 census tracts qualifying as WRA out of the state's total of 4,900.[16]

These two criteria (poverty and test scores) are widely used in similar studies and usually embody many other implicit characteristics of places. For example, poverty can be highly correlated with the average educational attainment or unemployment rate of the neighborhood. However, a number of challenges and limitations motivated the authors to expand the previous work on WRA designation, as described in the following.

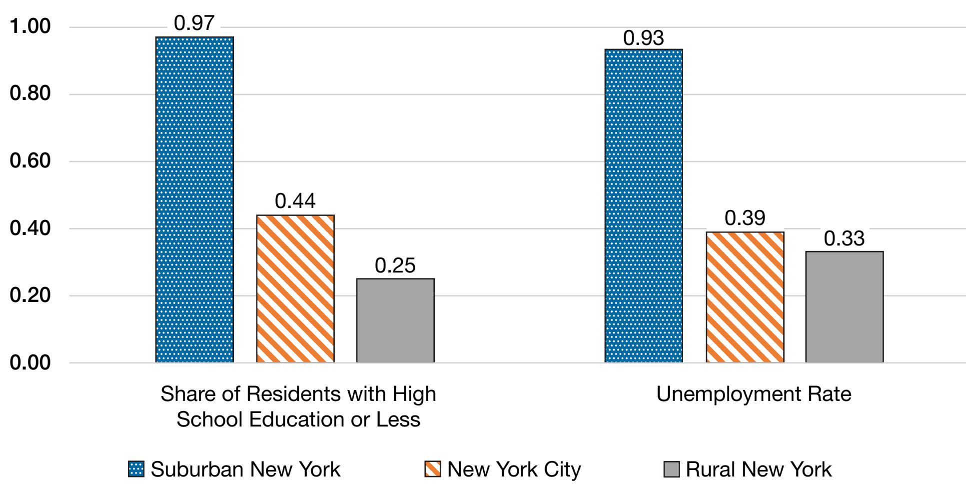
With regard to the school test score data, although they provide a widely accepted publicly available measure of school quality, the availability and nature of the data were noticeably disrupted by the COVID-19 pandemic. Educational assessments were not conducted in 2020, and subsequent assessments featured significantly lower participation rates than previous years, to the extent that the 2022 data were considered unreliable by the New York State Education Department (NYSED, 2022). In addition, the increasing number of students who opted out of standardized tests was another threat to test-score data credibility in terms of representing the state's students (Harris and Fessenden, 2015). These limitations necessitated using additional metrics to augment understanding of a neighborhood's education quality. These measures are described in the following section.

[16] The original analysis was using the Census Bureau's 2010 definition of census tract boundaries, when the total number of tracts in New York State was 4,900. According to the 2020 Census, which is used in the current study, the state includes 5,411 tracts.

Secondly, the nature of neighborhoods, and therefore ways of measuring their opportunity levels, can differ between dense urban areas like New York City, where an average of 31 people live in each acre of land, and small rural areas, with 1 resident living on every 10 acres. Within this context of diverse urban settings, the poverty rate—as a proxy for a multitude of other place-based characteristics—performs differently in different regions. Exhibit 4 shows two examples of this variation: Poverty rate at the tract level is more strongly correlated with educational attainment or unemployment rate in "suburban" New York State (the area comprised mostly of New York City's metropolitan area excluding the city itself) than in New York City or the state's rural areas.[17]

Exhibit 4

Correlation Between Poverty Rates and Other Neighborhood Attributes by Region



Source: American Community Survey 5-year estimates, 2021

To address the challenges imposed by the unavailability of reliable test score data and the regional variations in the correlation between poverty rate and other socioeconomic attributes, this study takes two steps: First, it broadens the number of indicators for neighborhood opportunity scores. After much testing and reviewing the available literature, the authors increased the number of indicators from the 2 original indicators to 24. These indicators were organized into four broad domains: income and employment; education; demographics; and housing and local amenities. The authors modeled the set of indicators and their classification generally after the categories used in the Child Opportunity Index and added new or more nuanced indicators in each domain.

The second remedy to potential risks of data unavailability was to classify neighborhoods into four groups according to their urban setting rather than treating them all similarly. This strategy aims to measure neighborhoods' resource levels relative to their regional contexts and counterparts. The authors' guide in classifying neighborhoods by region is the HUD Entitlement Communities, which are counties, major metropolitan areas, and cities with sizable populations eligible for

[17] For brevity, the article only presents the two examples of high school education and unemployment rate, but variation in poverty's correlation with other economic and social factors runs across many more indicators.

the Community Development Block Grant (CDBG) program.[18] In addition to identifying New York State's nonurban areas in this way, the study further subdivides the urban areas based on population densities. Exhibits 5 and 6 present the four regions and their characteristics.

Exhibit 5

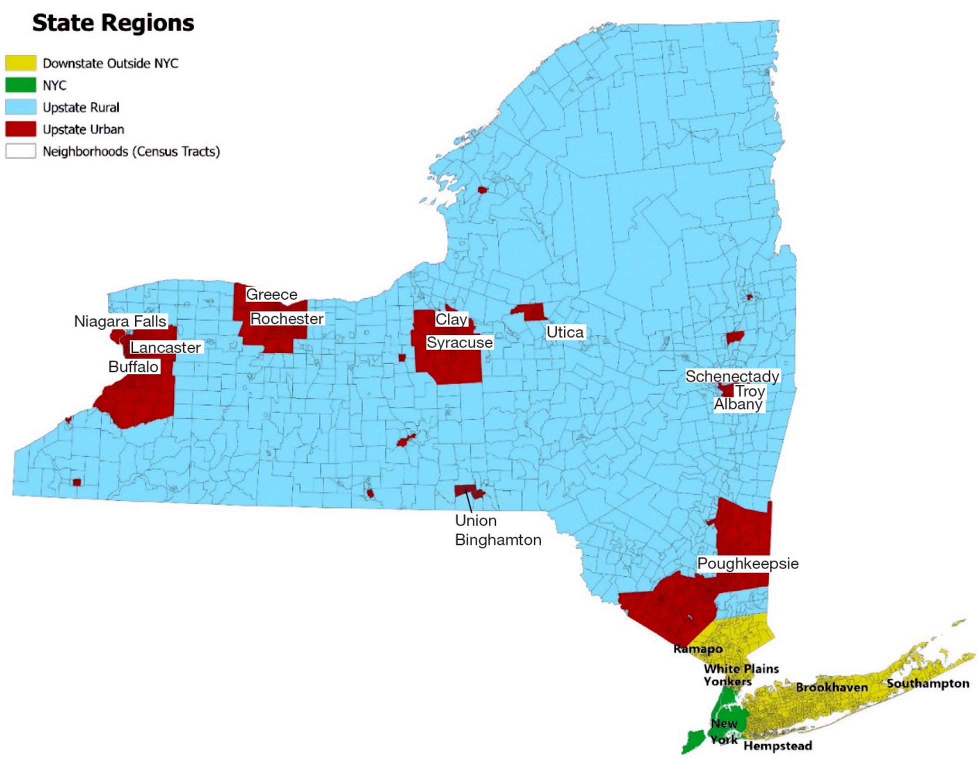
Four Regions Based on Urban Density

Region	Number of Census Tracts	Total Population	Population Density (persons per acre)
Downstate excluding New York City	980	4,057,833	2.58
New York City	2,237	8,238,733	31.05
Upstate Urban	1,013	3,365,462	1.12
Upstate Rural	1,046	3,348,282	0.12

Source: American Community Survey 2021 5-year estimates

Exhibit 6

Four Classes of Neighborhoods Across New York State



NYC = New York City.
Source: Authors' categorization of New York State regions guided by HUD Entitlement Communities

[18] For detailed information on Entitlement Communities, see https://www.hud.gov/program_offices/comm_planning/cdbg/entitlement-program.

The following section lays out the data sources for the four domains of neighborhood resourcefulness, and exhibit 7 provides the complete list of indicators.

A. **Income and Employment.** This group of indicators is motivated either by HCR's prior work to identify WRA tracts or other similar indices, such as the Child Opportunity Index (COI). The Income and Employment domain includes indicators for *poverty rate, level of reliance on cash transfers, labor force participation and unemployment rates, commute times, and the socioeconomic status of the residents and the workforce*. The latter set of indicators, including metrics like the proportion of high-wage earners living in the tract, are inspired by findings from Chetty et al. (2022) that attest to the positive impact of growing up among employed adults on children's future outcomes.

   Indicators in this category are constructed using data from the Census Bureau's American Community Survey (ACS) 5-year estimates for 2021 (the most recent year available as of this writing). For the workforce metrics, the authors used tract-level data from the Longitudinal Employer-Household Dynamics (LEHD), which is also a Census Bureau product providing information about the labor force in their tracts of residence. The latest year available for LEHD is 2020.

B. **Education.** In addition to using standardized test scores, the study includes a number of other indicators for approximating neighborhood educational qualities. Given the need for HCR to update the Neighborhood Resource Index with new data annually, the authors use publicly available data and construct indicators that approximate students, schools, and neighborhoods' educational performance. Those indicators are *student poverty, class size, school district funding, state aid to cover revenue gaps, and the share of proficient students in math and ELA*.

   In constructing the education indicators, the authors had to resolve the geographical mismatch inherent to the data. The education data used in the analysis are available either at the school or school district levels. Because the study defines neighborhoods as census tracts, it converted school or school district information to census tract level data.[19]

   Educational indicators are constructed using data from the New York State Division of the Budget (Enacted Budget Financial Plan for Fiscal Year 2023) and the New York State

[19] In the case of school-specific information, the authors faced three situations: (1) When a census tract had only one school, the authors assigned it the characteristics of that school; (2) when a census tract had more than one school, they assigned it the characteristics of the school that was closest to the tract's centroid; and (3) when a census tract did not have any schools in it, they assigned it the characteristics of the closest school to its boundaries, conditional on some overlap between the tract and the school's catchment area.

In the case of school district information, boundaries do not usually align with those of census tracts. Here, there were two possible situations: (1) When multiple census tracts fell within the same school district, the authors assigned all of them the same characteristics of the district; and (2) when one census tract spanned over multiple school districts, they assigned it the characteristics of the district that entailed the tract's centroid.

The authors undertake these reconciliation strategies to convert educational variable to census-tract-level ones. However, they acknowledge the imperfections that they introduce to the data. For example, by assigning the characteristics of the closest school to a census tract, the study makes the assumption that living in that particular tract is synonymous with attending that specific school, which may not always be the case. In the absence of a perfect solution, the authors chose to reconcile the education data with our other data as explained previously.

Education Department. Data on school poverty (free and reduced lunches) are used at the high school level, and the other metrics used data at the school district level.

C. **Local Demographics.** The demographic indicators are generally motivated by the COI and Chetty and colleagues' Opportunity Atlas (2018). They include *household structure, educational attainments, and access to health insurance*.

Demographic indicators are constructed using data from the Census Bureau's ACS 5-year estimates for 2021.

D. **Housing and Local Amenities.** Indicators in this domain include *housing vacancy, overcrowding, adequate utilities, neighborhood residential stability, and proximity to health and recreational facilities*.

Data for health facilities are from the New York State Facility Map (health.data.ny.gov), and data on recreational facilities are from the New York State GIS Clearinghouse. All other indicators in this domain use ACS 5-year estimates for 2021.

Exhibit 7

Neighborhood Resource Metrics (1 of 2)

Domain	Indicator	Description	Reference
Income and Employment	**Poverty rate**	Rank in distribution across all tracts	Acevedo-Garcia et al. (2014); Ellen et al. (2018)
	Public assistance share of income	Tract's aggregate transfer income relative to the total	Acevedo-Garcia et al. (2014)
	Median household income	Rank in distribution across all tracts	Acevedo-Garcia et al. (2014)
	Labor force participation		Chetty et al. (2018); McClure and Schwartz (2021)
	Unemployment rate		Acevedo-Garcia et al. (2014); McClure and Schwartz (2021)
	Average commute time	Rank in distribution across all tracts	Acevedo-Garcia et al. (2014); Ellen et al. (2018)
	Labor force socioeconomic status		Chetty et al. (2018); Lens (2014)
	Wages	Proportion earning $3,333 or more per month	
	Telecommuters	Proportion working remotely	
	Education	Proportion with a high school degree or less	

Exhibit 7

Neighborhood Resource Metrics (2 of 2)

Domain	Indicator	Description	Reference
Education	**Share of free & reduced lunches**		Acevedo-Garcia et al. (2014)
	Student-to-teacher ratio		Chetty et al. (2018)
	Spending per pupil		
	State school aid	Distributed in inverse proportion to local fiscal capacity	
	Standardized Test Proficiency	Share of students in the top 2 tiers of math and ELA	HCR; McClure and Schwartz (2021)
Demo-graphics	**Married couples**		Acevedo-Garcia et al. (2014)
	Education level	Share with a high school degree	Acevedo-Garcia et al. (2014)
	Health insurance coverage	Insured-uninsured ratio	Acevedo-Garcia et al. (2014)
Housing and Amenities	**Vacancy rates**		Acevedo-Garcia et al. (2014)
	owners & renters		
	Overcrowding	Share of units with 1.5 or more people per room	
	Inadequate access to utilities	Lacking adequate plumbing, kitchen, telephone	
	Residential stability	Proportion of residents that moved in before 2010	Chetty et al. 2018
	Access to health facilities	Miles to first hospital	
	Access to parks	Miles to the first state park	Acevedo-Garcia et al. (2014); Ellen et al. (2018)

ELA = English Language Arts. HCR = Homes and Community Renewal.

Sources: American Community Survey 5-year estimates, 2021; Longitudinal Employer-Household Dynamics, 2020; New York State Division of the Budget Enacted Budget Financial Plan for Fiscal Year, 2023; New York State Education Department, 2023; New York State GIS Clearinghouse

Indexing and Weights

After compiling data on all indicators, the study generates resource index scores for every census tract within each of the four regional classes in the following manner: First, it standardizes individual indicators—which have various units of analysis or scales—by generating *z*-score equivalents. Subsequently, it generates domain-specific scores by adding up the individual indicators within each domain and then converts the domain scores to a scaled range of 0 to 25. This range is a product of having four domains, and the study assigns equal weights to each of them in constructing the aggregate score. As exhibit 7 shows, the number of indicators each has among the four domains is not equal. In the absence of definitive theories or empirical findings about how metrics should be weighed,[20] the analysis is based on the assumption that domains bear equal importance in defining neighborhoods' level of resource, or, in other words, have equal weights.

[20] Other studies, such as Acevedo-Garcia et al. (2014) generate weights based on regressing future child outcomes on individual indicators. This article does not engage longitudinal data on child outcomes.

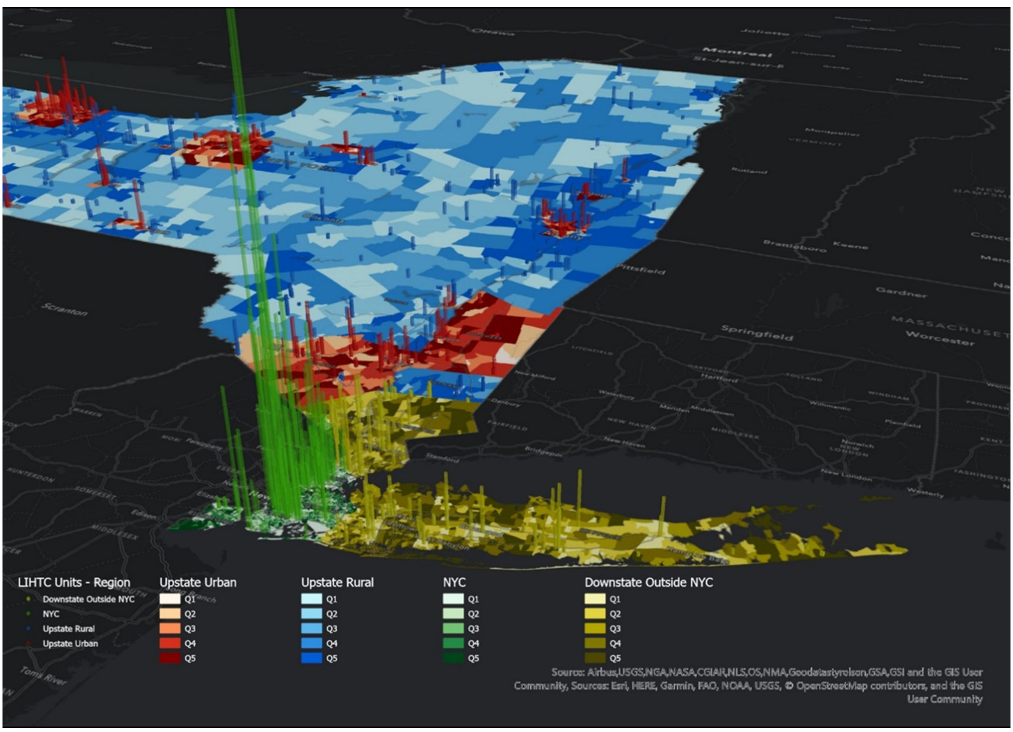
The final resource score is the sum of the four domain-specific scores, and it ranges between 0 and 100. In the conversion and summation process, the scoring scheme takes into account the different nature of various metrics because some imply a higher resource status for tracts that rank higher (e.g., median household income), and some others imply a lower status when tracts rank higher (e.g., unemployment rate). The outcome of this procedure, i.e., the four domain-specific scores and the final aggregate score, indicate the level of neighborhood resources on an escalating basis, which means higher-resource neighborhoods have scores closer to 100.

To exclude neighborhoods that are not residential or are irrelevant to the purposes of this study, the scoring scheme excludes census tracts that had populations less than 10 or were entirely undevelopable. This adjustment reduces the sample size from 5,411 to 5,270 tracts.

Exhibit 8 provides a view of the four regions (in different colors), Neighborhood Resource Index (in shades), and the number of LIHTC units (bars). Darker shades indicate higher resource scores, and bars represent tens of units.

Exhibit 8

LIHTC Units and Neighborhood Opportunity Scores



NYC = New York City. Q = quintile.

Sources: Homes and Community Renewal's (HCR) administrative Low-Income Housing Tax Credits (LIHTC) data; Authors' analysis of the following data sources: American Community Survey 5-year estimates, 2021; Longitudinal Employer-Household Dynamics, 2020; New York State Division of the Budget Enacted Budget Financial Plan for Fiscal Year, 2023; New York State Education Department, 2023; New York State GIS Clearinghouse

Discussion of Findings

Profile of Neighborhoods in New York State

Resources are distributed less equitably in high-resource areas than elsewhere. As is implied by the indicators laid out in the previous section of this article, neighborhoods with higher scores have wealthier and more educated residents, better schools, greater shares of homeowners, and lower vacancy rates. Although these attributes may signify a more desirable neighborhood, they can also characterize highly exclusive and homogenous ones. Social and developmental gains from diversity are highlighted in the academic literature (e.g., Chetty et al., 2022), and the authors want to know whether equity and diversity are implied by the article's resource score. This section presents a description of census tracts at the bottom and top of the Neighborhood Resource Index distribution from the standpoint of equity and diversity. The authors use a number of simple indicators of equity, listed in exhibit 9, that aim to demonstrate differences in the distribution of resources between the top and bottom groups. These indicators are set up so that higher values would represent higher inequality or lower diversity (the only exception is the Herfindahl–Hirschman Index of industrial concentration, with which lower scores imply more diversity).

Exhibit 9 compares the average census tract in the bottom and top quintiles of the resource score distribution (or low- versus high-resource neighborhoods) in terms of the nine equity indicators in each of the four regions. Two indicators (the ratio of high-wage workers to low-wage ones and the number of high earners relative to low earners) specifically aim to measure income inequality. As demonstrated by both measures, neighborhoods with higher scores have much more inequitable income distributions compared with low-resource neighborhoods.

Gender and racial compositions, captured by the ratio of male to female workers and White residents to non-White, are also more biased toward men and White residents in neighborhoods with high scores. Upstate rural can be considered an exception here, where the White to non-White ratio is lower in highest-opportunity areas, perhaps because of the different composition of agricultural workers in that region.

With regard to housing inequality, a few points are worth highlighting. First, high-scoring neighborhoods are predominantly owner-occupied across all regions, which shows a lack of tenure diversity to various extents. Secondly, the ratio of single-family homes to other housing types is much larger in well-resourced neighborhoods, especially in the upstate urban and suburban downstate areas. Differentials between the highest- and lowest-value homes are also larger in the higher-scoring neighborhoods. Together, these measures indicate a much greater inequitable distribution of housing resources in areas that score high on resources.

Regions rank differently in terms of different equity metrics. Among the four regions, the first one (Downstate Outside NYC) ranks least equitable in four out of nine equity and diversity metrics: It has the most disproportionate number of top earners relative to low earners, homeowners relative to renters, single-family homes relative to multifamily ones, and high-value homes relative to low-value ones. The New York City region comes second, ranking most disproportionate in three out of nine equity metrics.

Exhibit 9

Equity and Diversity in Low- Versus High-Resource Neighborhoods

	Downstate Outside NYC		NYC		Upstate Rural		Upstate Urban	
	Bottom Quintile	Top Quintile	Bottom Quintile	Top Quintile	Bottom Quintile	Top Quintile	Bottom Quintile	Top Quintile
High- to Low-Wage Worker Ratio	2.1	3.6	1.6	3.9	1.7	2.8	1.0	2.7
White to Non-White Worker Ratio	2.4	6.5	0.8	3.1	18.5	17.1	1.5	11.9
Male to Female Worker Ratio	0.9	1.0	0.8	1.0	1.0	1.0	0.9	1.0
Top to Bottom Earners Ratio (Count)	4.1	18.0	1.0	7.9	1.0	7.6	0.4	8.6
Herfindahl–Hirschman Index of Industry Mix	0.11	0.09	0.12	0.10	0.11	0.10	0.11	0.09
Owner to Renter Ratio	2.7	21.2	0.5	2.4	2.9	7.0	0.6	11.8
Single-Family to Multifamily Ratio	15.7	63.0	1.5	5.5	58.3	63.1	11.6	49.1
Rent to Mortgage Ratio (Median)	0.7	0.6	0.6	0.7	0.7	0.6	0.9	0.7
High to Low Home Value Ratio	5.3	54.4	8.6	36.9	0.3	3.8	0.1	2.4

NYC = New York City.

Sources: Authors' analysis of the following data sources: American Community Survey 5-year estimates, 2021; Longitudinal Employer-Household Dynamics, 2020; New York State Division of the Budget Enacted Budget Financial Plan for Fiscal Year, 2023; New York State Education Department, 2023; New York State GIS Clearinghouse

To go beyond just the average tract and learn about the covariation between the Neighborhood Resource Index and equity, exhibit 10 presents nine scatterplots that visualize each equity metric in relation to resource scores. This exercise shows that when the regression coefficient (or the slope of the best-fit line) is considered, the Upstate Urban region ranks least equitable in six out of nine indicators. The slopes shown on each scatter plot indicate the rate of change in equity relative to changes in resource scores. For example, when the ratio between high- and low-wage workers increases by 1 unit, the neighborhood resource score increases by 18.3 units (i.e., 18.3 percent) in the Upstate Urban region and 9.02 percent in the Downstate region.

Exhibit 10

Regional Variations in the Associations Between Neighborhood Resource Score and Equity (1 of 2)

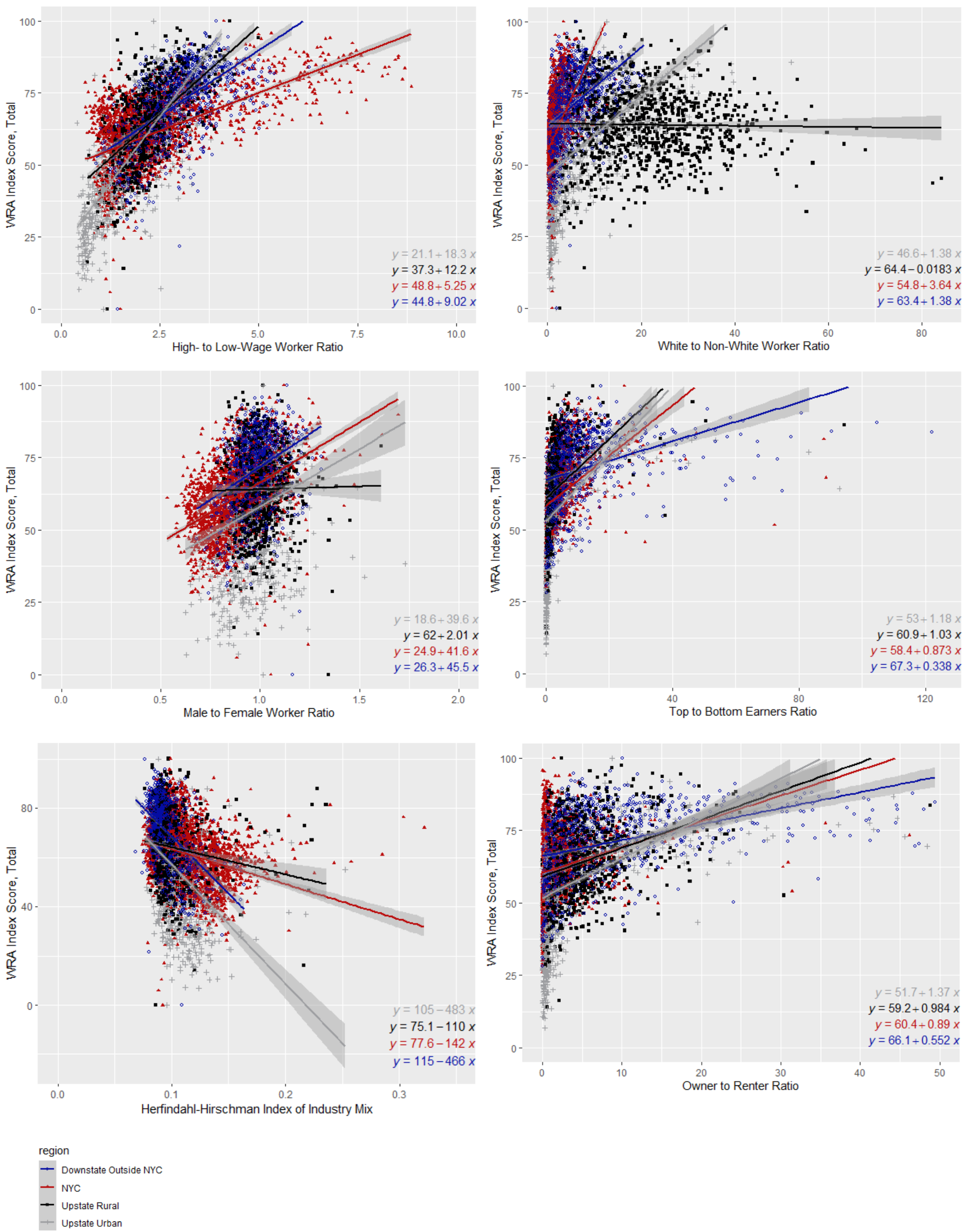
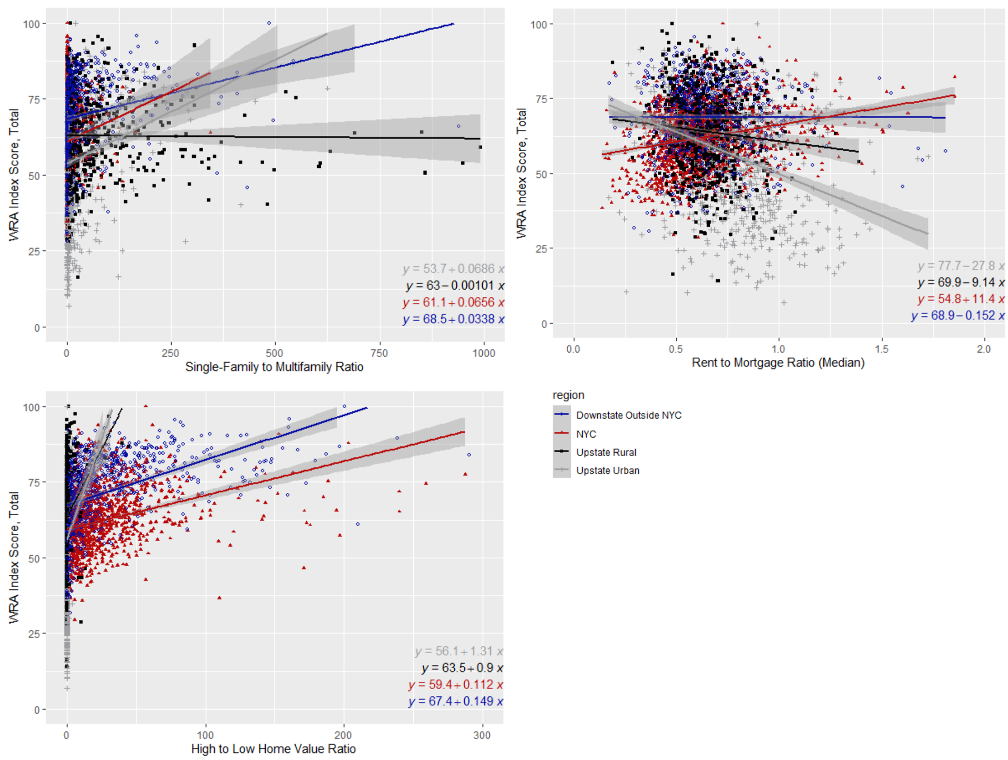


Exhibit 10

Regional Variations in the Associations Between Neighborhood Resource Score and Equity (2 of 2)



NYC = New York City. WRA = Well-Resourced Areas.

Sources: Authors' analysis of the following data sources: American Community Survey 5-year estimates, 2021; Longitudinal Employer-Household Dynamics, 2020; New York State Division of the Budget Enacted Budget Financial Plan for Fiscal Year, 2023; New York State Education Department, 2023; New York State GIS Clearinghouse

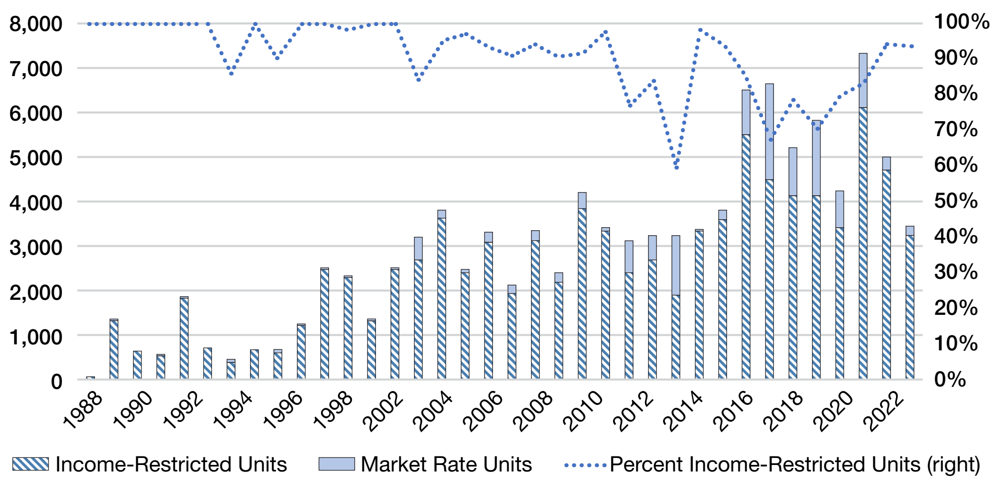
Although the presented analysis of equity and diversity is broad and general, it implies a noteworthy reality about the siting of low-income housing units: Resources in highest-scoring neighborhoods are distributed rather inequitably, which confirms what previous studies have highlighted. As an example, Reid's (2019) study of residents from 18 LIHTC buildings in California showed that children of lower-income families face barriers to opportunity that are driven by a lack of access to local resources more than by neighborhood characteristics. This situation suggests that policymakers need to think beyond the siting of affordable housing to ensure that low-income families have access to the resources shown to improve future outcomes. The appendix provides a detailed description of each equity/diversity indicator.

LIHTC Buildings vis-à-vis Neighborhood Resource Scores

The income-restricted share of LIHTC units has been increasing recently. Developers of LIHTC properties are allowed to rent a subset of units at the market rate to supplement the below-market rate rents. Although the share of income-restricted units in the past few years in the sample of state-administered LIHTC properties has been below the historical average, the analysis shows an upward trend in that share since 2016. That timeframe follows the state legislature enacting additional funding for affordable housing as part of the state's housing plans. Exhibit 11 demonstrates the number of income-restricted and market-rate units over time.

Exhibit 11

LIHTC New Construction in New York



Source: Authors' analysis of Homes and Community Renewal's (HCR) administrative Low-Income Housing Tax Credits (LIHTC) data

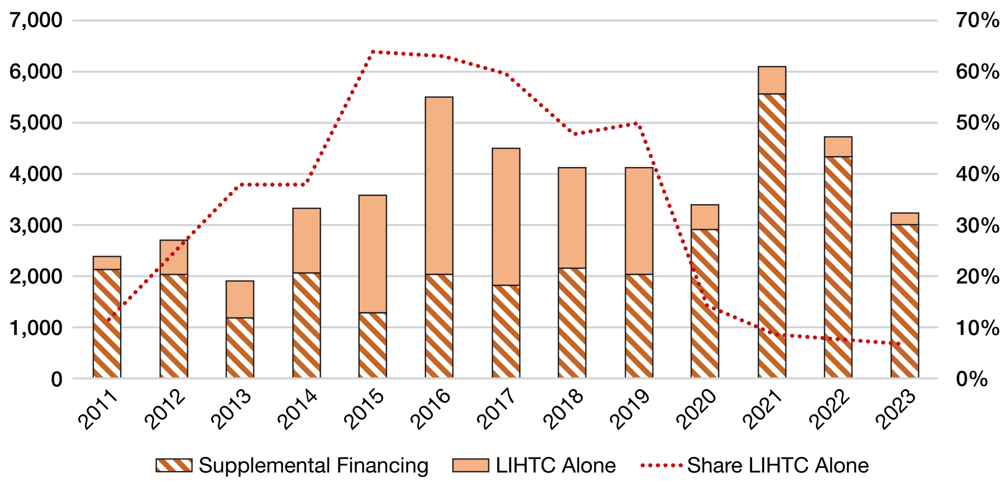
Most developments augment LIHTC with other capital financing. Construction costs can be particularly high in high-resource neighborhoods, including those for land acquisition and permit fees. Therefore, building affordable housing in these areas would require substantial gap financing. As confirmed by the study's data, most developers avail themselves of additional capital subsidies from the state and other sources to supplement LIHTC credits.[21] Across properties for which the authors could confirm this information (properties built after 2010), approximately 34 percent of units were in buildings that used only LIHTC to finance the development. The share of these units peaked during the 2015–19 period and then most recently declined to less than 10 percent. This trajectory is primarily explained by the rise and fall in the prevalence of developments with smaller shares of affordable units. In the earlier years of the 2010s, when New York City's residential rental market was rebounding from its Great Recession-era slump, developers showed noticeable interest in new developments with a small number of affordable units. The large share of market-rate units in these developments would lower the need for substantial subsidies, which increases the likelihood of using LIHTC as the only source of subsidy. However, over time, New York State decided to focus its support on developments with larger shares of affordable homes, which

[21] These credits include funding through the State's Housing Trust Fund, capital funding appropriated by the State Legislature as part of the Five-Year Capital Plan, or federal sources such as the Federal Housing Trust Fund and the HOME Investment Partnership.

typically require additional financial support due to their restricted rent income.[22] As a result, units in LIHTC-only properties comprise a smaller share of all newly built units after 2019. Exhibit 12 shows the distribution of units by source of subsidy over time.

Exhibit 12

Affordable Units by Type of Subsidy



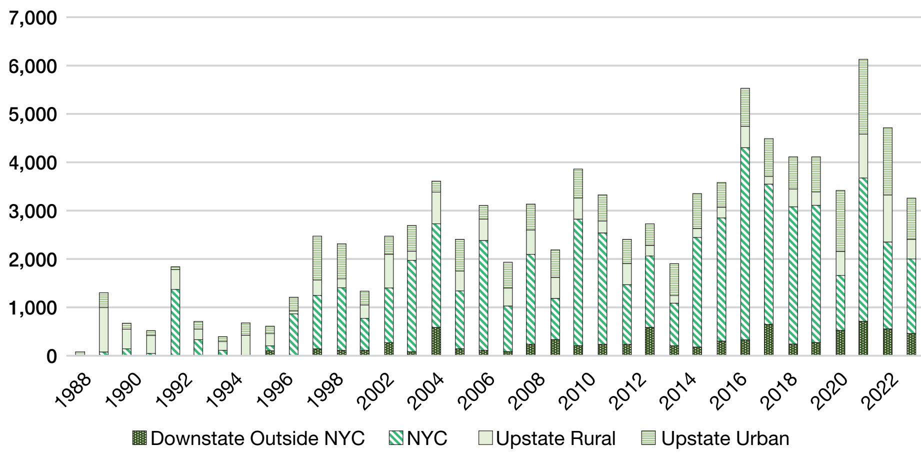
Source: Authors' analysis of Homes and Community Renewal's (HCR) administrative Low-Income Housing Tax Credits (LIHTC) data

New York City leads other regions in number of affordable units by far. Given that more than 43 percent of the state's population and two-thirds of its rental units are within the boundaries of New York City, it is not surprising that the city has the largest number of LIHTC units among the four regions. Factors such as a large low-income population, a history of local low-income housing assistance programs, strong tenant political representation, and land use policies that accommodate multifamily developments also make New York City a relatively easier place to develop subsidized housing. In the study sample, 56 percent of all units are located in New York City, as demonstrated in exhibit 13.

[22] New York City's local property tax structure, zoning code, and robust local administrative capacity make it somewhat unique in the state. A number of substantial incentives or programs exist for developers to build affordable housing in the city, including in its well-resourced neighborhoods. Examples include Mandatory Inclusionary Housing, which aims to permanently increase the stock of affordable units through zoning modifications, and the 421-a program, which provides property tax exemptions in return for developers setting aside a share of newly built units as affordable. These programs have shown sizeable impact, especially in motivating developments with a mix of market-rate and affordable units in well-resourced neighborhoods. Over time, New York State and New York City determined that these programs were more efficient and effective tools than LIHTC in encouraging multifamily housing in high-score areas, and gradually diverted away from using LITHC as an incentive for those developments in New York City. This gradual shift of approach does not mean that new affordable units are not being built in New York City's well-resourced areas, but rather than other policy tools are used more heavily than LIHTC to incentivize them. Details of the State's policy change are explained at https://www.crainsnewyork.com/article/20140127/REAL_ESTATE/140129891/state-tweaks-affordable-housing-incentive.

Exhibit 13

Affordable Units by Region

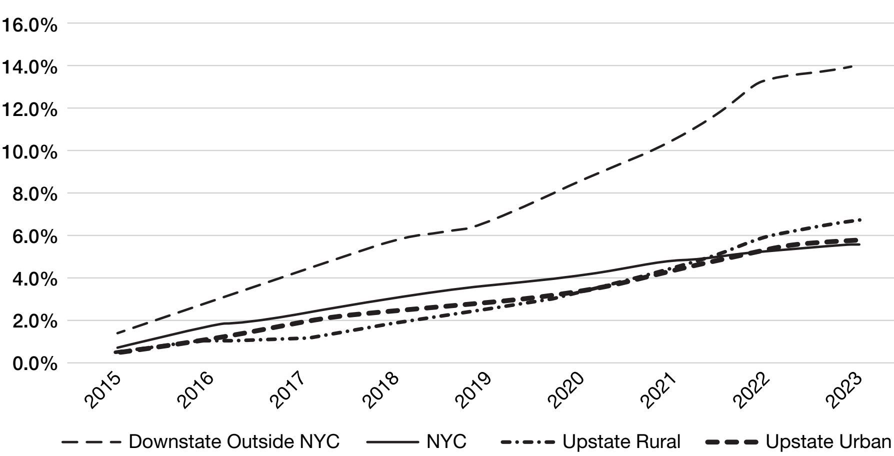


NYC = New York City.
Source: Authors' analysis of Homes and Community Renewal's (HCR) administrative Low-Income Housing Tax Credits (LIHTC) data

Siting of LIHTC developments has been expanding to more neighborhoods. On an annual basis, LIHTC developments occur in a range between 0.4 and 4.5 percent of neighborhoods, depending on the region. The authors' analysis shows that over time, developments have been expanding over a growing number of neighborhoods since 2015. Exhibit 14 shows the growth in the share of neighborhoods with LIHTC development each year—or the percent change in the new neighborhoods with LIHTC developments. Upstate Urban leads other regions in diversifying LIHTC buildings across space. This trend may be driven in part by the noncontiguous geography of this region according to the authors' definition, which is comprised of cities in different locations of the state.

Exhibit 14

Cumulative Share of Neighborhoods with LIHTC Developments



Source: Authors' analysis of Homes and Community Renewal's (HCR) administrative Low-Income Housing Tax Credits (LIHTC) data

More homes are built in high-resource areas now than before. Overlaying geocoded LIHTC developments with census tracts reveals two points. First, neighborhoods with any LIHTC developments since 2015 have lower average resource scores than their counterparts without LITHC buildings. This finding can be explained in part by the structure of the LIHTC program, as described in the Review of Theoretical and Empirical Literature section of this article. This score differential between neighborhoods with and without LIHTC developments ranges from 4 points in the Upstate Rural region to nearly 16 points in Upstate Urban. Given the scaling of the Neighborhood Resource Index in this article, this means that LIHTC buildings are in neighborhoods with an average of 4 to 16 percent fewer resources. When quintiles of resource score are considered, the majority of LIHTC units built after 2015 are in census tracts in the bottom quintile. This observation is shown in exhibit 15, which documents that the percentage of LIHTC units built in the highest quintile of the Neighborhood Resource Index is small in all regions and usually disproportionate to the renter share of residents. New York City stands out because it has placed a noticeable 28 percent of its units in top-scoring neighborhoods. However, as can be calculated from the information in exhibit 15, the average number of units per building is much higher in the top quintile tracts of NYC than in lower quintiles. This result implies that a relatively small number of high-density developments in a high-scoring neighborhood could considerably move the share of total units in high-resource neighborhoods in that region.[23]

Secondly, the analysis also shows that despite the scarcity of LIHTC in high-scoring neighborhoods, more units have been built in well-resourced neighborhoods after 2015 than before. As exhibit 15 lays out, the share of new units built in the highest-scoring neighborhoods has grown in all regions except New York City, with some variation among them. Gains are more pronounced in the downstate suburban and upstate rural regions, where the top-scoring neighborhoods go from having no affordable units in the 2010–15 period to hosting 4 to 8 percent of the region's units after 2015. As previously discussed in the Incorporation of Well-Resourced Areas into HCR Programs section of this article, these gains can at least partly be attributed to New York State's proactive efforts to motivate building in well-resourced neighborhoods in recent years.

New York City is the exception again. As explained previously, the availability of LIHTC credits in the early years of the 2010s for developments with very small shares of affordable housing allowed developers to place income-restricted units in well-resourced neighborhoods—in the same building as the upscale ones—without much need for substantial additional subsidies. This situation created a higher baseline in New York City for the share of new units in well-resourced neighborhoods in the early 2010s. As the state has gradually moved away from supporting this type of development, New York City has been converging toward the typical pattern in other regions. Therefore, the city does not demonstrate the same growth observed in all other regions in the share of new LIHTC units in well-resourced neighborhoods. It is important to note that this analysis includes only the affordable units supported by LIHTC, whereas New York City has several other mechanisms

[23] The study's data confirm that this was evidently a common pattern in NYC: Of the 6,237 units located in the top quintile of resource score in NYC, 5,975 (96 percent) were in 80/20 buildings with relatively small shares of affordable housing (20 percent or less), which enabled developers to build in high-income neighborhoods.

for promoting the development of new affordable units in well-resourced areas (see footnote 21). Exhibit 16 visualizes this growth in each region.[24]

Exhibit 15

Distribution of LIHTC Units in Quintiles of Neighborhood Resource Score

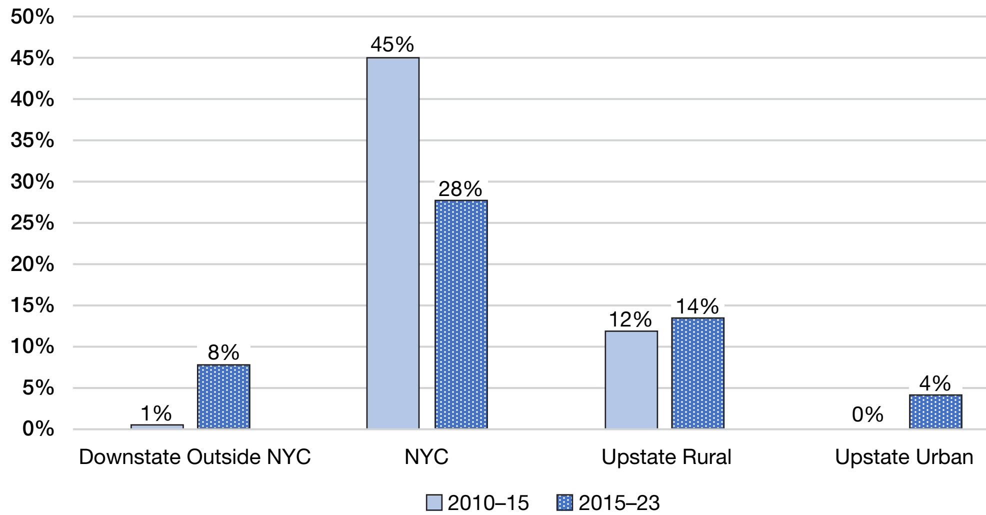
	Downstate Outside NYC					NYC				
	Lowest Quintile of Resource Score	**Quintile 2**	**Quintile 3**	**Quintile 4**	**Highest Quintile of Resource Score**	**Lowest Quintile of Resource Score**	**Quintile 2**	**Quintile 3**	**Quintile 4**	**Highest Quintile of Resource Score**
Mean Resource Score	51.2	64.8	72.8	79.1	85.9	44.3	55.2	61.7	67.6	77.8
Number of Tracts	196	196	196	196	196	447	447	447	447	447
Share of Households	0.21	0.40	0.12	0.18	0.09	0.49	0.18	0.08	0.05	0.20
Share of Renters	0.36	0.36	0.11	0.13	0.04	0.51	0.17	0.08	0.05	0.19
LIHTC Buildings	39	57	22	28	22	102	41	11	9	24
Percent LIHTC Buildings	*23%*	*34%*	*13%*	*17%*	*13%*	*55%*	*22%*	*6%*	*5%*	*13%*
LIHTC Units	1,698	1,388	383	303	319	9,288	2,939	1,464	2,485	6,237
Percent LIHTC Units: 2015–23	***42%***	*34%*	*9%*	*7%*	***8%***	***41%***	*13%*	*7%*	*11%*	***28%***
Percent LIHTC Units: 2010–14	***61%***	*24%*	*9%*	*6%*	***1%***	***18%***	*13%*	*7%*	*17%*	***45%***
	Upstate Rural					**Upstate Urban**				
	Lowest Quintile of Resource Score	**Quintile 2**	**Quintile 3**	**Quintile 4**	**Highest Quintile of Resource Score**	**Lowest Quintile of Resource Score**	**Quintile 2**	**Quintile 3**	**Quintile 4**	**Highest Quintile of Resource Score**
Mean Resource Score	45.9	57.3	64.2	71.0	81.5	28.5	46.9	60.2	69.2	79.4
Number of Tracts	208	210	209	209	208	202	202	203	202	202
Share of Households	0.22	0.12	0.13	0.37	0.16	0.33	0.25	0.20	0.14	0.08
Share of Renters	0.33	0.10	0.13	0.34	0.10	0.44	0.30	0.14	0.09	0.03
LIHTC Buildings	74	39	55	97	49	341	185	104	86	40
Percent LIHTC Buildings	*24%*	*12%*	*18%*	*31%*	*16%*	*45%*	*24%*	*14%*	*11%*	*5%*
LIHTC Units	1,694	645	423	1,111	611	3,230	1,472	1,500	1,688	343
Percent LIHTC Units: 2015–23	***38%***	*14%*	*9%*	*25%*	***14%***	***39%***	*18%*	*18%*	*21%*	***4%***
Percent LIHTC Units: 2010–14	***32%***	*28%*	*17%*	*11%*	***12%***	***34%***	*26%*	*19%*	*20%*	***0%***

NYC = New York City.
Sources: Homes and Community Renewal's (HCR) administrative Low-Income Housing Tax Credits (LIHTC) data; Authors' analysis of the following data sources: American Community Survey 5-year estimates, 2021; Longitudinal Employer-Household Dynamics, 2020; New York State Division of the Budget Enacted Budget Financial Plan for Fiscal Year, 2023; New York State Education Department, 2023; New York State GIS Clearinghouse

[24] Ideally, the comparison between the two periods shown on exhibit 7 (2010–15 vs. 2015–23) would allow the neighborhood resource scores to vary by year or to reflect the existing local condition of resources in any given year. However, given that the score draws on data from multiple sources, the authors were not able to obtain consistent data for the earlier period to replicate the indexed score for 2010 to 2015. Given that limitation, the analysis applies the measure of neighborhood resources in the 2015–23 period to the earlier period, which is imperfect but the only practical option.

Exhibit 16

Gains and Losses of New LIHTC Units in the Top Quintile of Neighborhood Resource Score Distribution



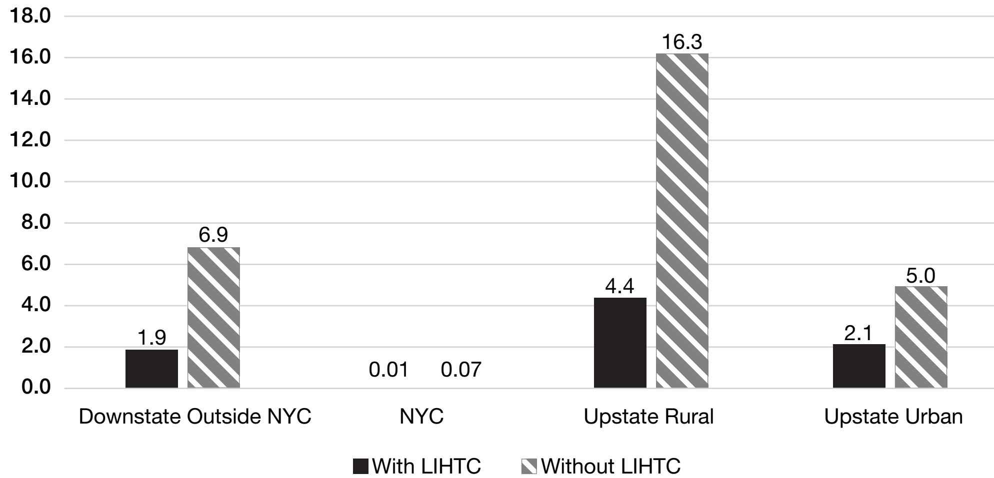
NYC = New York City.

Sources: Homes and Community Renewal's (HCR) administrative Low-Income Housing Tax Credits (LIHTC) data; Authors' analysis of the following data sources: American Community Survey 5-year estimates, 2021; Longitudinal Employer-Household Dynamics, 2020; New York State Division of the Budget Enacted Budget Financial Plan for Fiscal Year, 2023; New York State Education Department, 2023; New York State GIS Clearinghouse

Affordable units are built in areas with an existing stock of multifamily housing. The authors propose that the scarcity of apartment buildings in a neighborhood could be a proxy for restrictive land use regulations that hinder the development of multifamily buildings in general and low-income multifamily housing in particular. This hindrance of development is well documented in the urban economics literature, as reviewed in a previous section. To examine this proposition, the authors compare the prevalence of multifamily buildings in neighborhoods with LIHTC units with those without. Exhibit 17 visualizes the comparison in terms of the ratio of single-family to multifamily structures (multifamily here is defined as structures with five or more units; see the description of the equity and diversity metrics in the appendix). Across all regions of the state, more LIHTC units have been built in neighborhoods with a larger stock of multifamily residences. This circumstance highlights the likely role of local regulatory infrastructure in the development of low-income rental homes.

Exhibit 17

Multifamily Stock in Neighborhoods With and Without LIHTC Units



NYC = New York City.

Sources: Homes and Community Renewal's (HCR) administrative Low-Income Housing Tax Credits (LIHTC) data; Authors' analysis of the following data sources: American Community Survey 5-year estimates, 2021; Longitudinal Employer-Household Dynamics, 2020; New York State Division of the Budget Enacted Budget Financial Plan for Fiscal Year, 2023; New York State Education Department, 2023; New York State GIS Clearinghouse

Conclusions

This study's main contribution was to set up an inclusive, regionally focused measure of neighborhood resources in New York State and to demonstrate the measure's practical application by using it to assess the siting pattern of LIHTC projects in the state. The analyses presented in this article augment existing measures of local opportunity by diversifying the underlying metrics and allowing the Neighborhood Resource Index to vary based on regional conditions. The article finds that the state's LIHTC units have been concentrated in a relatively small number of neighborhoods and that those neighborhoods typically scored lower in terms of neighborhood resources, as measured by the presented index. However, the analysis of recent siting patterns suggests that units are being built in an increasing number of neighborhoods, including more areas with higher Neighborhood Resource Index scores, with some regional variability. These findings offer some useful insights into the challenges to developing affordable housing in well-resourced areas, the strategies to address this disparity, and the issue of equitable access to resources more broadly.

The article offers some potential explanations for why LIHTC units have not been built where local resources are greatest. The most straightforward explanation is higher financing, land acquisition, and construction costs. Developing housing with below-market rents is more burdensome in high-resource, high-cost neighborhoods than elsewhere due to higher opportunity costs (the market rent that could be charged absent the subsidy) and general costs attached to land acquisition, permit fees, and construction. The article's findings acknowledge that, at least in New York State, most newly built units augment LIHTC with other capital funding. Nonetheless, the ability of the program to move low-income renters to high-resource areas, as measured by the share of units built in those areas, is limited, especially outside of New York City.

Another possible explanation for the concentration of affordable units in low-resource neighborhoods—apart from the structure of the program itself—is rooted in land use regulations. As the article documents, new LIHTC units are rare in neighborhoods where single-family homes dominate and are built mostly in census tracts with low existing ratios of single-to-multifamily housing. This situation implies that restrictive zoning may play a role in the distribution pattern of LIHTC developments. Although the analysis does not include a formal analysis of land use regulations in the study areas, it finds evidence that high-resource neighborhoods are exclusive and homogenous, including with respect to housing types and income groups.

Based on these findings, the article recommends a few strategies that can facilitate achieving the goals of Fair Housing and moving to opportunity. States should be proactive in rewarding new developments in high-resource areas to alleviate the cost burden for potential developers. New York State has taken this initiative by announcing a specific housing goal dedicated to building homes in Well-Resourced Areas in its Qualified Allocation Plans and setting aside capital funding for developments in these areas.[25] The results from the analysis comparing siting patterns of a subset of New York's LIHTC units produced before and after 2015 suggest that these efforts are having positive impacts.

Obstacles that arise from zoning restrictions can also be addressed at various levels. Where local governments are pressured by their electoral base to deflect new multifamily units for the sake of property-value protection, states may have a role to play. Several states, including California, Massachusetts, and Oregon, have passed laws in recent years to either incentivize local communities to allow for more multifamily housing or denser development patterns or provide a remedy for developers seeking to develop qualifying multifamily projects in areas with restrictive zoning codes.[26] It should be noted, however, that states' ability to impose such influence over local land use regulations may be limited by home rule, which grants municipalities the ability to pass laws to govern themselves with respect to issues not expressly prescribed by the state in its constitution. New York State is one such home rule state.

Finally, the article's findings—and the scholarly literature—suggest that improving equitable access to resources and bettering future outcomes for low-income families cannot be achieved solely through increasing access to affordable housing in well-resourced neighborhoods. If, as the article's findings suggest, resources are inequitably distributed in well-resourced areas, and if, as the literature suggests, living in a well-resourced area does not guarantee access to opportunity, efforts to develop more affordable housing in well-resourced areas must be complemented with other

[25] See https://hcr.ny.gov/system/files/documents/2023/05/2023-fair-housing-matters-ny-public-cmt-draft-june-16-deadline.pdf.

[26] With regard to land use reforms, Oregon requires all cities of more than 10,000 residents to allow two- to four-unit homes on all residentially zoned lots (Oregon Department of Land Conservation and Development, 2020). With regard to builder remedies, Massachusetts has passed laws that offer expedited permitting for sustainable development in designated sustainable growth areas (Galante, Reid, and Weinstein-Carnes, 2016); California has restricted the number of procedural steps that local governments could require developers to go through, among other policies (California Legislative Information, 2020).

strategies to address equity. In New York State, HCR employs mobility counseling programs[27] to help families move to well-resourced areas and to help connect them with the resources necessary to meet their needs.

Furthermore, investments in affordable housing in well-resourced areas should not come at the expense of neighborhoods with fewer resources. In other words, states must also continue to invest in these areas, but with a concerted effort to ensure that those investments improve equity and opportunity for new and existing residents. Indeed, as mentioned previously, the LIHTC program also provides financing for developments in high-poverty areas, and researchers have found that LIHTC developments can benefit low-income neighborhoods (e.g., Schwartz, 2010). The IRS specified in 2016 that LIHTC developments in high-poverty areas must contribute to a concerted community revitalization plan,[28] and New York State's QAP specifically awards points to LIHTC projects that complement neighborhood revitalization efforts that "seek to fundamentally improve the quality of life and opportunities for neighborhood residents."[29] Strategies of this nature can help more low-income households find homes in well-resourced areas while also facilitating improvements to neighborhoods that have historically experienced long-term disinvestment and a shortage of adequate housing. The scoring tool presented in this article helps policymakers navigate their options for achieving those objectives and affirmatively furthering fair housing.

Appendix

Description of Equity and Diversity Metrics

High- to Low-Wage Workers Ratio

The ratio for high- to low-wage workers was calculated by dividing the number of jobs with earnings greater than $3,333/month by the number of jobs with earnings of $1,250/month or less in each tract.

White to Non-White workers Ratio

The ratio of White to non-White workers was calculated by dividing the number of jobs held by White people by the number of jobs held by non-White people. The variable for number of jobs held by non-White people was generated by subtracting the number of jobs held by White people from the total number of jobs.

Male to Female Workers Ratio

The ratio of male to female workers was calculated by dividing the number of jobs held by males by the number of jobs held by females.

[27] See https://hcr.ny.gov/mobility.

[28] See Rev. Rul. 2016-77, 2016-52 I.R.B. 914 (https://www.irs.gov/irb/2016-52_IRB).

[29] See https://hcr.ny.gov/system/files/documents/2021/05/qap-9-lihtc-part-2040.1-2040.13.pdf.

Top to Bottom Income Ratio

The ratio of top to bottom earners was calculated by splitting earners into two groups. Top earners were those with a household income of $100,000 or more. Bottom earners were those with a household income of $25,000 or less. The ratio between the groups was created by dividing the number of top earners by the number of bottom earners.

Herfindahl–Hirschman Index

The Herfindahl–Hirschman Index describes the mix of jobs from various industries within each tract. The industry mix in a tract of residence was calculated by determining the proportion each North American Industry Classification System (NAICS) sector makes up in the total number of jobs in a tract. The NAICS sectors are as follows: 11 (Agriculture, Forestry, Fishing and Hunting), 21 (Mining, Quarrying, and Oil and Gas Extraction), 22 (Utilities), 23 (Construction), 31-31 (Manufacturing), 42 (Wholesale Trade), 44-45 (Retail Trade), 48-49 (Transportation and Warehousing), 51 (Information), 52 (Finance and Insurance), 53 (Real Estate and Rental and Leasing), 54 (Professional, Scientific, and Technical Services), 55 (Management of Companies and Enterprises), 56 (Administrative and Support and Waste Management and Remediation Services), 61 (Educational Services), 62 (Health Care and Social Assistance), 71 (Arts, Entertainment, and Recreation), 72 (Accommodation and Food Services), 81 (Other Services [except Public Administration]), and 92 (Public Administration). The Herfindahl–Hirschman Index of industry mix was then calculated by summing the squares of the portion of jobs each sector makes up.

Homeowner to Renter Ratio

The homeowner to renter ratio was calculated by dividing the number of homeowners by the number of renters.

Single to Multifamily Ratio

The single to multifamily ratio was calculated by dividing the number of residences with one detached unit by the number of multifamily residences. The variable for multifamily residences was created by summing the number of structures with 10 or more units in each tract.

Rent to Mortgage Ratio (Housing Cost Burdens)

The ratio between owner and renter housing cost burdens was calculated by dividing median gross rent by median mortgage cost in each tract.

High to Low Home Value Ratio

The ratio for top to bottom home value was calculated by dividing the number of high-value owner-occupied units by the number of low-value owner-occupied units. High-value units were determined by summing the number of owner-occupied units worth at least $500,000 in each tract. Low-value units were determined by summing the number of owner-occupied units with a maximum value of $149,999 in each tract.

Exhibit A1

Descriptive Statistics for Equity and Diversity Metrics

	Min	Max	Median	Mean	N	SD
High- to Low-Wage Worker Ratio	0.40	8.9	2.18	2.35	5,270	1.09
White to Non-White Worker Ratio	0.11	84.1	3.30	7.23	5,270	9.45
Male to Female Worker Ratio	0.53	1.7	0.94	0.94	5,270	0.13
Top to Bottom Earner Ratio (Count)	0.00	411.0	2.42	4.93	5,270	0.13
Herfindahl–Hirschman Index of Industry Mix	0.07	0.3	0.10	0.11	5,270	0.02
Owner to Renter Ratio	0.00	265.5	1.54	4.52	5,227	11.36
Single-Family to Multifamily Ratio	0.00	2,329.0	2.36	24.49	4,313	86.29
Rent to Mortgage Ratio (Median)	0.14	7.4	0.63	0.66	4,211	0.23
High to Low Home Value Ratio	0.00	1,078.0	1.57	14.16	4,288	35.39

SD = Standard Deviation.

Sources: Authors' analysis of the following data sources: American Community Survey 5-year estimates, 2021; Longitudinal Employer-Household Dynamics, 2020

Acknowledgments

The authors wish to thank agency staff at New York State Homes and Community Renewal for ongoing support and feedback.

Authors

Pooya Ghorbani, Courtney Wolf, Ben Wetzler, Simon McDonnell, Bobbetta Davis, and Parker Pence are members of the Office of Research and Strategic Analysis (ORSA) at New York State Homes and Community Renewal (HCR).

References

Acevedo-Garcia, Dolores, Nancy McArdle, Erin F. Hardy, Unda I. Crisan, Bethany Romano, David Norris, Mikyung Baek, and Jason Reece. 2014. "The Child Opportunity Index: Improving Collaboration Between Community Development and Public Health," *Health Affairs* 33 (11): 1948–1957.

Asquith, Brian, Evan Mast, and Davin Reed. 2019. Supply Shock Versus Demand Shock: The Local Effects of New Housing in Low-Income Areas. Working paper 19-316. Kalamazoo, MI: W.E. Upjohn Institute for Employment Research.

Bassett, Edward M. 1922. *Zoning*. New York: National Municipal League.

Been, Vicky, Ingrid Gould Ellen, and Katherine O'Regan. 2019. "Supply Skepticism: Housing Supply and Affordability," *Housing Policy Debate* 29 (1): 25–40.

Been, Vicky, Jowiqh Madar, and Sunib McDonnell. 2014. "Urban Land-Use Regulation: Are Homevoters Overtaking the Growth Machine?" *Journal of Empirical Legal Studies* 11 (2): 227–265.

Bergman, Peter, Raj Chetty, Stefanie DeLuca, Nathaniel Hendren, Lawrence F. Katz, and Christopher Palmer. 2019. Creating Moves to Opportunity: Experimental Evidence on Barriers to Neighborhood Choice. National Bureau of Economic Research Working Paper No. 26164. Cambridge: MA: National Bureau of Economic Research.

California Legislative Information. 2020. SB-330 Housing Crisis Act of 2019. https://leginfo.legislature.ca.gov/faces/billTextClient.xhtml?bill_id=201920200SB330&search_keywords=housing.

Capps, Kriston. 2023. "Developers Forecast Major Affordable Housing Drought in 2025." Bloomberg. https://www.bloomberg.com/news/articles/2023-07-19/affordable-housing-shortage-looms-amid-inflation-high-construction-costs?cmpid=BBD071923_CITYLAB&utm_medium=email&utm_source=newsletter&utm_term=230719&utm_campaign=citylabdaily.

Chetty, Raj, John N. Friedman, Nathaniel Hendren, Maggie R. Jones, and Sonya R. Porter. 2018. The Opportunity Atlas: Mapping the Childhood Roots of Social Mobility. Working paper No. w25147. Cambridge, MA: National Bureau of Economic Research.

Chetty, Raj, John N. Friedman, Nathaniel Hilger, Emmanuel Saez, Diane Whitmore Schanzenbach, and Danny Yagan. 2011. "How Does Your Kindergarten Classroom Affect Your Earnings? Evidence From Project STAR," *The Quarterly Journal of Economics* 126 (4): 1593–1660.

Chetty, Raj, and Nathaniel Hendren. 2015. The Impacts of Neighborhoods on Intergenerational Mobility: Childhood Exposure Effects and County-Level Estimates. NBER Working Paper. Cambridge, MA: National Bureau of Economic Research.

Chetty, Raj, Nathaniel Hendren, and Lawrence F. Katz. 2016. "The Effects of Exposure to Better Neighborhoods on Children: New Evidence From the Moving to Opportunity Experiment," *American Economic Review* 106 (4): 855–902.

Chetty, Raj, Matthew O. Jackson, Theresa Kuchler, Johannes Stroebel, Nathaniel Hendren, Robert B. Fluegge, Sara Gong, Federico Gonzalez, Armelle Grondin, Matthew Jacob, Drew Johnston, Martin Koenen, Eduardo Laguna-Muggenburg, Florian Mudekereza, Tom Rutter, Nicolaj Thor, Wilbur Townsend, Ruby Zhang, Mike Baily, Pablo Barberá, Monica Bhole, and Nils Wernerfelt. 2022. "Social Capital I: Measurement and Associations With Economic Mobility," *Nature* 608 (7921): 108–121.

Clancy, P. E. 1990. "Tax Incentives and Federal Housing Programs: Proposed Principles for the 1980s." In *Building Foundations: Housing and Federal Policy*, edited by D. DiPasquale & L.C. Keyes. Philadelphia: University of Pennsylvania Press.

Climaco, Carissa, Meryl Finkel, Bulbul Kaul, Ken Lam, and Chris Roger. 2009. Updating the Low-Income Housing Tax Credit (LIHTC) Database: Projects Placed in Service Through 2006. Abt. Associates Inc. https://www.huduser.gov/portal/datasets/lihtc/report9506.pdf.

de Souza Briggs, Xavier, Susan J. Popkin, and John Goering., 2010. *Moving to Opportunity: The story of an American Experiment to Fight Ghetto Poverty*. Oxford University Press.

Ellen, Ingrid Gould, Keren Mertens Horn, and Yiwen Kuai. 2018. "Gateway to Opportunity? Disparities in Neighborhood Conditions Among Low-Income Housing Tax Credit Residents," *Housing Policy Debate* 28: 572–591.

Ellickson, Robert C. 2022. *America's Frozen Neighborhoods: The Abuse of Zoning*. New Haven, CT: Yale University Press.

Elmendorf, Christopher S. 2022. *A Primer on California's "Builder's Remedy" for Housing-Element Noncompliance*. Los Angeles: The UCLA Lewis Center for Regional Policy Studies.

Eriksen, Michael D. 2009. "The Market Price of Low-Income Housing Tax Credits," *Journal of Urban Economics* 66: 141–149.

Eriksen, Michael D., and Bree J. Lang. 2018. *Overview and Proposed Reforms of the Low-Income Housing Tax Credit Program*. Cincinnati, OH: University of Cincinnati Lindner College of Business Research Paper.

Fischel, William A. 2005. *The Homevoter Hypothesis: How Home Values Influence Local Government Taxation, School Finance, and Land-Use Policies*. Cambridge, MA: Harvard University Press.

———. 2004. "An Economic History of Zoning and a Cure for its Exclusionary Effects," *Urban Studies* 41 (2): 317–340.

Fischler, Raphaël. 1998. "Toward a Genealogy of Planning: Zoning and the Welfare State." *Planning Perspectives* 13 (4): 389–410.

Galante, Carol, Carolina Reid, and Ashley Weinstein-Carnes. 2016. *Borrowing Innovation, Achieving Affordability: What We Can Learn From Massachusetts Chapter 40B*. Berkeley, CA: Terner Center for Housing Innovation, University of California Berkeley.

Glaeser, Edward L., and Joseph Gyourko. 2002. The Impact of Zoning on Housing Affordability. NBER Working Paper No. 8835. Cambridge, MA: National Bureau of Economic Research.

Godinez-Puig, Luisa, Gabriella Garriga, and Yonah Freemark. 2023. *Tracing the Money*. Washington, DC: Urban Institute.

Haar, Charles M., and Jerold S. Kayden. 1989. *Zoning and the American Dream: Promises Still to Keep*. Chicago, IL: Planners Press.

Harris, Elizabeth, and Ford Fessenden. 2015. "'Opt-Out' Becomes Anti-Test Rallying Cry in New York State." *The New York Times*. https://www.nytimes.com/2015/05/21/nyregion/opt-out-movement-against-common-core-testing-grows-in-new-york-state.html.

Hickey, Robert, Lisa Sturtevant, and Emily Thaden. 2014. Achieving Lasting Affordability Through Inclusionary Housing. Working paper. Cambridge, MA: Lincoln Institute of Land Policy.

Hirt, Sonia A. 2014. *Zoned in the USA: The Origins and Implications of American Land-Use Regulation*. Ithaca, NY: Cornell University Press.

Jacobs, Jane. 1961. *The Death and Life of Great American Cities*. New York: Random House.

Jencks, Christopher, Susan B. Mayer. 1990. "The Social Consequences of Growing Up in a Poor Neighborhood." In *Inner-City Poverty in the United States*, edited by L.E. Lynn, M.F.H. McGeary. Washington, DC: National Academy Press: 111–186.

Lens, Michael. 2014. "Employment Accessibility Among Housing Subsidy Recipients," *Housing Policy Debate* 24: 671–691.

Li, Xiaodi. 2022. "Do New Housing Units in Your Backyard Raise Your Rents?" *Journal of Economic Geography* 22 (6): 1309–1352.

Mast, Evan. 2023. "JUE Insight: The Effect of New Market-Rate Housing Construction on the Low-Income Housing Market," *Journal of Urban Economics* 133 (103383).

McClure, Kirk, and Alex F. Schwartz. 2021. "Neighbourhood Opportunity, Racial Segregation, and the Low-Income Housing Tax Credit Program in the United States," *Housing Studies* 38: 1459–1481.

Micklow, A. 2008. The Gender Implications of Euclidian Zoning. Unpublished thesis. Blacksburg, VA: Virginia Tech.

New York State Education Department (NYSED). 2022. *New York State Department of Education, Note on State Report Card 2020-2021*. https://data.nysed.gov/essa.php?year=2021&state=yes.

New York State Division of Housing and Community Renewal. 2022. *Request for Proposals: Multifamily Programs HCR Multifamily Finance 9% LIHTC RFP*. https://hcr.ny.gov/system/files/documents/2022/10/rev.10.3.2022.fall-2022-multifamily-finance-9-rfp.pdf.

———. 2016. *Analysis of Impediments to Fair Housing Choice*. https://hcr.ny.gov/system/files/documents/2023/05/2023-fair-housing-matters-ny-public-cmt-draft-june-16-deadline.pdf.

New York State Division of the Budget. 2022. *FY 2023 Enacted Budget Financial Plan*. https://www.budget.ny.gov/pubs/archive/fy23/en/fy23en-fp.pdf.

Oregon Department of Land Conservation and Development. 2020. Land Conservation and Development Commission Adopts Middle Housing Rules to Boost Housing Choice and Supply. https://www.oregon.gov/lcd/NN/Press%20Releases/20201210_Large_Cities_Middle_Housing_Rules_Adopted.pdf.

Pollakowski, Henry O., Daniel H. Weinberg, Fredrik Andersson, John C. Haltiwanger, Giordano Palloni, and Mark J. Kutzbach. 2022. "Childhood Housing and Adult Outcomes: A Between-Siblings Analysis of Housing Vouchers and Public Housing," *American Economic Journal: Economic Policy* 14 (3): 235–272.

Quigley, John M., and Steven Raphael. 2004. "Is Housing Unaffordable? Why Isn't it More Affordable?" *Journal of Economic Perspectives* 18 (1): 191–214.

Reid, Carolina K. 2019. "Rethinking Opportunity in the Siting of Affordable Housing in California: Resident Perspectives on the Low-Income Housing Tax Credit," *Housing Policy Debate* 67: 279–292.

Reid, Carolina K., Carol Galante, and Ashley F. Weinstein-Carnes. 2017. "Addressing California's Housing Shortage: Lessons from Massachusetts Chapter 40B," *Journal of Affordable Housing & Community Development Law* 25 (2): 241–274.

Saiz, Albert. 2010. "The Geographic Determinants of Housing Supply," *The Quarterly Journal of Economics* 125 (3): 1253–1296.

Schuetz, Jenny. 2022. *Fixer-Upper: How to Repair America's Broken Housing Systems*. Washington, DC: Brookings Institution Press.

Schwartz, Alex F. 2010. *Housing Policy in the United States*. Second Edition. New York: Routledge.

Schwartz, Alex, and Edwin Meléndez. 2008. "After Year 15: Challenges to the Preservation of Housing Financed with Low-Income Housing Tax Credits," *Housing Policy Debate* 19 (2): 261–294.

Stacy, Christina Plerhoples, Christopher Davis, Yonah Freemark, Lydia Lo, Graham MacDonald, Vivian Zheng, and Rolf Pendall. 2023. "Land-Use Reforms and Housing Costs: Does Allowing for Increased Density Lead to Greater Affordability?" *Urban Studies* 60 (14): 2919–2940.

Stegman, Michael A. 1991. "The Excessive Costs of Creative Finance: Growing Inefficiencies in the Production of Low-Income Housing," *Housing Policy Debate* 2 (2): 357–373.

Tolley, George S. 1974. "The Welfare Economics of City Bigness," *Journal of Urban Economics* 1 (3): 324–345.

Turner, Margery Austin. 2011. "Moving to Opportunity – or Not?" Urban Wire. The Urban Institute. https://www.urban.org/urban-wire/moving-opportunity-or-not.

U.S. Department of Housing and Urban Development (HUD). 2023. *The U.S. Department of Housing and Urban Development. Low-Income Housing Tax Credit*. https://www.huduser.gov/portal/datasets/lihtc/property.html.

Zahirovich-Herbert, Velma, and Karen M. Gibler. 2014. "The Effect of New Residential Construction on Housing Prices," *Journal of Housing Economics* 26: 1–18.

Using Administrative Data Linkage to Drive Homelessness Policy: Experiences From Wales

Ian Thomas
Administrative Data Research Wales, Cardiff University

Peter Mackie
Cardiff University

Abstract

This article profiles Administrative Data Research Wales (ADR-W) and its use of data linkage to support homelessness policy and practice in Wales, United Kingdom. Despite having a national integrated data system for nearly a decade, Wales lacks the capacity—and, more important, the homelessness administrative data—necessary to engage in linkage-based research. However, the formation of ADR-W, with a remit to make better use of public-sector data, has instigated a rapid shift in the use of administrative data linkage in Wales. This article introduces the ADR-W, situating it as part of a wider turn to administrative data in the United Kingdom before providing an overview of the operation of the integrated data system ADR-W uses to conduct research—the Secure Anonymised Information Linkage Databank. This article offers insights into some of the homelessness research ADR-W conducts, highlighting key policy-relevant findings—including the effectiveness of the COVID-19 response. This article also critically reflects on some of the challenges with the current homelessness administrative data landscape in Wales, concluding with a note on its future direction.

Introduction

Homelessness represents a violation of human rights, being the deprivation of adequate housing, privacy, and security (United Nations, 2016). Accordingly, the government of Wales has committed to creating a nation in which homelessness is rare and, if people become homeless, to ensure it is a brief and nonrecurrent experience (Welsh Government, 2021a). What makes Wales—and the United Kingdom—unique internationally is that, since 1977, local authorities have had a legal duty to ensure that accommodation is made available to certain "priority need" households experiencing homelessness: families with dependent children (Fitzpatrick and Davies, 2021). Subsequent amendments to homelessness legislation in Wales that came into force in 2015 increased the legal duty to provide assistance to prevent homelessness (Mackie, Thomas, and Bibbings, 2017). Most recent data suggest that of the estimated 1.4 million households in Wales in 2022 and 2023, roughly 9,250 households were assisted to prevent homelessness, and 12,540 homeless households were assisted (StatsWales, 2023).

Although policymakers and practitioners in Wales base their decisionmaking on up-to-date evidence, quantitative research on homelessness lags behind that of other nations, particularly the United States (Culhane, Fitzpatrick, and Treglia, 2020). U.K. homelessness research has historically leaned more heavily toward qualitative methods, whereas the United States has stronger research links to psychological and health sciences and, therefore, a greater affinity for quantitative methods (Fitzpatrick and Christian, 2006). When quantitative research does take place in Wales, United Kingdom, it is largely based on small-scale nonrandom surveys that charities fund for political advocacy purposes—rather than larger-scale studies with greater levels of generalizability beyond particular at-risk subgroups within the homeless population (Pleace and Quilgars, 2003). Apart from surveys, administrative data enable interesting opportunities to expand the repertoire of quantitative homelessness research in Wales.

Being the "data exhaust" of day-to-day processes, administrative data can provide insight into the experiences of people whom housing and homelessness services assist (Hand, 2018). Furthermore, by linking different administrative datasets, researchers can gain insight into people's interactions across multiple systems, that is, housing, health, and education (Culhane, 2016). During the past decade, devolved governments and the research community in the United Kingdom have shown increased interest in administrative data and their linkage (Elias, 2018).

An Administrative Data Taskforce report in 2012 recommended the formation of a U.K.-wide network to facilitate access to and linkage of administrative data (ESRC, 2012), leading to the formation of the Administrative Data Research Network (ADRN) in 2014. Over time, the ADRN has transitioned into Administrative Data Research (ADR) United Kingdom, shifting its remit from assisting the research community access administrative data to directly engaging in policy-relevant research using administrative data and its linkage (Gordon, 2020). Each of the U.K. nations has its own national ADR center that generates evidence relevant to the specific national context. This article relates to the ADR Wales (ADR-W) center. The ADR centers each adopt a different infrastructure for storing and linking together administrative data. ADR-W uses the Secure Anonymised Information Linkage (SAIL) Databank (Ford et al., 2009).

Data Linkage Infrastructure in Wales: The SAIL Databank

Initially piloted in 2006, the SAIL Databank acts as an "integrated data system" for Wales (Zanti et al., 2022), storing de-identified yet linkable individual-level data and facilitating access to those data for research purposes. When data are ingested into SAIL, they undergo de-identification, whereby personal data are replaced with an identification number unique to each person in Wales. The de-identification process involves matching to a population spine created when people register their addresses with general practitioners or family doctors in Wales (Lyons et al., 2009).

Matching can be achieved deterministically using national healthcare numbers or exact matching on name, date of birth, gender, and postcode. Alternatively, matching is achieved probabilistically using combinations of name, date of birth, gender, and postcode. Once matched to the population address spine, the person's national healthcare number is extracted and forms the basis of his or her unique identifier. That identifier, also known as an Anonymised Linkage Field, links information about the same person across SAIL data. As an added layer of privacy protection, a "trusted third party" de-identifies the data so that SAIL cannot see personal data along with the "clinical information" relating to people's service interactions.

The original basis for the SAIL Databank was the study of population health. Therefore, SAIL holds a comprehensive array of healthcare information for the population of Wales, from primary care and family medicine to hospitalizations. Over time, data sources have diversified, and SAIL currently holds a range of other datasets, including substance misuse treatment service data, education outcomes (from schools to universities), census records, and criminal and civil court records. Researchers can apply to the SAIL Information Governance Review Panel to use those data, with projects assessed on their ability to generate new knowledge of scientific and practical value—that is, research that has public benefit. As a condition of accessing data in SAIL, researchers must undergo training in information governance to be considered "safe researchers." Once a research team receives approvals and meets the conditions of access, the team may access data within a secure virtual environment. To ensure that outputs are "safe," any analysis requested from the virtual environment undergoes disclosure control checks to ensure that individuals cannot be identified or inferred from outputs.

In addition to having relatively streamlined access to de-identified data already in the SAIL Databank, researchers can also upload data they possess to the SAIL environment. The ability to bring data into SAIL can enable, among other things, the use of data linkage to obtain routinely collected data for participants involved in housing and homelessness interventions. For example, as part of the PHaCT randomized control trial of a critical time intervention with prison leavers facing homelessness (Lewsey, 2023), personal data related to trial participants are imported into SAIL to allow the trial team to extract their health records—with the aim of comparing health outcomes between trial arms. Outside this novel potential use of SAIL to conduct evaluative research, the SAIL infrastructure enables linkage between homelessness and other data sources. However, the limited availability of individual- and case-level linkable homelessness data in Wales has proved a challenge for ADR-W.

The State of Homelessness Administrative Data (Linkage) in Wales

Despite legal obligations placed on local authorities to help prevent homelessness and assist households experiencing homelessness, no mandate exists to collect case-level data relating to households. By contrast, Continuums of Care in the United States must collect "universal data elements" on people and households accessing services funding by the U.S. Department of Housing and Urban Development. However, local authorities in Wales submit aggregate information to the Welsh government for monitoring purposes and, as such, collect similar data on individual cases to complete these "aggregate returns." Although without a top-down mandate to collect the same data in the same format, local authorities have designed divergent data collections. Therefore, the statutory homelessness data landscape in Wales is fractured and inconsistent.

Without centralized, case-level national data collection for Wales, access to statutory homelessness data is possible only through negotiations with each of the 22 local authority housing services in Wales. However, in 2018, the ADRN obtained data from a single local authority housing service, forming the basis for ADR-W's pioneering research program on homelessness. Several studies using those data have centered on the theme of severe and multiple disadvantages, exploring the interactions of people experiencing homelessness with other public services, including health, substance misuse, and the police (Browne Gott, 2019; Thomas, 2021). A second research theme has been the educational experiences of children in homeless families. This program of work proved particularly challenging, because the available local authority homelessness data contain personal information only for heads of households, not for their children. Outputs from this work focused on attainment and absenteeism and found that becoming homeless or being at risk of homelessness was associated with a 7-percent increase in total half-day sessions absent from school (Welsh Government, 2020, 2021b).

From initial conversations to data deposited in SAIL, acquiring this single local authorities' homelessness data took roughly 2 years. A large part of the delay in data acquisition was due to uncertainty within the local authority housing team of legal obligations when sharing data, combined with the necessary bureaucratic steps to enable the sharing of personal data—for example, undertaking impact assessments, creating legal documents, and finding appropriate people to authorize and take ownership of the process. Dealing with bureaucracy is a not insignificant aspect of any data share. In this instance, it placed an additional burden on an already overstretched housing service, contributing to delays.

Since 2021, the Welsh Government, in partnership with ADR-W, has been engaging in a pilot project to secure statutory homelessness data from additional local authorities in Wales. The intention of this pilot has been to acquire additional datasets to enable ADR-W to engage in far more nuanced analyses of underserved groups within the statutory homeless population, such as minority ethnic-headed households and children in families accessing housing services. As with ADRN's previous acquisition of local authority homelessness data, delays beset the pilot project. Although local authorities, government analysts, senior policymakers in the Welsh Government, and academics express a strong desire to engage in this work, progress has been slow.

Lacking up-to-date administrative data on people assisted under the statutory homelessness system in Wales, ADR-W employs innovative methods to identify people experiencing homelessness using existing population-level data collections within the SAIL Databank. Many of the health datasets in SAIL contain indicators for homelessness as a social determinant of ill health. For example, the International Classification of Diseases (ICD)-10 system classifying diseases in hospitalization data contains a code specifying homelessness. Also, as part of the substance misuse data collection, service users are asked directly about their housing support needs and provide specific examples of homelessness experiences to guide their clinicians' recording of different levels of housing need. Within that measure, "urgent housing problems" and "housing problems" cover instances of severe homelessness, ranging from living on the streets to sleeping in different accommodations each night.

Using those codes and measures, ADR-W conducted research during the COVID-19 pandemic that generated evidence of the potential protective effect of the Welsh Government crisis response, which included accommodating people experiencing homelessness in "suitable" temporary accommodations. Researchers found that the prevalence of SARS-CoV-2 infection among people experiencing homelessness was 5.0 percent compared with 6.9 percent among a nonhomeless matched comparator (Thomas and Mackie, 2021). A similar methodology demonstrated the ineffectiveness of the Welsh Government policy to prioritize people experiencing homelessness for the COVID-19 vaccination. Thomas and Mackie (2023) found that the incidence of the COVID-19 vaccination after 350 days of followup was 60.4 percent among people experiencing homelessness compared with 81.4 percent among a matched adult comparator. Despite those groundbreaking—at least in Wales—insights, the linkage and analysis of administrative data have limitations, particularly when researching homelessness (Thomas, 2020a; Thomas and Tweed, 2021).

Limitation of ADR-W's Homelessness Research

Foremost, ADR-W research on people assisted under the statutory homelessness system has drawn on data from a single authority and covers a relatively short period: people assisted between 2011 and 2017. Therefore, analysis using this dataset has been limited to more general questions related to the association between homelessness and people's outcomes and access to public services. An added complication is that the statutory data cover a period during which Welsh homelessness legislation was being reformed, meaning that the categorizations of how households were assessed under legislation were in flux. As a result, ADR-W has been unable to look at the outcomes of households assessed under the current Housing (Wales) Act 2014 in any complex multivariate way due to insufficient sample size and followup time.

When ADR-W has resorted to using nonhousing administrative data to identify people experiencing homelessness, generalizability to some larger homeless population may be limited. Health diagnosis codes related to homelessness are likely used in instances in which the clinician believes homelessness was a factor in the health event, which may not be the case in all healthcare interactions, leading to an underreporting of homelessness. Although the housing need measure within the substance misuse dataset applies to all people within the data and is, therefore, potentially less biased by recording behavior, the population in this dataset is quite obviously limited to people with higher needs seeking assistance with problematic substance use. As a result, people experiencing homelessness identified in nonhousing data sources are potentially at the more precarious end of the homelessness spectrum due to underlying healthcare and substance misuse-related issues. However, this "bias" has proven useful in the COVID-19 vaccine analysis, because it demonstrated that vaccine prioritization could not reach the most vulnerable homeless people in Wales.

As with all data linkage research, "missed matches" can be problematic and a source of bias if matched and unmatched people have differing characteristics (Harron et al., 2017). In the case of ADR-W research, missed matches occur either when a record could not be de-identified and assigned a unique identifier or when records are excluded due to low matching quality, that is, when unique identifiers are assigned probabilistically and researchers cannot be certain that it is the correct identifier. From ADR-W's experience of using different data sources, healthcare data in SAIL demonstrated far lower rates of missed matches (~5 percent missed) when compared with the local authority homelessness dataset (~25 percent missed). The data maturity of the organizations collecting data and their different information needs may contribute to those differences in match rates.

Although local authorities in Wales are developing a culture of data use, gaps and weaknesses in their data remain (Audit Wales, 2018). This level of use compares with that of the national health service in Wales, for which data are core to its functioning, leading to higher-quality data. Furthermore, national healthcare numbers are collected when people access healthcare-related settings, which then allows deterministic matching to the population spine used when de-identifying data. Without healthcare numbers, local authority homelessness data were de-identified using probabilistic methods, which, the authors suggest, was affected by poorer quality collection of personal information. As data linkage becomes mainstreamed as part of the data processing and evidence landscape in Wales, the authors hope that the data maturity of local authorities and housing and homelessness support services improves, reducing missed-match rates.

Future Directions

Regardless of the challenges, ADR-W is slowly developing a portfolio of research that demonstrates the potential use of data linkage to support evidence-driven homelessness policy and practice. For more than a decade the ADR-W team has consistently argued for the need for national individual-level data collection related to statutory homelessness in Wales (Mackie, Thomas, and Hodgson, 2012; Thomas, 2020b). Putting this situation in context, Wales is now the outlier of the devolved U.K. nations, with Northern Ireland, Scotland, and England all having individual-level homelessness collections. However, some promising developments are occurring in this area. The Ending Homelessness Action Plan for Wales commits to improving data (Welsh Government, 2021a), with the Welsh Government engaging ADR-W to scope out a new data collection system (Thomas, 2020b), drawing inspiration from other U.K. approaches. Rather than focusing only on better measurement of homelessness and its effects, such a system would enable the key opportunity of evaluating interventions to establish what works in ending and, more importantly, preventing homelessness in Wales.

Acknowledgments

This article reports findings from research that have made use of anonymized data held in the Secure Anonymised Information Linkage (SAIL) Databank. We would like to acknowledge all the data providers who made anonymized data available for research via the SAIL Databank. Responsibility for the interpretation of the information supplied by SAIL is the authors' alone. Linked datasets cannot be made available to researchers due to data-sharing agreements limiting their access to the research team. However, the individual datasets used in this research can be accessed via application to the SAIL Databank. This analysis was undertaken as part of the Administrative Data Research Centre Wales, funded by the Economic and Social Research Council (ES/W012227/1).

Authors

Ian Thomas is a researcher at Administrative Data Research Wales, Cardiff University School of Social Sciences. Peter Mackie is a professor at the Cardiff University School of Geography and Planning.

References

Audit Wales. 2018. The Maturity of Local Government in Use of Data. Cardiff, Wales: Auditor General for Wales.

Browne Gott, Hannah. 2019. "Using Administrative Data to Understand the Service Interactions of People Experiencing Homelessness," *International Journal of Population Data Science* 4 (3): 169. https://doi.org/10.23889/ijpds.v4i3.1333.

Culhane, Dennis. 2016. "The Potential of Linked Administrative Data for Advancing Homelessness Research and Policy," *European Journal of Homelessness* 10 (3): 109–126.

Culhane, Dennis, Suzanne Fitzpatrick, and Dan Treglia. 2020. "Contrasting Traditions in Homelessness Research Between the UK and US." In *Using Evidence to End Homelessness*, edited by Lígia Teixeira and James Cartwright. Bristol: Policy Press.

Economic and Social Research Council (ESRC). 2012. *The UK Administrative Data Research Network: Improving Access for Research and Policy*. Swindon, UK: ESRC.

Elias, Peter. 2018. "The UK Administrative Data Research Network: Its Genesis, Progress, and Future," *The ANNALS of the American Academy of Political and Social Science* 675 (1): 184–201. https://doi.org/10.1177/0002716217741279.

Fitzpatrick, Suzanne, and Julie Christian. 2006. "Comparing Homelessness Research in the US and Britain," *European Journal of Housing Policy* 6 (3): 313–333. https://doi.org/10.1080/14616710600973151.

Fitzpatrick, Suzanne, and Liz Davies. 2021. "The 'Ideal' Homelessness Law: Balancing 'Rights Centred' and 'Professional-Centred' Social Policy," *Journal of Social Welfare and Family Law* 43 (2): 175–197. https://doi.org/10.1080/09649069.2021.1917712.

Ford, David V., Kerina H. Jones, Jean-Philippe Verplancke, Ronan A. Lyons, Gareth John, Ginevra Brown, Caroline J. Brooks, Simon Thompson, Owen Bodger, Tony Couch, and Ken Leake. 2009. "The SAIL Databank: Building a National Architecture for e-Health Research and Evaluation," *BMC Health Services Research* 9: 157. https://doi.org/10.1186/1472-6963-9-157.

Gordon, Emma. 2020. "Administrative Data Research UK," *Patterns* 1 (1): 1–2. https://doi.org/10.1016%2Fj.patter.2020.100010.

Hand, David J. 2018. "Statistical Challenges of Administrative and Transaction Data," *Journal of the Royal Statistical Society Series A: Statistics in Society* 181 (3): 555–605. https://doi.org/10.1111/rssa.12315.

Harron, Katie, Chris Dibben, James Boyd, Anders Hjern, Mahmoud Azimaee, Mauricio L Barreto, and Harvey Goldstein. 2017. "Challenges in Administrative Data Linkage for Research," *Big Data & Society* 4 (2): 1–12. https://doi.org/10.1177/2053951717745678.

Lewsey, Jim. 2023. *Preventing Homelessness, Improving Health for People Leaving Prison*. London, UK: BioMed Central. https://www.isrctn.com/ISRCTN46969988.

Lyons, Ronan A., Kerina H. Jones, Gareth John, Caroline J. Brooks, Jean-Philippe Verplancke, David V. Ford, Ginevra Brown, and Ken Leake. 2009. "The SAIL Databank: Linking Multiple Health and Social Care Datasets," *BMC Medical Informatics and Decision Making* 9 (3). https://doi.org/10.1186/1472-6947-9-3.

Mackie, Peter, Ian Thomas, and Jennie Bibbings. 2017. "Homelessness Prevention: Reflecting on a Year of Pioneering Welsh Legislation in Practice," *European Journal of Homelessness* 11 (1): 81–107.

Mackie, Peter, Ian Thomas, and Kate Hodgson. 2012. *Impact Analysis of Homelessness Legislation in Wales: A Report to Inform the Review of Homelessness Legislation in Wales*. Cardiff, Wales: Welsh Assembly Government.

Pleace, Nicholas, and Deborah Quilgars. 2003. "Led Rather Than Leading? Research on Homelessness in Britain," *Journal of Community & Applied Social Psychology* 13: 187–196.

StatsWales. 2023. "Households for Which Assistance Has Been Provided by Outcome and Household Type." https://statswales.gov.wales/Catalogue/Housing/Homelessness/Statutory-Homelessness-Prevention-and-Relief.

Thomas, Ian. 2021. "Severe and Multiple Disadvantage," *Data Insight*, March. Administrative Data Research Wales. https://www.adruk.org/fileadmin/uploads/adruk/Documents/Data_Insights_Severe_and_multiple_disadvantage__March_2021_.pdf.

———. 2020a. "Analysing Administrative Data." In *Sage Research Methods: Foundations*, edited by Paul Atkinson, Sara Delamont, Alexandru Cernat, Joseph W. Sakshaug, and Richard A. Williams. https://doi.org/10.4135/9781526421036871241.

———. 2020b. *Feasibility Study Into a New Case Level Homelessness Data System for Wales*. Glasgow, Scotland: UK Collaborative Centre for Housing Evidence. https://housingevidence.ac.uk/publications/feasibility-study-into-a-new-case-level-homelessness-data-system-for-wales/.

Thomas, Ian, and Peter Mackie. 2023. "Assessing the Coverage and Timeliness of Coronavirus Vaccination Among People Experiencing Homelessness in Wales, UK: A Population-Level Data-Linkage Study," *BMC Public Health* 23: 1–8. https://doi.org/10.1186/s12889-023-16432-x.

———. 2021. "A Population Level Study of SARS-CoV-2 Prevalence Amongst People Experiencing Homelessness in Wales, UK," *International Journal of Population Data Science Special Issue* 5 (4): 1–12. https://doi.org/10.23889/ijpds.v5i4.1695.

Thomas, Ian, and Emily Tweed. 2021. "The Promises and Pitfalls of Administrative Data Linkage for Tackling Homelessness," *European Journal of Homelessness* 15 (3): 177–188.

United Nations. 2016. *Report of the Special Rapporteur on Adequate Housing as a Component of the Right to an Adequate Standard of Living, and on the Right to Non-Discrimination in This Context*, 31st Session of the Human Rights Council, Report A/HRC/31/54/Add.2. Geneva: United Nations. https://www.ohchr.org/en/hr-bodies/hrc/regular-sessions/session31/list-reports.

Welsh Government. 2021a. "Ending Homelessness in Wales: A High Level Action Plan 2021–2026." https://www.gov.wales/ending-homelessness-wales-high-level-action-plan-2021-2026.

———. 2021b. "Absence From School Amongst Children Living in Homeless Households." https://www.gov.wales/absence-school-amongst-children-living-homeless-households-html#:~:text=Absenteeism%20varied%20with%20academic%20year,total%20sessions%20absent%20from%20school.

———. 2020. "Exploratory Analysis of Education Outcomes of Children and Young People Living in Homeless Households." https://www.gov.wales/exploratory-analysis-education-outcomes-children-and-young-people-living-homeless-households.

Zanti, Sharon, Emily Berkowitz, Matthew Katz, Amy Hawn Nelson, T.C. Burnett, Dennis Culhane, and Yixi Zhou. 2022. "Leveraging Integrated Data for Program Evaluation: Recommendations from the Field," *Evaluation and Program Planning* 95: 102093. https://doi.org/10.1016/j.evalprogplan.2022.102093.

Ownership and Displacement

Assessing How Gentrification and Disinvestment-Related Market Pressures Drive the Loss of Small Multiunit Housing in Chicago Neighborhoods

Housing Speculation, Affordable Investments, and Tenant Outcomes in New York City

Evidence-Based Policymaking to Address the Affordable Housing Crisis: The Potential of Local Data

Improving Housing Policy with Neighborhood Data

Assessing How Gentrification and Disinvestment-Related Market Pressures Drive the Loss of Small Multiunit Housing in Chicago Neighborhoods

Sarah Duda
Geoff Smith
Yiwen Jiao
Institute for Housing Studies at DePaul University

Abstract

In the city of Chicago, two- to four-unit buildings play a crucial role in the housing stock. These buildings provide affordable rental housing, homeownership, and income-generating opportunities and make up a substantial portion of Chicago's communities of color housing supply. However, there are substantial concerns that the city's supply of these buildings is disappearing. For this research, the authors use historical parcel-level data from the Cook County Assessor's Office to analyze changes in the housing stock makeup of neighborhoods throughout Chicago, quantify the losses in two- to four-unit buildings, identify what happened to these properties, and assess the underlying real estate market factors behind the loss of these buildings.

The authors find that nearly every neighborhood in Chicago is losing two- to four-unit buildings, but the reasons for the losses vary. In higher-cost neighborhoods, shifting housing demand toward higher-income owner-occupants means that older, smaller rental properties are targeted for conversion to either high-cost single-family homes or condominium buildings. This type of market activity reduces the supply of lower-cost rental housing, potentially making these communities less affordable for modest-income renters and driving displacement pressures. In lower-cost neighborhoods with long histories of disinvestment, two- to four-unit buildings are commonly demolished, becoming vacant land. This type of market activity also leads to a loss of housing in these neighborhoods, presenting potential barriers to reinvestment.

These results indicate that there is an urgent need for interventions to protect two- to four-unit buildings and that these interventions must be informed by an understanding of the neighborhood housing market conditions driving this activity. This research adds to the literature by highlighting the critical role that two- to four-unit buildings play in housing affordability for both renters and homeowners and also profiling the spectrum of challenges they face.

Introduction

Cities in the United States face a housing affordability crisis that makes it difficult for modest-income renters to find affordable housing or prospective homebuyers to find starter homes (Herbert, 2023; NLIHC, 2023). Although the reasons behind this affordability crisis are multifaceted and complex, there is a growing awareness that small, multiunit properties with two to four units play an important role in the current housing supply by providing both affordable rental housing and homeownership opportunities (Garcia et al., 2022). However, market pressures on existing two- to four-unit buildings in various neighborhood contexts may pose challenges to the preservation of this unique and critical housing type. This article uses Chicago as a case study to highlight how shifting housing demand and underlying real estate market factors in both higher-cost and historically disinvested neighborhoods are driving the loss of these buildings.

The authors find that virtually every neighborhood in Chicago is losing two- to four-unit buildings, but the reasons for the losses vary. In higher-cost neighborhoods, shifting housing demand toward higher-income owner occupants means that older, smaller rental properties are targeted for conversion or demolition and replaced with either high-cost single-family homes or condominium buildings. This type of market activity reduces the broader supply of lower-cost rental housing and rental housing, potentially making these communities less affordable for modest-income renters and driving displacement pressures. In lower-cost neighborhoods with long histories of disinvestment, two- to four-unit buildings are commonly demolished, becoming vacant land. This type of market activity also leads to a loss of housing in these neighborhoods, limiting pathways to redevelopment and presenting potential barriers to community reinvestment (Parolek and Parolek, 2021).

Understanding this market context is critical to developing targeted policy interventions to preserve the stock of small rental buildings across neighborhoods. For example, proactive policies to preserve the existing affordable two- to four-unit rental stock in higher-cost neighborhoods are needed before this stock is lost to gentrification pressures. Meanwhile, in lower-cost areas, the loss of two to four units to deterioration and demolition highlights the need for investment in both the broader community and existing housing stock. This investment is needed to reverse the tide of long-term population loss, historic disinvestment, the ongoing legacies of the foreclosure crisis and Great Recession, and the economic impact of the recent COVID-19 pandemic.

Context

Largely built to provide homeownership opportunities for immigrant families in the early 20th century, two- to four-unit properties make up a significant portion of the housing stock in many older U.S. cities, including Boston, Buffalo, Chicago, Milwaukee, New Orleans, Newark, and St. Louis (U.S. Census Bureau, 2021). Two- to four-unit buildings often have a similar scale and form as single-family homes, meaning that their presence maintains the appearance of lower-density neighborhoods (Parolek, 2020: 15). However, these multiunit buildings also sustain an urban density, making small multifamily buildings a crucial part of pushes for denser and more walkable neighborhoods (Holeywell, 2016). Exhibit 1 shows an example of a block with two- to four-unit buildings in Chicago.

Exhibit 1

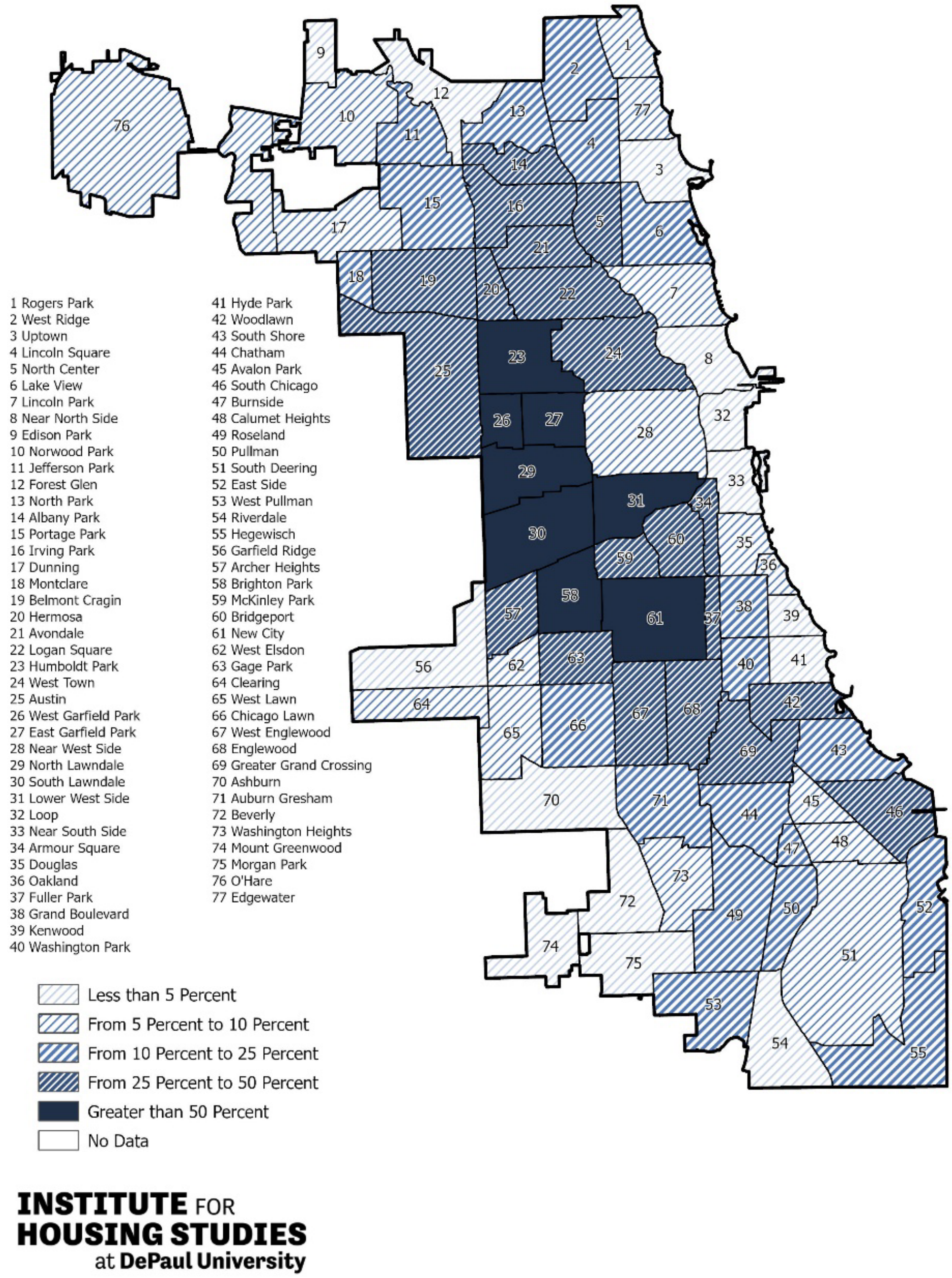
Chicago Block With Two- to Four-Unit Buildings

A block with two- to four-unit buildings in Chicago. Photo credit: Institute for Housing Studies at DePaul University.

In Chicago, buildings with two to four units are a critical component of the city's housing stock and play a unique role in providing both renter- and owner-occupied housing, constituting 26 percent of all residential units in the city (IHS, 2021a). Two- to four-unit buildings are also the most common type of rental housing in the city, making up more than 32 percent of Chicago's rental housing supply, accounting for more rental units than buildings with 5 to 49 or 50 units or greater (IHS, 2023a). Two- to four-unit properties are in most of the city's neighborhoods but are also highly geographically concentrated in certain communities. For example, as exhibit 2 illustrates, more than one-half of housing units in 8 of Chicago's 77 community areas, primarily on Chicago's South and West Sides, are in two- to four-unit buildings (IHS, 2021a).

Exhibit 2

Share of Residential Parcels in Two- to Four-Unit Properties in Chicago, Tax Year 2020



Sources: Cook County Assessor's Office; Institute for Housing Studies Data Clearinghouse

In Chicago and nationally, two- to four-unit buildings are a critical component of the supply of unsubsidized affordable rental housing for lower-income households, particularly for households and communities of color. According to national research from the University of Southern California and Enterprise Community Partners, rental units in two- to four-unit buildings tend to have the lowest rents of any building type and also serve renters with the lowest incomes (An et al., 2017). In Chicago, two- to four-unit buildings have the highest share of rental units with rents less than $900 (IHS, 2023b). Compared with larger multifamily rental properties, two- to four-unit buildings are much more likely to offer family-sized units, with roughly 40 percent of units in these properties containing three or more bedrooms (IHS, 2021a). In addition, two- to four-unit buildings are particularly important to the housing stock in communities of color and to residents of color citywide. Roughly 60 percent of Chicago's Latino-headed and 32 percent of African-American-headed renter households live in two- to four-unit buildings (IHS, 2021a). Nearly 46 percent of the total housing units in Chicago's majority-Latino census tracts and nearly 30 percent in predominantly African-American census tracts are in buildings with two to four units (IHS, 2021a). These local patterns are mirrored in national research. For example, the Urban Institute recently found that African-American and Latino renters make up a higher share of tenants in two- to four-unit buildings than any other type of rental property (Theodos et al., 2019).

Two- to four-unit buildings are unique because they are often owned by small, "mom and pop" landlords who may also occupy units as their homes while renting the remaining units. This arrangement provides affordable homeownership and wealth-building opportunities for the owner households, particularly for households and communities of color (Choi and Young, 2020). Nationally, Black or Latino landlords are more likely to own two- to four-unit buildings than any other size of rental building, and small rental buildings have been found to provide a pathway to homeownership for immigrant households (Choi and Young, 2020; Cornelissen, Hermann, and Whitney, 2023). In Chicago, roughly 54 percent of all two- to four-unit buildings had an active homeowner exemption in the 2019 tax year, suggesting that they are owner occupied. By comparison, when considering only majority-Latino neighborhoods, the share of two- to four-unit parcels with homeowner exemptions rises to more than 62 percent (IHS, 2021a). In addition, two- to four-unit properties have proven popular with recent homebuyers of color in Chicago. Between 2019 and 2021, 11.5 percent of all home purchase loans in Chicago were for two- to four-unit properties. However, the rates for African-American and Latino homebuyers were more than 24 percent and nearly 20 percent, respectively, compared with 6 percent for White homebuyers. These percentages illustrate that two- to four-unit buildings are a critical pathway to homeownership for borrowers of color in Chicago (IHS, 2023d).

Despite the importance of two- to four-unit buildings to Chicago renters and homeowners, the challenges facing this segment of the housing stock have been long term. Of all types of residential property in Chicago, two- to four-unit buildings were most affected by foreclosure in the years leading up to and following the Great Recession (IHS, 2021a). Between 2005 and 2019, nearly 30 percent of two- to four-unit parcels in Chicago were associated with at least one foreclosure filing. Properties in the city's predominantly African-American communities were hit hardest, with more than 47 percent of two- to four-unit parcels being associated with at least one foreclosure filing (IHS, 2021a).

In the years following the Great Recession, the city saw losses to the two- to four-unit rental stock despite overall growth in rental housing demand and overall rental supply. Between 2012 and 2021, the city lost nearly 12 percent of its rental units in two- to four-unit buildings during a period when the number of rental housing units in all other building types, besides single-family homes, grew (IHS, 2023a). As Chicago lost two- to four-unit rental buildings, it also lost lower-cost rental housing units. During the same period, Chicago saw a more than 15-percent decline in the number of rental units affordable to lower-income households, widening the gap between demand for affordable housing from lower-income renters and the supply of affordable rental units (IHS, 2023a). This correlation between lost two- to four-unit rentals and declines in Chicago's affordable rental supply highlights the key role that two- to four-unit buildings play in providing lower-cost rental housing. It is also consistent with national evidence that the loss of lower-cost rental housing in cities across the country is one of the biggest housing affordability challenges nationally (JCHS, 2023).

For many years, Chicago housing and community development stakeholders have led advocacy campaigns regarding the loss of two- to four-unit buildings citywide and in their neighborhoods (IHS, 2021b). In 2014, Communities United, a community-based organization working in Chicago's Albany Park neighborhood, began raising concerns that two- to four-unit buildings were being lost to foreclosure and acquired by investors for conversion into single-family homes (Shropshire, 2016). The Logan Square Neighborhood Association and Latin United Community Housing Association also raised concerns that single-family conversions were accelerating displacement pressure near The 606, a linear park system on the border of Humboldt Park and Logan Square built in 2015 (Black, 2020). In Chicago's South Side communities, such as Auburn Gresham, Chatham, Englewood, and Greater Grand Crossing, advocates seeking to leverage the two- to four-unit housing stock to attract and retain residents struggled to contend with a limited stock of these buildings after significant post-Great Recession demolition campaigns, severe tax delinquency, or the extensive rehabilitation needs of many remaining two- to four-unit buildings (Caine, 2022).

Analysis

Since 2012, the Institute for Housing Studies (IHS) has written extensively about the importance of two- to four-unit properties to Chicago's supply of unsubsidized affordable housing. The institute used reports from community partners and American Community Survey aggregated data to indicate that the stock was disappearing during the post-Great Recession housing market recovery. IHS staff worked also with local partners to identify the likely reasons for this loss, including pressures to convert two- to four-unit buildings to single-family homes and the loss of the stock through deterioration (IHS, 2019).

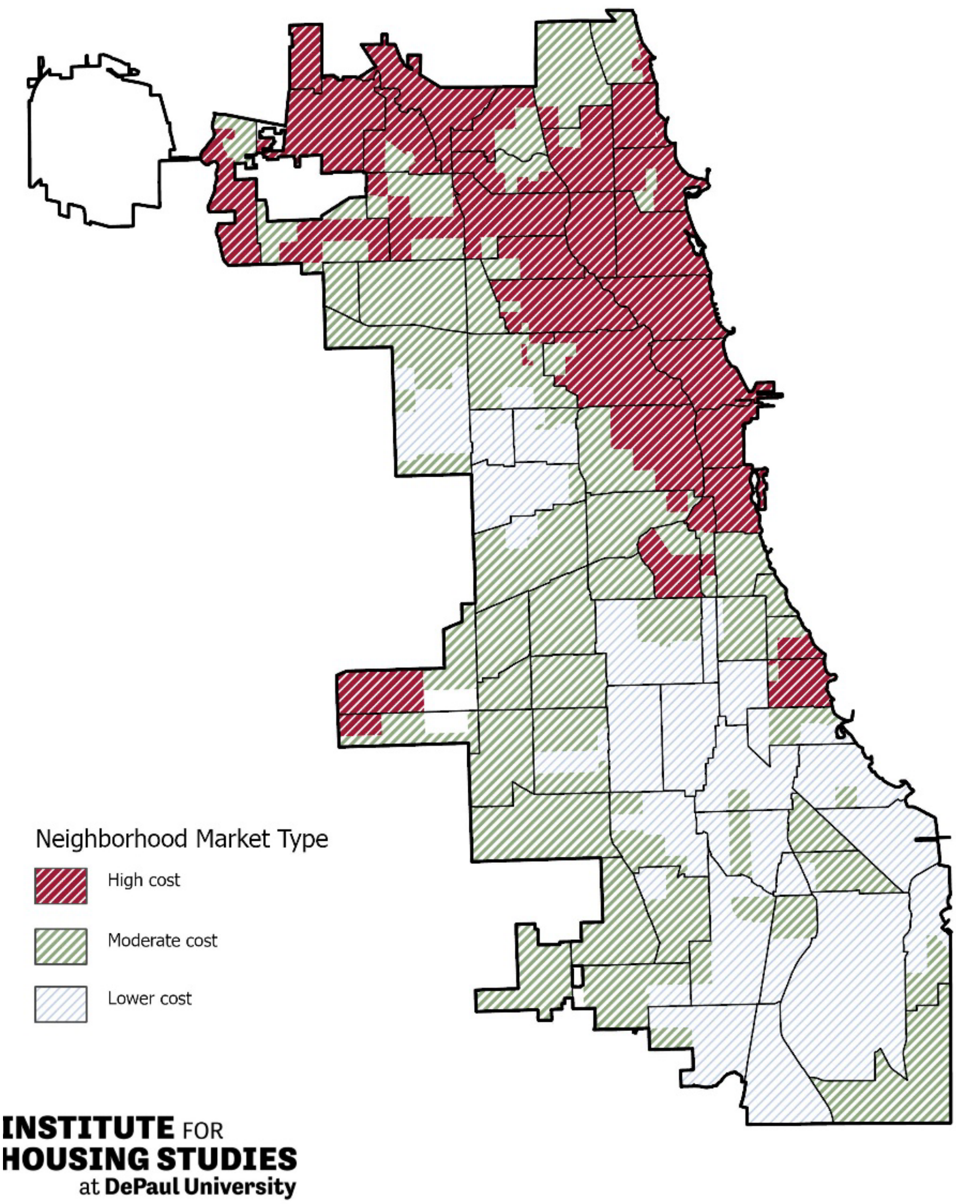
IHS's use of local, parcel-level administrative data to understand the nature and market context for this lost supply began in 2019 with data-focused technical assistance engagements with nonprofit community partners, representatives from the City of Chicago Department of Housing, and citywide aldermanic offices. For example, IHS built a dataset of former two- to four-unit properties with recent tax class changes to quantify the scale of conversion activity of two- to four-unit buildings to single-family homes in and around the North Center neighborhood of Chicago, testing initial results against walking surveys collected by residents.

This article's analysis uses citywide, historical parcel-level property assessment data from the Cook County Assessor's Office to analyze changes in the housing stock makeup of neighborhoods throughout Chicago and to quantify and categorize losses in two- to four-unit buildings. It identifies a group of properties that the Cook County Assessor's Office categorized as two- to four-unit buildings in 2013. It then traces those properties' minor tax class and other characteristics into 2019 to determine whether or not buildings remained two to four units and, if not, how the property tax class changed during the period (IHS, 2021d). This analysis focuses on the period from 2013 to 2019 to allow for two triannual property tax reassessments in 2015 and 2018 to ensure that the Cook County Assessor's Office had sufficient time to capture changes in the tax class through research, data collection, and appeals processes. IHS used iterations of this methodology and preliminary dataset for multiple technical assistance projects completed for neighborhood-based nonprofits and used these engagements to "ground-truth" the results and refine the methodology in this analysis.

This analysis also uses a lens of neighborhood market value to highlight the variation in the different ways that two- to four-unit buildings are lost in different market contexts. To build a typology of market types, the authors use parcel-level administrative data on one- to four-unit property sales activity from the Cook County Recorder of Deeds and geospatial techniques to derive a granular assessment of neighborhood-level prices in 2020 relative to surrounding areas and to the city of Chicago as a whole (IHS, 2017). Using these data, this analysis classifies census tracts on the basis of current market conditions (high, moderate, and lower cost) and the distribution of census tracts by price (IHS, 2021c). As exhibit 3 illustrates, high-, moderate-, and lower-cost neighborhoods in 2020 were highly geographically clustered, with high-cost areas concentrated on the North and Northwest Sides of the city; lower-cost areas concentrated on the South, Southwest, and Far South Sides of the city; and moderate-cost areas concentrated on the West Side and scattered in North and Southside neighborhoods citywide.

Exhibit 3

Neighborhood Housing Market Typology of One- to Four-Unit Property Sales Prices for the City of Chicago, 2020



Sources: Cook County Assessor's Office; Black Knight Financial Services; Institute for Housing Studies Data Clearinghouse

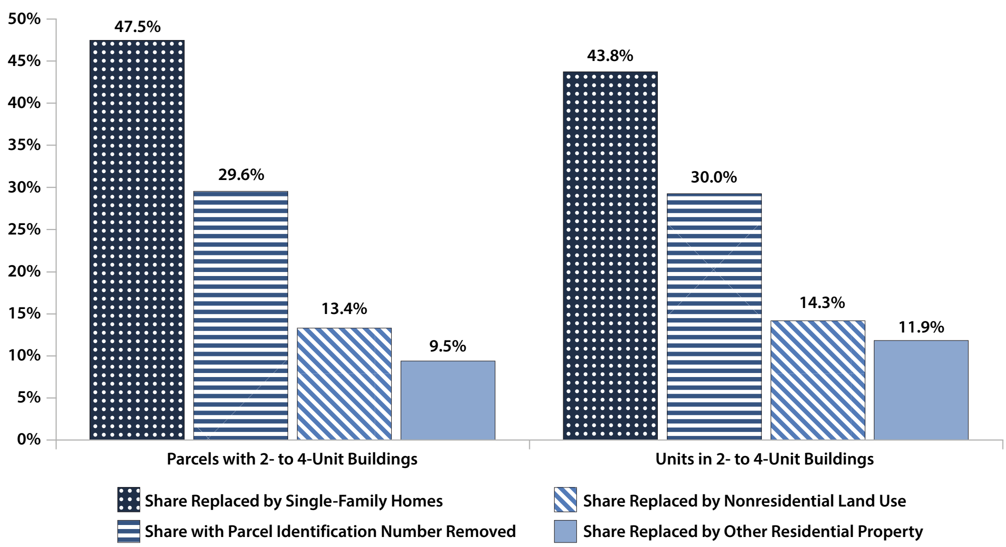
Key Findings

Chicago has lost two- to four-unit housing stock to a range of other types of residential properties and nonresidential land uses.

Since 2013, Chicago has lost more than 4,800 two- to four-unit buildings, representing 11,775 housing units and a loss of 4.2 percent of Chicago's stock of two- to four-unit parcels. Exhibit 4 illustrates that 47.5 percent of these lost buildings were replaced by single-family homes, either through conversion of the existing structure or demolition and new construction. In addition, 29.6 percent of lost buildings were replaced by nonresidential uses, such as vacant land. At the same time, 13.4 percent of lost two- to four-unit buildings had "disappeared" from the data by 2019.[1] Finally, 9.5 percent of lost two- to four-unit properties became some other residential type but not a single-family home. In this situation, many of the changes detected were the result of additional housing units being added to the building, such as a legal basement unit. Such a change would mean that the property became a five-or-more-unit building.

Exhibit 4

Minor Class Changes for Two- to Four-Unit Properties in the City of Chicago, 2013 to 2019



Sources: Cook County Assessor's Office; Institute for Housing Studies Data Clearinghouse

Although all neighborhood market types have lost a share of two- to four-unit buildings since 2013, high-cost areas lost the largest share of their 2013 stock.

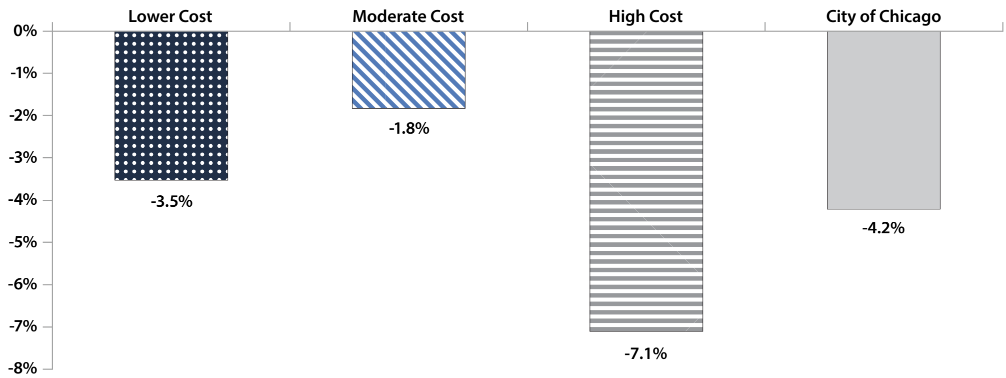
Nearly 61 percent of all lost two- to four-unit buildings were in high-cost areas, 23.9 percent were in lower-cost areas, and 15.3 percent were in moderate-cost areas. Exhibit 5 illustrates the share of

[1] The reasons a property may "disappear" from the data are numerous, but in many cases, it can be traced to the redevelopment of two- to four-unit properties into condominium buildings or a larger multifamily rental building. The new property has a new parcel identification number or is split into multiple parcel identification numbers. In these cases, because the property identification number changed, it impedes the type of analysis in this study.

the 2013 two- to four-unit building stock lost between 2013 and 2019 by neighborhood market type. It shows that high-cost neighborhoods lost 7.1 percent of the 2013 two- to four-unit building stock—the largest share of any market type. By comparison, 3.5 and 1.8 percent of the 2013 two- to four-unit building stock has been lost in lower- and moderate-cost communities, respectively, during the period.

Exhibit 5

Change in Two- to Four-Unit Parcels in the City of Chicago by Neighborhood Market Type, 2013 to 2019



Sources: Cook County Assessor's Office; Institute for Housing Studies Data Clearinghouse

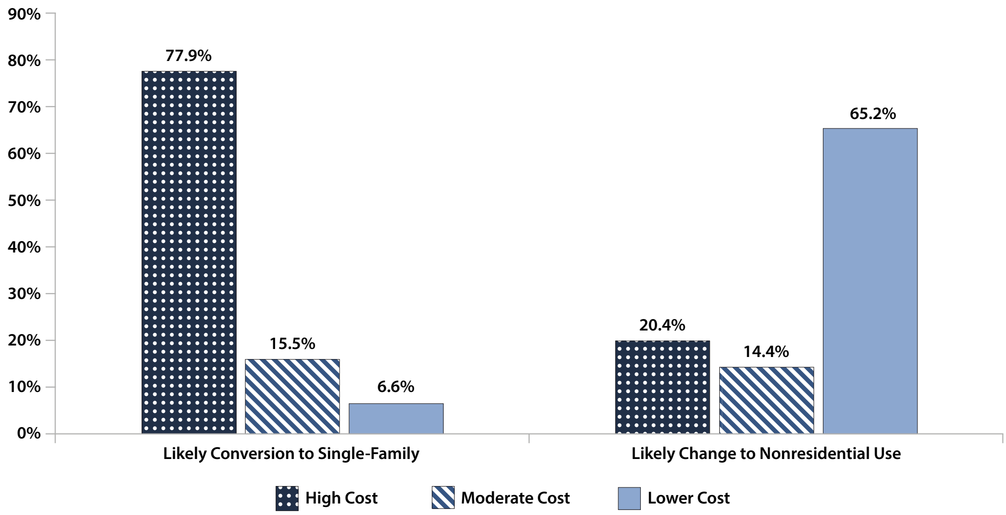
The nature of the lost two- to four-unit stock differs by neighborhood market condition.

The threats facing two- to four-unit buildings are largely related to a mismatch between market demand and the supply of certain types of housing in a neighborhood. For example, many higher-cost neighborhoods on Chicago's North Side have experienced recent increases in both higher-income households and families with children, indicating increased demand for housing from households that may want more space and have the means to afford expensive single-family homes (IHS, 2018b). At the same time, other neighborhoods are dealing with the effect of historic disinvestment, long-term population loss, and the continued legacy of the foreclosure crisis, particularly in lower-cost areas (IHS, 2018a).

These factors have led to very different pathways for the lost two- to four-unit stock in Chicago neighborhoods connected to patterns of gentrification, displacement, and disinvestment. Exhibit 6 illustrates that nearly 78 percent of the lost two- to four-unit buildings that single-family homes replaced were in high-cost markets, and more than 65 percent of all two- to four-unit buildings lost to nonresidential uses were in lower-cost neighborhoods. Exhibit 7 maps the total lost two- to four-unit stock by neighborhood market type. Exhibit 8 maps the lost two- to four-unit stock by type of loss and illustrates the distinct geographic patterns of this loss in Chicago neighborhoods. It shows that the loss of two- to four-unit buildings through single-family home replacement is largely a phenomenon in higher-cost areas of the city and pockets of moderate-cost areas near amenities, such as transit access, although the loss to nonresidential use is highly concentrated in lower-cost neighborhoods.

Exhibit 6

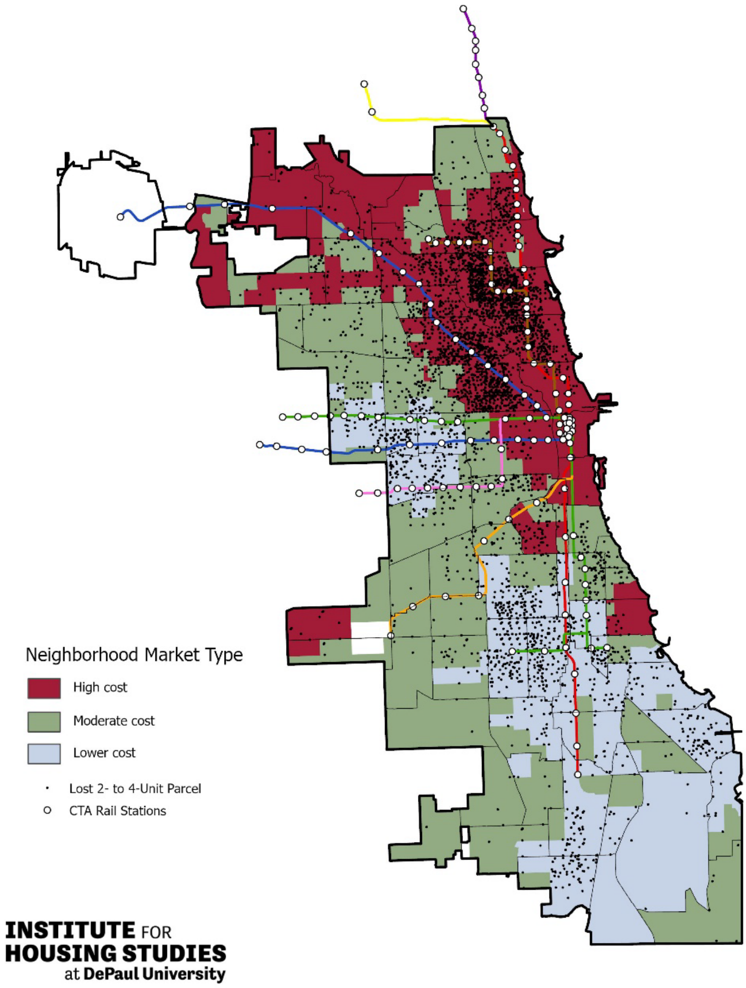
Distribution of Minor Class Changes for Two- to Four-Unit Properties in the City of Chicago by Neighborhood Market Type, 2013 to 2019



Sources: Cook County Assessor's Office; Institute for Housing Studies Data Clearinghouse

Exhibit 7

Map of Lost Two- to Four-Unit Parcels in the City of Chicago by Neighborhood Market Type, 2013 to 2019

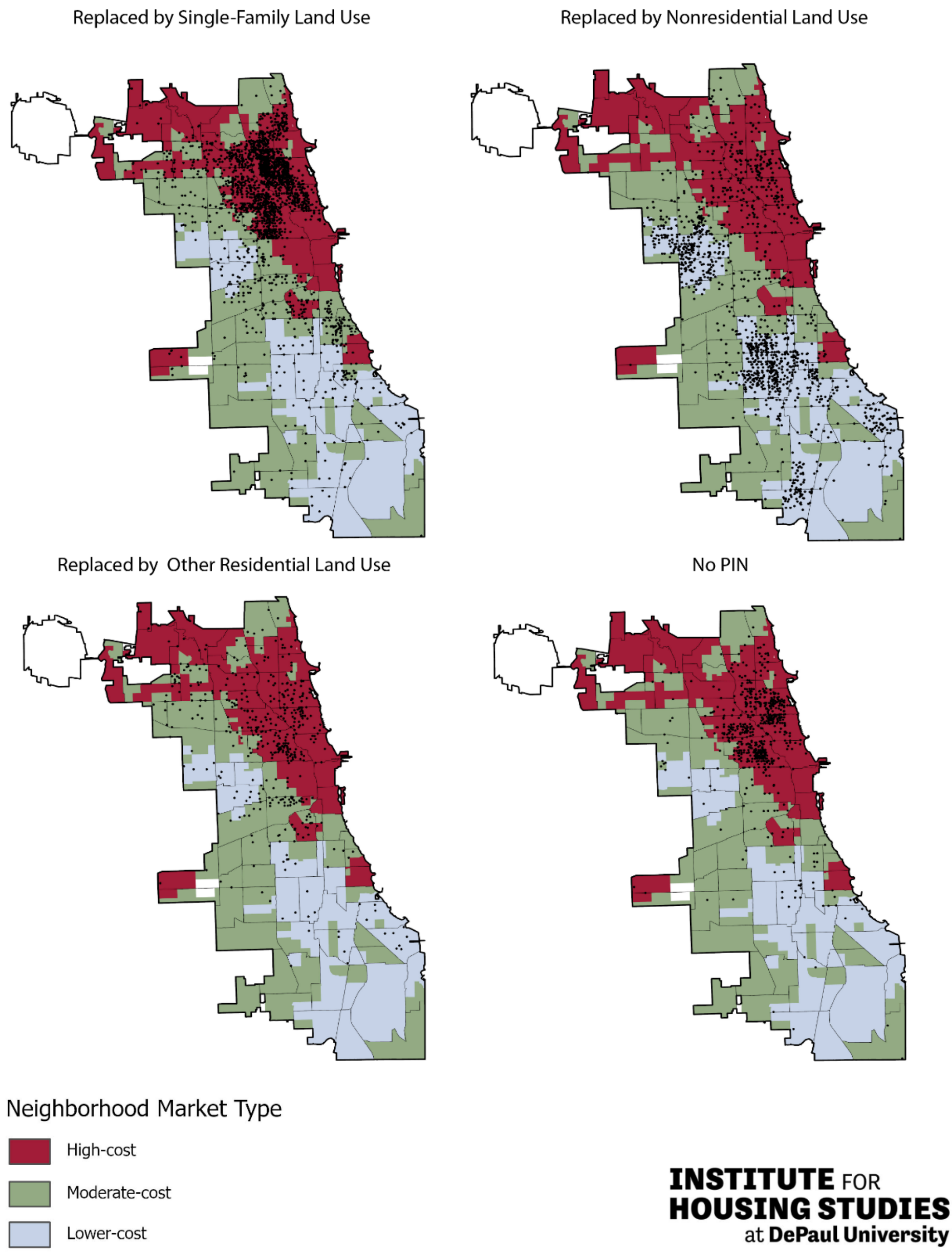


CTA = Chicago Transit Authority.

Sources: Cook County Assessor's Office; Institute for Housing Studies Data Clearinghouse

Exhibit 8

Map of Lost Two- to Four-Unit Parcels by Type of Loss and Neighborhood Market in City of Chicago Community Areas, 2013 to 2019



PIN = parcel identification number.

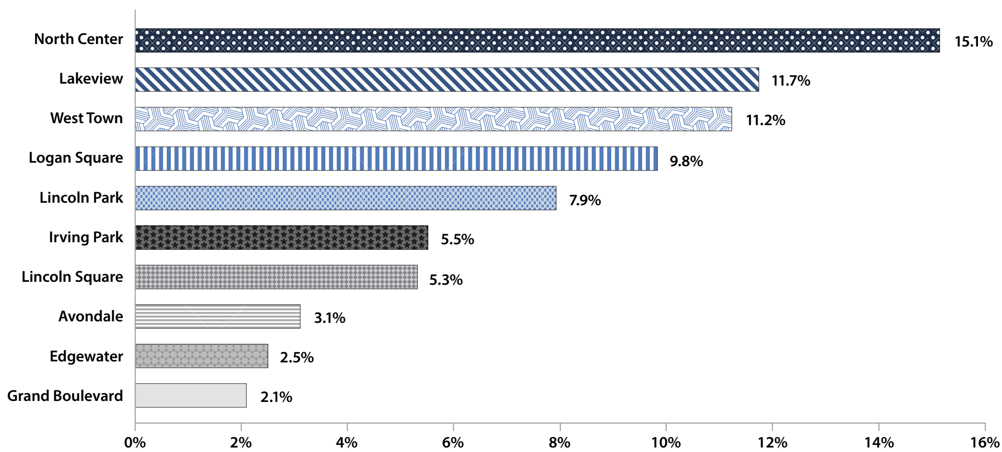
Sources: Cook County Assessor's Office; Institute for Housing Studies Data Clearinghouse

A small group of neighborhoods on the city's North and Northwest Sides account for most of the two- to four-unit properties that single-family homes replaced.

Exhibit 9 highlights Chicago's top 10 community areas with the largest concentration of two- to four-unit properties lost to single-family homes. Of all two- to four-unit buildings that single-family homes replaced citywide, nearly 75 percent occurred in the 10 community areas in exhibit 9, with most of this type of activity taking place in the top five neighborhoods. Neighborhoods with the highest concentration of two- to four-unit buildings lost to single-family home conversion or new construction include high-cost neighborhoods such as North Center (15.1 percent of Chicago's total), Lakeview (11.7 percent), West Town (11.2 percent), Logan Square (9.8 percent), and Lincoln Park (7.9 percent). Exhibit 10 maps the loss of two- to four-unit properties to single-family homes in Chicago's community areas.

Exhibit 9

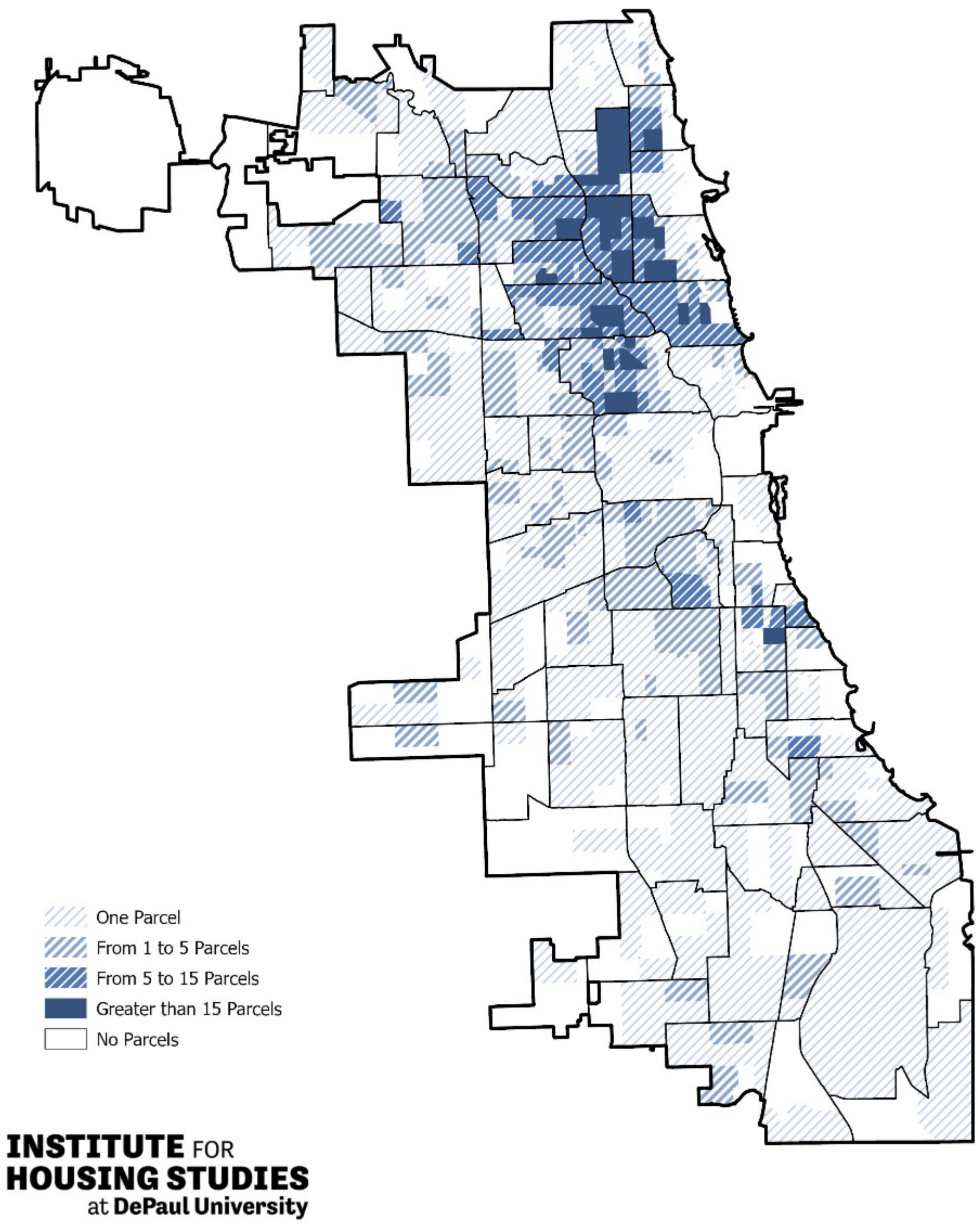
Distribution of Total Two- to Four-Unit Parcels Lost to Single-Family Homes in the City of Chicago by Top 10 Community Areas, 2013 to 2019



Sources: Cook County Assessor's Office; Institute for Housing Studies Data Clearinghouse

Exhibit 10

Map of Two- to Four-Unit Parcels Lost to Single-Family Homes in City of Chicago Community Areas, 2013 to 2019



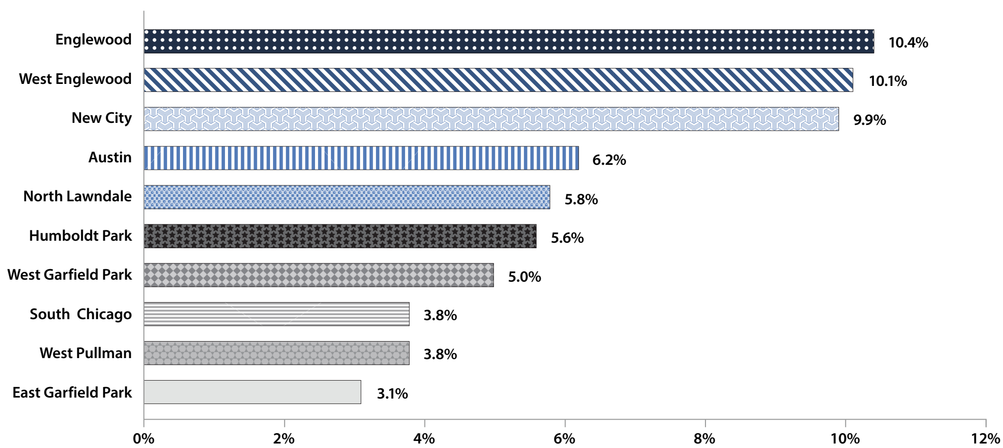
Sources: Cook County Assessor's Office; Institute for Housing Studies Data Clearinghouse

Conversely, two- to four-unit buildings lost to nonresidential land use, typically to vacant land, are concentrated in lower-cost communities on Chicago's South and West Sides.

As a result of historic disinvestment, the impact of the foreclosure and economic crisis, and a slow housing market recovery, many two- to four-unit properties are now vacant land in Chicago's lower-cost neighborhoods. More than 1,150 two- to four-unit buildings in lower-cost communities have been lost since 2013, 80.6 percent of them to nonresidential land uses. The Cook County Assessor's Office classified 89.1 percent of them as vacant land. Exhibit 11 highlights the distribution of two- to four-unit properties lost to vacant land by Chicago's top 10 community areas and illustrates that 53 percent of all two- to four-unit buildings lost to vacant land are in just 7 community areas. The top 5 community areas with the highest shares of the city's two- to four-unit stock lost to vacant land include Englewood (10.4 percent), West Englewood (10.1 percent), New City (9.9 percent), Austin (6.2 percent), and North Lawndale (5.8 percent). This type of activity is also seen in higher-cost areas and some moderate-cost areas, but two- to four-unit stock lost to vacant land in these market types often appears to be side lots or temporarily vacant lots awaiting development. As of the 2021 tax year, more than 96 percent of these former two- to four-unit buildings were still vacant land in lower-cost areas, although slightly less than 43 percent were still vacant land in high-cost neighborhoods. Exhibit 12 maps the lost two- to four-unit stock that became vacant land in 2019 by Chicago's community areas.

Exhibit 11

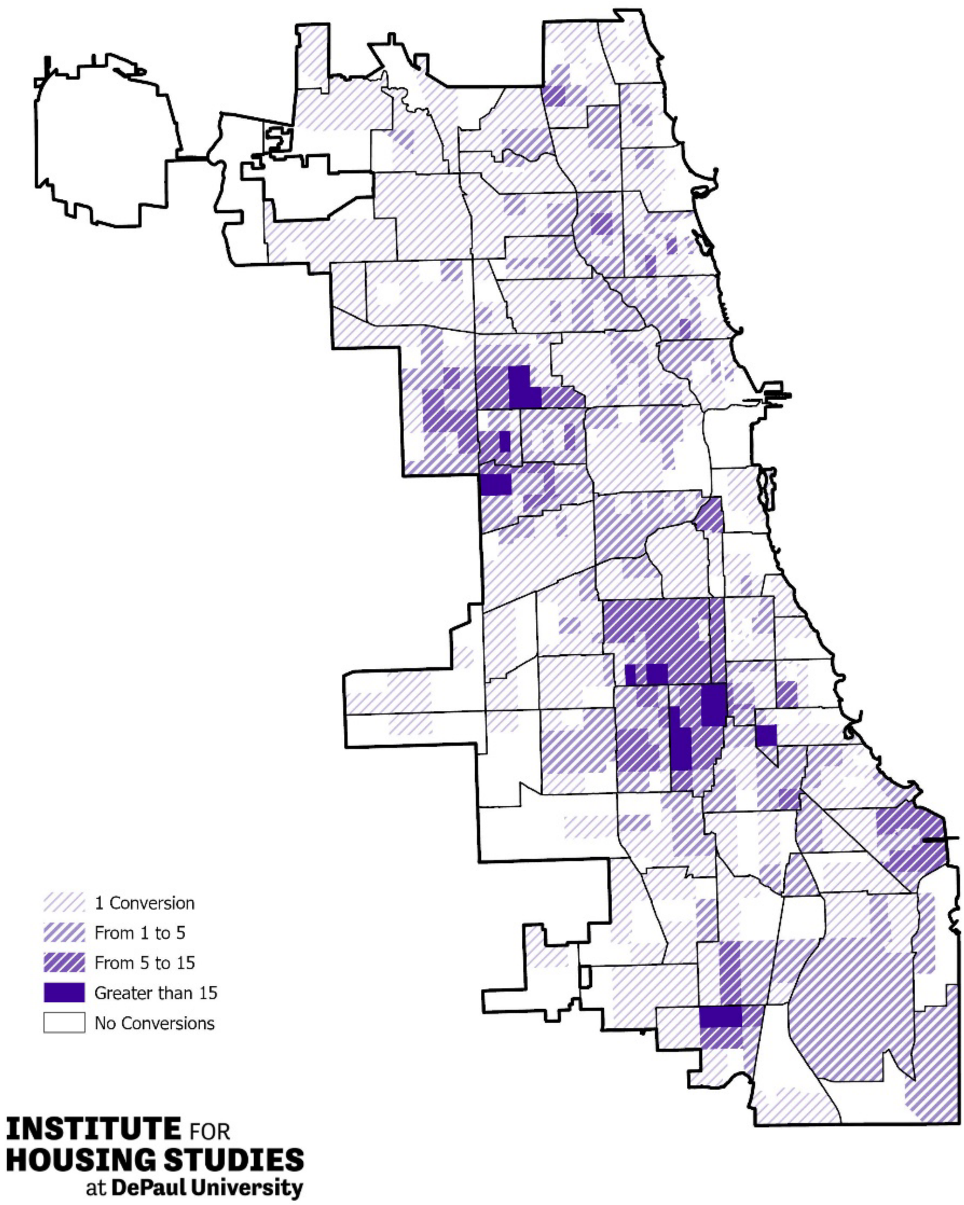
Distribution of Total Two- to Four-Unit Parcels Lost to Vacant Land in the City of Chicago by Top 10 Community Areas, 2013 to 2019



Sources: Cook County Assessor's Office; Institute for Housing Studies Data Clearinghouse

Exhibit 12

Map of Two- to Four-Unit Parcels Lost to Vacant Land in City of Chicago Community Areas, 2013 to 2019



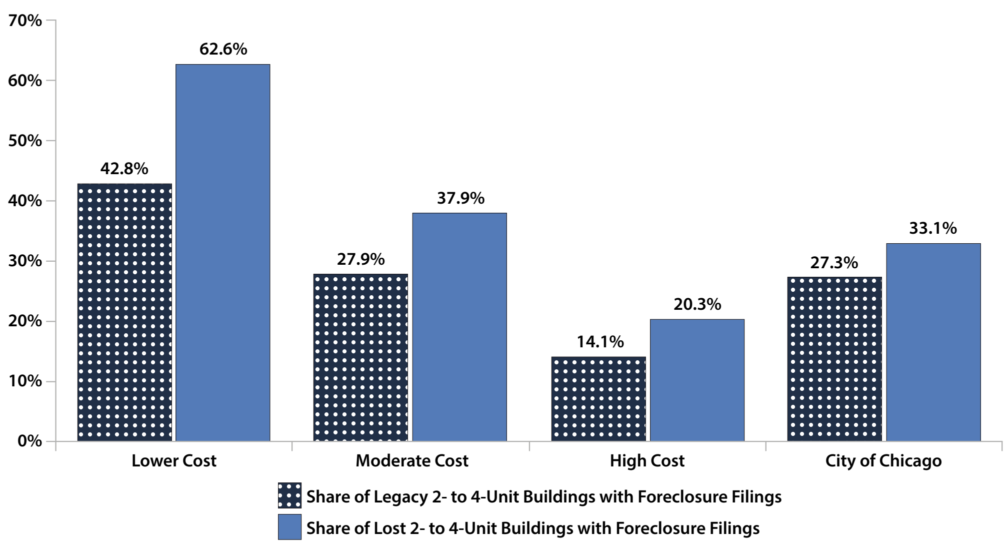
Sources: Cook County Assessor's Office; Institute for Housing Studies Data Clearinghouse

Across all neighborhood market types, lost two- to four-unit parcels are more likely to be associated with foreclosure filings than legacy two- to four-unit buildings active in 2013 and 2019.

As noted previously, research illustrated that two- to four-unit properties had the highest share of parcels associated with foreclosure filings compared with single-family or larger multifamily rental properties in Chicago. Examining foreclosure activity on the lost two- to four-unit stock compared with stable "legacy" two- to four-unit properties, exhibit 13 shows that a higher share of lost two- to four-unit properties have been associated with foreclosure filings since 2005 than two- to four-unit properties that remained in the stock as of 2019. In Chicago, 33.1 percent of lost two- to four-unit buildings were associated with foreclosures compared with 27.3 percent of legacy two- to four-unit buildings. This pattern holds across neighborhood market types, with the highest levels of foreclosure activity in lower-cost neighborhoods.

Exhibit 13

Share of Legacy and Lost Two- to Four-Unit Buildings With at Least One Foreclosure Filing in the City of Chicago by Neighborhood Market Type, 2005 to 2019



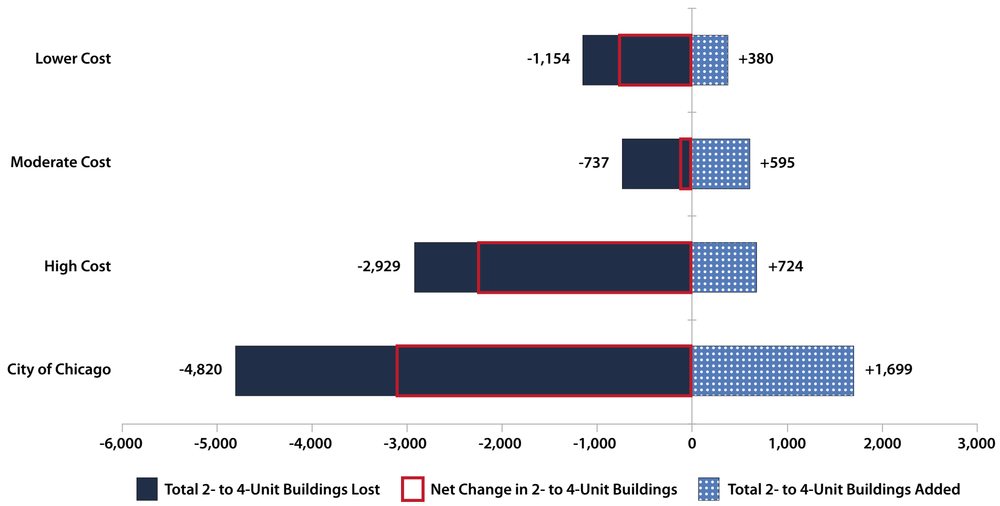
Sources: Cook County Assessor's Office; Institute for Housing Studies Data Clearinghouse

Although some two- to four-unit buildings were added to the stock between 2013 and 2019, these gains do not offset losses in lower- and high-cost neighborhoods.

The housing stock is not static, and two- to four-unit buildings are added to the stock through a number of channels. These channels include adding legal basement apartments in existing single-family structures, converting and reclassifying single-family homes and former commercial or industrial properties into two- to four-unit residential properties, new construction, or reclassifying five- and six-unit properties into two- to four-unit properties in some cases. Exhibit 14 shows that between 2013 and 2019, 1,699 two- to four-unit properties were added to the Chicago housing stock. Despite this addition, the city still had a net loss of more than 3,100 two- to four-unit buildings. Although higher-cost areas accounted for the largest number of "new" two- to four-unit buildings, these neighborhoods still had a net loss of more than 2,200 two- to four-unit buildings. Lower-cost neighborhoods had a net loss of nearly 800 two- to four-unit parcels despite the addition of 380 "new" two- to four-unit buildings.

Exhibit 14

Newer Two- to Four-Unit Buildings (Added 2013–19) Compared With Lost Two- to Four-Unit Buildings in the City of Chicago by Neighborhood Market Type, 2013 to 2019



Sources: Cook County Assessor's Office; Institute for Housing Studies Data Clearinghouse

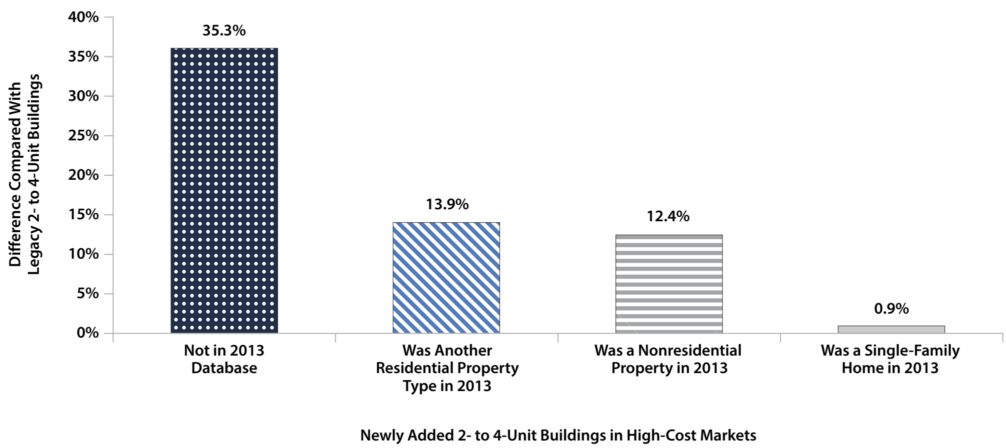
Many newly added two- to four-unit buildings in high-cost markets are likely less affordable than existing "legacy" two- to four-unit buildings.

Exhibit 15 compares the 2019 assessed values for two- to four-unit buildings added to the stock after 2013 with legacy two- to four-unit buildings in high-cost markets. This analysis groups new two- to four-unit stock by former 2013 property class and shows that added two- to four-unit buildings have higher median assessed values than legacy two- to four-unit buildings, except for newly added two- to four-unit buildings that were formerly single-family homes. For example, the median assessed value of newly added two- to four-unit buildings in high-cost neighborhoods compared with parcel identification numbers (PINs) not in the 2013 assessor data was more than 35 percent higher than the median assessed value of legacy two- to four-unit buildings in these neighborhoods.[2] In Chicago's lower- and moderate-cost neighborhoods, assessed values for newly added and established two- to four-unit buildings are generally similar. More research is needed to fully understand the characteristics of newly added two- to four-unit properties. However, higher assessed valuations of these properties relative to the legacy two- to four-unit housing stock indicate that many of the newly added two- to four-unit buildings in higher-cost markets are likely to be less affordable than their peer properties.

[2] Because these 2019 two- to four-unit PINs were not in the 2013 data, the most likely assumption is that they are new construction properties.

Exhibit 15

Difference in Median 2019 Assessed Values of Legacy Two- to Four-Unit Buildings (Active in 2013) Compared With Newer Two- to Four-Unit Buildings by Former Property Class (Added in 2013–19) in High-Cost Markets, 2019



Sources: Cook County Assessor's Office; Institute for Housing Studies Data Clearinghouse

Discussion

This analysis highlights the importance of two- to four-unit properties to Chicago's overall housing stock, quantifies the loss of this stock in neighborhoods across the city, and illustrates the pathways these lost properties take in different market contexts. The analysis highlights the market pressures facing the stock in lower- and high-cost markets citywide and shows that being associated with a foreclosure seems to increase the risk of two- to four-unit properties being lost. These findings highlight the need for a comprehensive approach to preserving two- to four-unit buildings that (1) recognizes the stock's critical importance to providing affordable rental housing, homeownership, and wealth-building opportunities in all Chicago neighborhoods and (2) addresses the spectrum of challenges facing these buildings, their owners, and tenants in different market contexts.

The analysis shows that every type of neighborhood housing market in Chicago is losing two- to four-unit buildings, but the loss is most acute in higher-cost neighborhoods on the city's North and Northwest Sides. In these neighborhoods, this loss is typically due to single-family homes replacing two- to four-unit properties through conversion of the existing buildings or demolition and new construction. This phenomenon highlights how the changing demand for housing, particularly the demand for expensive single-family homes, affects the overall housing supply. Limited additions of "new" two- to four-unit properties do not offset this loss in higher-cost neighborhoods, and data indicate that many of these properties may be less affordable. Meanwhile, in the city's more affordable, moderate-cost neighborhoods, this analysis shows that although the two- to four-unit housing stock is generally stable, losses of two- to four-unit buildings are concentrated in a small number of census tracts, with rising values or near ongoing or planned catalytic investment projects. These findings amplify housing advocates' calls regarding the need for proactive policies to preserve the existing lower-cost rental stock before it is lost to gentrification pressures.

Recent policies in Chicago have attempted to slow down this type of conversion activity through zoning rules. For example, in July 2022, the Chicago City Council passed the Connected Communities Ordinance, a sweeping overhaul of the city's transit-oriented development policies. The ordinance includes reforms to zoning, parking minimums, and accessibility guidelines and also includes provisions that make it more difficult to convert two- and three-unit buildings to single-family homes in neighborhoods with rising prices. In designated Community Preservation Areas near transit and zoned for higher densities, detached houses cannot be constructed, and two- to four-unit properties cannot be converted to single-family homes without a zoning change request, providing an additional level of protection for two- to four-unit buildings close to transit (City of Chicago, 2022).

In Chicago's lower-cost neighborhoods, this analysis found that the loss of two- to four-unit stock is most common through demolition and nonresidential-use replacement, often vacant land. This phenomenon highlights the need for investment in both the broader community and existing housing stock to reverse the tide of long-term population loss, historic disinvestment, and the ongoing legacies of the foreclosure crisis and the Great Recession. Recent efforts in Chicago's lower-cost communities involve policies that help stabilize the existing two- to four-unit housing stock and also remove barriers to redeveloping vacant land, often where two- to four-unit properties once stood. For example, the Renew Woodlawn program helped to restore vacant, foreclosed two- to four-unit buildings in a community with substantial, historic disinvestment and recent concerns about gentrification pressures surrounding the nearby development of the Barack Obama Presidential Center (POAH, 2018). Other programs have leveraged an abundance of city-owned vacant land to develop affordable, new-construction housing to reverse population loss and attract homeowners of color back to some city neighborhoods (IHS, 2023c).

Although this analysis focuses on Chicago, the findings have relevance to other cities attempting to address the loss of older, lower-cost housing and reinvestment-related challenges in different neighborhood market contexts. The loss of this stock has implications for rental housing affordability and homeownership opportunities for modest-income homebuyers and homebuyers of color. Recent research from New York highlights the loss of thousands of housing units as existing units are consolidated into larger, more expensive homes (Brodheim, 2023). Other reporting highlights that beyond Chicago, cities such as Detroit and Pittsburgh struggle with redeveloping vacant land in historically disinvested communities despite the need for housing (Barrett, 2023). For policymakers in communities looking to develop policies to incentivize investment in and the preservation of two- to four-unit buildings, understanding the market pressures that lead to losing this stock and how they vary geographically is essential. This analysis highlights the unique ways that public administrative data can be leveraged to understand conditions on the ground and inform local policy development.

Acknowledgments

The authors would like to thank Evelyn Ryan and Lily Baird for their contributions in preparing this article for publication, as well as Communities United, Neighborhood Housing Services of Chicago, and members of the Chicago Flats Initiative for their continued partnership with the Institute for Housing Studies at DePaul University on issues facing Chicago's two- to four-unit stock.

Authors

Sarah Duda is the deputy director at the Institute for Housing Studies at DePaul University. Geoff Smith is the executive director at the Institute for Housing Studies at DePaul University. Yiwen Jiao is a research and data analyst at the Institute for Housing Studies at DePaul University.

References

An, Brian, Raphael W. Bostic, Andrew Jakabovics, Anthony W. Orlando, and Seva Rodnyansky. 2017. "Understanding the Small and Medium Multifamily Housing Stock." Enterprise Community Partners. https://www.enterprisecommunity.org/resources/understanding-small-and-medium-multifamily-housing-stock.

Barrett, Joe. 2023. "Too Many Vacant Lots, Not Enough Housing: The U.S. Real-Estate Puzzle," *The Wall Street Journal*, August 17.

Black, Curtis. 2020. "'Green Gentrification' and lessons of the 606," *The Chicago Reporter*, January 30.

Brodheim, Adam Oscar. 2023. "Bigger Houses, Fewer Homes: Dwelling Unit Consolidation in New York City." Unpublished master's thesis. Columbia University.

Caine, Paul. 2022. "Legacy of Redlining Continues to Blight Communities of Color," *PBS WTTW*, July 25.

Choi, Jung Hyun, and Caitlin Young. 2020. "Owners and Renters of 6.2 Million Units in Small Buildings Are Particularly Vulnerable During the Pandemic." Urban Institute. https://www.urban.org/urban-wire/owners-and-renters-62-million-units-small-buildings-are-particularly-vulnerable-during-pandemic.

City of Chicago. 2022. "Connected Communities Ordinance." https://www.chicago.gov/city/en/sites/equitable-transit-oriented-development/home/connected-communities-ordinance.html.

Cornelissen, Sharon, Alexander Hermann, and Peyton Whitney. 2023. "Rethinking the American Dream: Small Multifamily Housing Remains Popular Among Immigrant Owners." Cambridge, MA: Harvard University, Joint Center for Housing Studies.

Garcia, David, Muhammad Alameldin, Ben Metcalf, and William Fulton. 2022. *Unlocking the Potential of Missing Middle Housing*. Berkeley: University of California, Terner Center for Housing Innovation. https://ternercenter.berkeley.edu/research-and-policy/unlocking-missing-middle/.

Herbert, Chris. 2023. *Home Prices and Rents Remain High, As Steep Interest Rates Lock Homeowners in Place and Slow Construction*. Cambridge, MA: Harvard University, Joint Center for Housing Studies.

Holeywell, Ryan. 2016. "How the 'Missing Middle' Can Make Neighborhoods More Walkable." Rice University, Kinder Institute for Urban Research. https://kinder.rice.edu/urbanedge/how-missing-middle-can-make-neighborhoods-more-walkable.

Institute for Housing Studies (IHS). 2023a. "2023 State of Rental Housing in the City of Chicago." DePaul University. https://www.housingstudies.org/releases/2023-state-rental-housing-city-chicago/.

———. 2023b. "2023 State of Rental Housing in the City of Chicago Data and Methodology." DePaul University. https://ihs-website-v2-production.s3.amazonaws.com/filer_public/82/dd/82dd78d1-e836-42ac-970a-07c51508e3f0/ihs_state_of_rental_housing_in_chicago_methodology.pdf.

———. 2023c. "Chicago Programs Helping to Expand the Supply of 2 to 4 Unit Properties in Black and Latino Neighborhoods." DePaul University. https://housingstudies.org/blog/chicago-programs-helping-expand-supply-2-4-propert/.

———. 2023d. "Recent Homebuying Patterns for 2 to 4 Unit Properties." DePaul University. https://housingstudies.org/releases/recent-homebuying-patterns-2-4-unit-properties/.

———. 2021a. "Characteristics of the 2 to 4 Stock in Chicago Neighborhoods." DePaul University. https://housingstudies.org/releases/characteristics-2-4-stock-chicago-neighborhoods/.

———. 2021b. "IHS Technical Assistance to Support Preservation of 2 to 4 Unit Properties in Chicago." DePaul University. https://www.housingstudies.org/blog/ihs-technical-assistance-support-preservation-2-4-/.

———. 2021c. "Technical Addendum: Mapping Displacement Pressure in Chicago, 2021." DePaul University. https://ihs-website-v2-production.s3.amazonaws.com/filer_public/4f/50/4f50feea-7f4f-4285-b315-ffa887fcf8e3/ihs_technical_addendum_displacement_pressure_2021.pdf.

———. 2021d. "Technical Addendum: Patterns of Lost 2 to 4 Unit Buildings in Chicago." DePaul University. https://ihs-website-v2-production.s3.amazonaws.com/filer_public/0a/bf/0abf1d16-8c37-4460-b03c-c42307fb4d08/ihs_lost_2to4_methodology_final.pdf.

———. 2019. "2019 State of Rental Housing in Cook County." DePaul University. https://www.housingstudies.org/releases/state-rental-2019/.

———. 2018a. "Understanding Household Income Shifts in Chicago Neighborhoods." DePaul University. https://www.housingstudies.org/blog/household-change-chicago-2010-to-2016/.

———. 2018b. "Why Some Chicago Neighborhoods Are Losing Their Children." DePaul University. https://www.housingstudies.org/blog/some-chicago-neighborhoods-losing-children/.

——— (IHS). 2017. “Mapping Displacement Pressure in Chicago.” DePaul University. https://housingstudies.org/releases/Mapping-Displacement-Pressure-in-Chicago/.

Joint Center for Housing Studies (JCHS). 2023. The State of the Nation’s Housing 2023. Cambridge, MA: Harvard University, JCHS.

National Low Income Housing Coalition (NLIHC). 2023. “The Gap: A Shortage of Affordable Homes.” https://nlihc.org/gap.

Parolek, Daniel. 2020. *Missing Middle Housing: Thinking Big and Building Small to Respond to Today’s Housing Crisis*. Washington, DC: Island Press.

Parolek, Daniel, and Karen Parolek. 2021. “The Middle Housing Affordability Solution.” Alliance for Housing Solutions. https://allianceforhousingsolutions.org/the-middle-housing-affordability-solution/.

Preservation of Affordable Housing (POAH). 2018. *Renew Woodlawn Promoting Homeownership on Chicago’s South Side*. Boston, MA: POAH. https://www.poah.org/sites/default/files/Renew_Woodlawn_Case_Final_1.12.2018.pdf.

Shropshire, Corilyn. 2016. “Group Aims to Preserve Affordable Housing on Gentrifying Northwest Side,” *Chicago Tribune*, November 11.

Theodos, Brett, Eric Hange, Brady Meixell, and Prasanna Rajasekaran. 2019. *Neighborhood Disparities in Investment Flows in Chicago*. Washington, DC: Urban Institute.

U.S. Census Bureau. 2021. “B25024: Units in Structure—1 Year.” American Community Survey. https://data.census.gov/table/ACSDT1Y2022.B25024.

Housing Speculation, Affordable Investments, and Tenant Outcomes in New York City

David M. Greenberg
Julia Duranti-Martínez
Francisca Winston
Spenser Anderson
Local Initiatives Support Corporation

Jacob Udell
Caroline Kirk
University Neighborhood Housing Program

Richard D. Hendra
New School University

Abstract

Housing speculation has been generally understood to be a major driver of displacement and hardship for Black, Indigenous, and People of Color (BIPOC) communities. To explore the impact of speculation, this research assesses tenant outcomes in buildings with the fastest-rising property values in New York City. In so doing, it builds on administrative data on mortgage transactions, sales prices, housing maintenance violations, and marshal's evictions to analyze the association between apartment building finances and tenant well-being. Combining these data with building-level information on affordable housing investments, the article also explores how acquisition of distressed housing by nonprofits, tenant cooperatives, and other responsible owners of affordable housing may disrupt speculative cycles and contribute to positive tenant outcomes. It finds that 1) sales price and mortgage debt increased the most steeply in neighborhoods with higher poverty, higher Black and Latinx populations, a growing percentage of adults with college degrees, and a growing population (in other words, neighborhoods showing signs of gentrification); 2) controlling for community characteristics, buildings with the highest increase in debt had about 0.78 more maintenance violations per unit per year than those that did not; 3) building owners who took on the most additional debt or bought at steepest price increases successfully evicted their tenants at 1.5 times the rates of others who owned properties in similar neighborhoods; and 4) buildings receiving affordable housing investments are less likely to be subject to speculation and are significantly better maintained than comparable properties in similar neighborhoods. Taken together, these findings suggest that speculation, and especially speculative finance, disproportionately impacts BIPOC communities and tenant quality of life, and that affordable housing investments can both protect buildings from speculative practices and improve tenant well-being.

Introduction

Housing speculation is not a new phenomenon. In many ways, speculation has driven the settlement and development of the United States, influencing everything from the economic motivations of settler colonialism and the American Revolution to the explosive growth of major cities like Chicago and Los Angeles (Glaeser, 2013). In recent years, however, institutional investors and private equity have accelerated speculative dynamics in the housing market, in some cases contributing to housing bubbles, such as the one that sparked the Great Recession (Gao, Sockin, and Xiong, 2020). After the Great Recession, institutional investors and private equity also capitalized on homeowner distress, particularly among homeowners of color, who suffered much higher rates of foreclosure than White homeowners and lost $400 billion in collective wealth (Bocian, Li, and Ernst, 2010). In Las Vegas, where corporate landlords' holding in single-family rentals increased by 34 times between 2009 and 2019, some of these larger investors were up to 6 times more likely to evict than a small or medium-sized landlord (Seymour and Akers, 2021). In the multifamily market, the subject of this article, advocates have drawn considerable attention to predatory actors fueling speculation, including private equity, where, supported in part by investments from pension funds, hedge funds, and wealthy individuals, large investors bought hundreds of thousands of units from local landlords (Hornbach et al., 2020). Although not every large investor engages in speculation, these kinds of activities have also been shown to harm tenants: Atlanta neighborhoods with more corporate owners of rental housing are one-third more likely to experience an eviction spike and are also more likely to gentrify (Raymond et al., 2021). The research conducted by Raymond et al. is particularly important for this project because it controls for neighborhood characteristics that might otherwise impact outcomes of interest. This approach is similar in that it examines the net effect of speculation on tenants, controlling for factors such as neighborhood characteristics and building type and size.

Housing speculation is defined in different ways, but it is often applied to the acquisition of properties at some risk to the investor, which also offers an opportunity for greater returns than can be expected from safer investments. Other than the case studies of corporate ownership cited previously, less research has been conducted on speculative practices as a whole on tenant outcomes. To assess the impact of speculation empirically and to explore what tools can promote positive outcomes for tenants, this article examines the purchase and financing of New York City apartment buildings and their association with tenant outcomes, asking three major questions:

1. Which neighborhoods have seen the most speculative activity in the multifamily market? What neighborhood characteristics are associated with higher levels of speculation?

2. What are the consequences of speculation for tenants, for the quality of their homes, and for their likelihood of being evicted?

3. What is the role of affordable housing investments in promoting tenant outcomes or in interacting with speculation?

Several features of this study build on publicly accessible data to contribute to the field's understanding of the interplay between speculation and tenant outcomes. First, since 2003, the

University Neighborhood Housing Program (UNHP) in the Bronx has used City of New York records to create the groundbreaking Building Indicator Project (BIP). The BIP tracks physical and financial distress indicators on more than 70,000 multifamily properties—those with five or more residential units—throughout New York City. More recently, UNHP has added a database of sales and mortgages since 2003 to BIP for that same universe of multifamily properties, relying on raw property-record data from the Automated City Register Information System (ACRIS). BIP data were combined with Census records that matched apartments to their community characteristics, and to these were added building-level records of executed evictions carried out by New York City marshals, drawn from a database taken from public court records maintained by the Housing Data Coalition.[1] These combined data permit an understanding of where speculation occurs and its potential impact on evictions and maintenance quality. Finally, to address the article's third question, the authors combined these data with information from the Subsidized Housing Information Project (SHIP) of the Furman Center of New York University.[2] Because SHIP also records publicly accessible investments in affordable housing at the building level, these additional data can show how community investments may interrupt negative outcomes for tenants and promote positive ones.

Operationalizing Speculative Dynamics in New York With Public Data

Although the research seeks to distinguish properties subject to speculation from other properties in New York's super-heated market, speculation is difficult to operationalize at the building level because there are factors that are not always observable in public data that may contribute to higher sale prices, such as an undervalued property that is well-located or has other unobservable amenities. In this article, changes in asset values of the *same* property over time are central to its operationalization of this definition of speculation. This approach is justified by the practices of speculators themselves. In many cases, net income (rental income after building expenses) drives profit for speculators, and tenants have long drawn attention to ways that speculators realize profits by increasing rents and cutting expenses. However, for speculators who treat apartment buildings as an asset class, landlords and investors see the rising value of their buildings both as a reflection of potential profit and as the main mechanism through which they actually profit (Hornbach et al., 2020).

In rental housing, two types of speculative strategies are predicated on a rapid increase in asset values. One involves purchasing a property and expecting that its value will rise quickly, simply because it is a desirable asset in the current housing market. In this case, the business strategy relies primarily on the assumption that, as property values rise, another investor will be willing to pay a premium for the building in a few years. Between 2000 and 2018, multifamily property values in Queens, Brooklyn, Manhattan, and the Bronx increased between 400 and 600 percent. In the Bronx, the average sale price per unit rose from about $17,325 in 1996 to approximately $175,000 in inflation-adjusted, 2020 terms, during a period when the median household income in the borough actually dropped from about $44,000 to $42,000 (Hornbach et al., 2020). In this

[1] See Housing Data Coalition. n.d. *Housing Data Coalition*. https://www.housingdatanyc.org/.

[2] See NYU Furman Center. n.d. *CoreData.nyc User Guide*. https://furmancenter.org/coredata/userguide/data-downloads.

study, these properties are identified through a focus on buildings that are resold in ways that capitalize on their increased value, because this indicates that they were previously purchased with the goal of realizing windfalls—a hypothesis that is borne out by the fact that these properties often *continue* to be sold for higher amounts, as shown in the following section. This article focuses on the top quartile of repeated sales, which during the study period rose in value at a rate of about 30 percent per unit per year.

The second type of speculative strategy involves debt. When buyers acquire buildings at ever-higher prices, they often do so with loans from a bank or nonbank financial institution. In this situation, the financial institution is incentivized to agree that the market value of a property has risen because it profits from the higher loan amount if it is repaid. Over time, the same owner may come back to the institution to claim that the value of the property has risen again, which justifies adding to the mortgage to reflect its new assumed value. Many owners refinance their mortgages as often as every few years and profit by taking out those new debt proceeds as payouts or to cheaply fund other profitable investments—often while neglecting the properties themselves (Miranova et al., 2022). This financing mechanism, referred to in the real estate industry as "cashing out" or "pulling equity out," is the most common instance of converting an increased asset value into profit and, as such, figures prominently in the analysis. Focusing on the top quartile of refinancing events identified buildings where debt increased by 50 percent per unit per year, a striking figure which in itself suggests the role debt plays in speculation.

This approach to identify speculative investments was adopted because the business strategies and extent of financial risk assumed by an owner can be defined in multiple ways and may inhere in characteristics of a property or strategies for its management that are not easily observable in public data—for example, in rising, realistic projections about net operating income. When assessing speculative risk, neighborhood and building context also matters. Multifamily buildings in prime locations, with higher-income tenants arriving who may pay higher rents, or of particularly high maintenance quality, may be seen as safer investments. To factor in these characteristics of neighborhood and building context, this article's measure of speculation builds from the insight that asset value increases are a measure of profit in housing, examining how much the *same* property increases its sales price from one sale to the next, adjusting for the length of time between the sales. Employing the additional insight that mortgage refinancings are the most common way for landlords to realize asset price increases, the research employs the same strategy for debt, measuring how much additional debt a property takes on, adjusted for the time between debt events.

For example, a 12-unit building that doubled in sales price after a year (2005 to 2006) would be treated the same as a 12-unit building with a sales price that quadrupled in 2 years between 2014 and 2016. The strategy is similar to other paired-sales indexes (such as the Case-Shiller index), which are used to understand asset inflation in relative terms. Although imperfect, the approach holds constant the property itself and its location, and across all multifamily buildings, it is reasonable to assume that higher leaps in sales price or in debt are likely to be signals of greater speculative risk or signs that equity is being extracted based on relatively inflated assumptions of value. (Although it is possible that increased debt is being reinvested into the property, public data do not indicate the use of funds, and the article's analysis of the association between maintenance quality and speculative finance suggests that overall, this kind of reinvestment is not occurring, as described in the section, "Speculative Debt and Housing Quality."

This article focuses on properties in the top quartile of increased sales price and increased debt to provide a clean "cut point" that can be used to describe cohorts of properties over time. This cut-off, while arbitrary, provides a way to identify properties that have been assigned the greatest additional amount of value over time, even in the overheated New York City housing market. This identification strategy also illustrates clear differences between this set of properties and others, but it is important to note that findings hold when examining more linear relationships, that is, when increased sales or debt values are expressed as more continuous measures. In other words, logistic regression results were similar to linear regression results, with the former method employed in this article examining whether a building is in the top quartile of sales-price or debt increases, with the latter employing more continuous measures of time-adjusted, per-unit increases in price or debt.

Understanding Where Speculation Occurs

Using this article's measure of relative per-unit, time-adjusted sales-price increases, Manhattan and Brooklyn have the greatest share of repeated sales events, and they also have the highest proportion of repeated sales events in the highest quartile of asset inflation. Combined, they account for about two-thirds of properties and units in this top quartile of higher resale value (exhibit 1). The location of these properties in New York City's most expensive borough (Manhattan) conforms to the popular image of where already-high housing markets have become increasingly more expensive during the past two decades. The high number of these properties in Brooklyn reflects that, during this period, areas of Brooklyn accelerated their gentrification, and it also corresponds to the fact that Brooklyn is the most populous borough.

Exhibit 1

Proportion of Units by Borough in Time-Adjusted Sales Price Change Among Units Experiencing Repeated Sales

Borough	Lowest Quartile – Change in Sales Price (%)	2nd Quartile – Change in Sales Price (%)	3rd Quartile – Change in Sales Price (%)	Highest Quartile – Change in Sales Price (%)	All Repeated Sales (%)
Manhattan	40	31	34	32	34
Bronx	28	27	20	15	22
Brooklyn	21	28	38	41	32
Queens	11	14	7	12	11
All Boroughs	100	100	100	100	100

Data: Repeated sales, weighted by units in building, 2003–20.
Source: Building Indicator Project, UNHP, drawing on data from New York City

Considerable diversity of income, race, and ethnicity exists *within* boroughs. This broad story of Manhattan- and Brooklyn-driven increases obscures a more granular picture of where values are rising most. Accordingly, regressions linking properties to the characteristics of the census tracts in which they were located, using 2019 estimates from the American Community Survey, permit analysis of which community factors were associated with buildings that proportionally rose the most in price—in other words, which aspects of a neighborhood were associated with speculation. Over the entire study period (2003–20), multifamily buildings were most likely to be resold for the greatest increase in price in areas that have higher poverty, higher Black-identified populations, higher Latinx-identified populations, a higher percentage of adults with college degrees, and a growing population (exhibit 2).[3] This constellation of indicators (with college degrees often being an operational definition of gentrification) suggests unsurprisingly that gentrification is one driver of speculation. This finding cuts against the stereotype of the city's White and affluent neighborhoods becoming astronomically more expensive. In relative terms, gains in value occurred most in Black and brown neighborhoods. At the same time, it is very much in line with what lower-income Black, Indigenous, and People of Color (BIPOC) neighborhood residents and their advocates have been describing, especially in gentrifying areas: apartment buildings in their communities have been subject to rising prices, which, in many cases, have put extraordinary pressure on tenants, as described in later sections.[4]

Exhibit 2

Community-Level Correlates of Being in the Top Quartile of Increased Sales Values (1 of 2)

Logit Regression on Speculative Sale	All Covariates, On Repeat Sales	Subset Covariates, On Repeat Sales	Subset Covariates with Rent Change, On Repeat Sales	Subset Covariates with Rent Change, On All Sales
	(1)	(2)	(3)	(4)
Percent Poverty (ACS 2019)	1.3169*** (3.92)	1.0019*** (3.57)	1.0373*** (3.64)	1.3120*** (5.35)
Percent Poverty Change ACS 2014–ACS 2019	-0.0625 (-1.05)			
Percent Black/African-American (ACS 2019)	0.7381*** (4.39)	0.7635*** (5.84)	0.7697*** (5.84)	1.2255*** (11.19)
Percent Hispanic/Latino (ACS 2019)	0.5244** (2.62)	0.5769*** (3.53)	0.5240** (3.16)	1.1372*** (8.07)
Percent Asian (ACS 2019)	-0.1451 (-0.59)			
Percent Adults w/ College Degree (ACS 2019)	1.3144*** (4.89)	1.4845*** (7.17)	1.2318*** (5.56)	1.1058*** (5.89)
Percent College Degree Change ACS 2014–ACS 2019	-0.0638 (-1.35)			
Median Household Income (ACS 2019)	0.0000 (1.04)			
Population (ACS 2019)	-0.0000*** (-3.74)	-0.0000*** (-3.98)	-0.0000** (-2.94)	-0.0000 (-1.05)
Population Change ACS 2014–ACS 2019	0.5379*** (4.38)	0.4713*** (4.08)	0.3900** (2.65)	0.2816** (3.08)
Percent Rent Change ACS 2014–ACS 2019			0.5547*** (4.17)	0.5638*** (5.01)
Bronx	-0.5015*** (-4.21)	-0.4756*** (-4.07)	-0.4455*** (-3.45)	-0.4615*** (-4.10)

[3] For consistency within the model, these community characteristics were defined through American Community Survey data during the end of the study period (2014–19), so it is accurate to say that these are characteristics of the neighborhoods as they now exist. Preliminary analyses appeared to show the proportion of Asian-identified populations as not substantially influencing the model, so this variable was eliminated from pared-down regressions models.

[4] Linear models show similar results, in that a higher increase in debt occurs in neighborhoods with higher poverty, a higher Black population, a higher Latinx population, and a higher percentage of adults with college degrees.

Exhibit 2

Community-Level Correlates of Being in the Top Quartile of Increased Sales Values (2 of 2)

Logit Regression on Speculative Sale	All Covariates, On Repeat Sales (1)	Subset Covariates, On Repeat Sales (2)	Subset Covariates with Rent Change, On Repeat Sales (3)	Subset Covariates with Rent Change, On All Sales (4)
Brooklyn	0.0979 (1.08)	0.0741 (0.85)	0.0859 (0.82)	0.1984* (2.16)
Queens	0.0094 (0.08)	-0.0231 (-0.21)	0.0251 (0.20)	0.1096 (1.02)
Upper Manhattan	-0.2546* (-2.58)	-0.2647** (-2.75)	-0.2521* (-2.32)	-0.1455 (-1.53)
Year 2003	4.8139*** (6.44)	4.8096*** (6.44)	4.8530*** (6.47)	-1.1858*** (-4.94)
Year 2004	2.6882*** (14.09)	2.6888*** (14.10)	2.7410*** (13.63)	0.3910* (2.30)
Year 2005	2.4354*** (14.59)	2.4341*** (14.59)	2.4802*** (13.92)	1.0474*** (6.44)
Year 2006	1.9136*** (11.59)	1.9110*** (11.59)	1.9334*** (10.95)	1.0367*** (6.31)
Year 2007	1.1512*** (6.97)	1.1579*** (7.02)	1.1642*** (6.56)	0.7672*** (4.57)
Year 2008	0.6929*** (3.96)	0.6996*** (4.00)	0.7482*** (4.00)	0.5545** (3.12)
Year 2010	0.1849 (0.97)	0.2061 (1.08)	0.2339 (1.14)	0.3327 (1.70)
Year 2011	-0.1161 (-0.61)	-0.1218 (-0.64)	-0.0093 (-0.05)	0.1383 (0.71)
Year 2012	0.2371 (1.39)	0.2371 (1.39)	0.2568 (1.40)	0.4700** (2.66)
Year 2013	0.5228** (3.22)	0.5303** (3.27)	0.5508** (3.16)	0.8332*** (4.99)
Year 2014	0.7657*** (4.74)	0.7704*** (4.77)	0.7887*** (4.54)	0.9455*** (5.70)
Year 2015	1.1680*** (7.35)	1.1675*** (7.35)	1.2498*** (7.33)	1.3691*** (8.41)
Year 2016	0.8588*** (5.25)	0.8568*** (5.24)	0.9156*** (5.23)	1.0995*** (6.58)
Year 2017	0.5907*** (3.45)	0.6099*** (3.57)	0.6495*** (3.55)	0.8313*** (4.75)
Year 2018	0.2772 (1.58)	0.2823 (1.61)	0.3567 (1.90)	0.6979*** (3.88)
Year 2019	0.1048 (0.55)	0.1198 (0.63)	0.2056 (1.01)	0.6059** (3.09)
Year 2020	-0.1066 (-0.47)	-0.1216 (-0.54)	0.0399 (0.17)	0.4090 (1.79)
Constant	-2.9738*** (-9.68)	-2.9040*** (-11.21)	-3.0485*** (-11.07)	-4.6876*** (-18.96)
Observations	15193	15233	14229	41734
Pseudo R-squared	0.1024	0.1012	0.1020	0.0389

t statistics in parentheses
** p<0.05. ** p<0.01. *** p<0.001.*
ACS = American Community Survey.
Notes: 1 is full model; 2 is with a trimmed set of covariates; 3 adds a rent-change variable; and 4 identifies predictors of speculative sales, taking all buildings (not just those with repeat sales) into account.
Sources: Building Indicator Project, UNHP, drawing on data from New York City; US Census

This analysis also shows the role of housing market dynamics in driving speculation. The model explores the role of the market in two main ways. First, it adds variables to account for the year in which a property was resold to see whether hot-market periods helped predict speculative sales. This appears to be true: speculative sales were more likely to occur between 2003 and 2008 and between 2013 and 2017, which were hot-market periods broken by the Great Recession. For example, the odds of a speculative sale occurring in 2014 are 2.16 times that of another year in the study (2003–20). Similarly, the odds of a speculative sale occurring in 2015 are 3.21 times higher (exhibit 2, column 2). Second, it adds a variable that accounted for rising rents in the census tract in which the sale took place to explore whether higher sales prices may be driven by purchasers' expectations of higher rental income, based on market dynamics in the surrounding area.

Rising local rents also contribute to the likelihood that a building will be resold for higher amounts (exhibit 2, column 3). However, even when taking both market-cyclical factors and local rent changes into account, race, poverty, and gentrification indicators still predicted speculative sales, although their predictive value decreased modestly. This finding suggests that trends within the housing market at a given point in time do not tell the entire story of sales-price increases. In other words, indicators of a "hot" market are associated with greater increases in a property's value, but signals of race, community distress, and gentrification remain important predictors, even when these market signals are factored in.

A similar pattern emerges when examining characteristics of neighborhoods where the greatest amount of additional debt is taken out on the same property over time. As described previously, taking out more debt on a property is another dynamic of speculation because an owner leverages the asset with the expectation of its increasing value for relatively inexpensive capital, especially when interest rates are low. Debt can obviously be used to improve the property, as an individual homeowner does when taking out a line of credit secured by their home to invest in repairs or amenities. However, landlords overall do not effectively reinvest resources in this way if improved housing maintenance is an indicator of this reinvestment.

At the borough level, as in the case of rising sales prices, Manhattan and Brooklyn are where the highest amount of increased debt occurs, accounting for about two-thirds of the highest per-unit, time-adjusted transactions (exhibit 3). However, when factoring in the role of neighborhood characteristics, a pattern emerges that is observed in speculative sales. More debt is taken out on properties in areas with higher poverty and higher Black and Latinx populations (exhibit 4). For example, an increase in census tract poverty rate from 20 to 30 percent is associated with a 14-percent increase in the odds of a speculative debt event, whereas other factors held constant. Because the relationship is not strictly linear, an increase in poverty rate from 20 to 40 percent is associated with a 30-percent increase in the odds of a speculative debt event. Community-level signals of gentrification—declining poverty, higher proportions of people with college degrees, and increasing populations—are also associated with higher increases in debt.

Exhibit 3

Proportion of Units by Borough in Time-Adjusted Sales Debt Change Among Units Experiencing Repeated Sales

Borough	Lowest Quartile – Change in Debt (%)	2nd Quartile – Change in Debt (%)	3rd Quartile – Change in Debt (%)	Highest Quartile – Change in Debt (%)	All BBLs with Change in Debt (%)
Manhattan	45	34	33	38	38
Bronx	15	21	21	18	18
Brooklyn	26	25	30	30	28
Queens	14	19	16	14	16
All Boroughs	100	100	100	100	100

Data: BBLs (properties identified by borough-block-lot) with change in debt, weighted by units in building, 2003–20.

Source: Building Indicator Project, UNHP, drawing on data from New York City

Exhibit 4

Community-Level Correlates of Being in the Top Quartile of Increased Debt Values (1 of 2)

Logit Regression on Speculative Debt	All Covariates, On Change in Debt (1)	Subset Covariates, On Change in Debt (2)	Subset Covariates with Rent Change, On Change in Debt (3)	Subset Covariates with Rent Change, On All Debt (4)
Percent Poverty (ACS 2019)	1.4479*** (8.04)	1.3080*** (8.24)	1.2951*** (7.97)	0.7968*** (5.43)
Percent Poverty Change ACS 2014–ACS 2019	-0.0929*** (-3.42)			
Percent Black/African-American (ACS 2019)	0.7077*** (8.12)	0.6901*** (9.71)	0.6982*** (9.71)	0.5743*** (8.79)
Percent Hispanic/Latino (ACS 2019)	0.9020*** (8.58)	0.8834*** (9.83)	0.8840*** (9.64)	0.7165*** (8.68)
Percent Asian (ACS 2019)	0.1316 (1.03)			
Percent Adults w/ College Degree (ACS 2019)	0.9520*** (6.67)	0.8122*** (7.27)	0.7800*** (6.51)	0.6156*** (5.71)
Percent College Degree Change ACS 2014–ACS 2019	0.0271 (0.94)			
Median Household Income (ACS 2019)	-0.0000 (-0.68)			
Population (ACS 2019)	-0.0000** (-3.19)	-0.0000** (-3.18)	-0.0000** (-2.94)	0.0000 (1.15)
Population Change ACS 2014–ACS 2019	0.2012** (2.76)	0.2004** (2.79)	0.0344 (0.36)	-0.0059 (-0.07)
Percent Rent Change ACS 2014–ACS 2019			0.2977*** (4.02)	0.1710* (2.54)
Bronx	-0.5164*** (-8.44)	-0.5271*** (-8.81)	-0.5461*** (-8.44)	-0.2825*** (-4.84)
Brooklyn	0.1479*** (3.47)	0.1483*** (3.64)	0.1086* (2.26)	-0.0003 (-0.01)
Queens	-0.1729** (-3.10)	-0.1633** (-2.94)	-0.2086*** (-3.47)	-0.2620*** (-4.74)
Upper Manhattan	-0.2034*** (-4.14)	-0.2007*** (-4.13)	-0.2349*** (-4.51)	-0.0876 (-1.85)

Exhibit 4

Community-Level Correlates of Being in the Top Quartile of Increased Debt Values (2 of 2)

Logit Regression on Speculative Debt	All Covariates, On Change in Debt (1)	Subset Covariates, On Change in Debt (2)	Subset Covariates with Rent Change, On Change in Debt (3)	Subset Covariates with Rent Change, On All Debt (4)
Year 2003	2.2307** (3.13)	2.2262** (3.12)	1.8812* (2.45)	-3.9386*** (-7.80)
Year 2004	1.4488*** (10.44)	1.4381*** (10.41)	1.4279*** (9.58)	-1.4394*** (-12.35)
Year 2005	2.1445*** (20.87)	2.1450*** (20.91)	2.2400*** (20.26)	0.3615*** (4.26)
Year 2006	1.9280*** (21.12)	1.9235*** (21.13)	1.9870*** (20.16)	0.8226*** (10.05)
Year 2007	1.4254*** (17.19)	1.4274*** (17.26)	1.4475*** (16.20)	0.9712*** (12.14)
Year 2008	0.7354*** (8.82)	0.7373*** (8.87)	0.7707*** (8.56)	0.6459*** (7.78)
Year 2010	-0.0655 (-0.74)	-0.0721 (-0.82)	0.0037 (0.04)	0.1914* (2.14)
Year 2011	-0.4565*** (-5.43)	-0.4648*** (-5.54)	-0.4721*** (-5.18)	-0.0664 (-0.76)
Year 2012	-0.1520 (-1.94)	-0.1548* (-1.98)	-0.1397 (-1.65)	0.2863*** (3.57)
Year 2013	0.1930* (2.56)	0.1873* (2.49)	0.2189** (2.68)	0.6720*** (8.71)
Year 2014	0.6140*** (8.13)	0.6116*** (8.13)	0.6291*** (7.68)	1.0518*** (13.68)
Year 2015	0.7489*** (10.07)	0.7433*** (10.02)	0.8066*** (10.00)	1.2679*** (16.74)
Year 2016	0.6987*** (9.24)	0.6940*** (9.20)	0.7371*** (8.99)	1.1365*** (14.78)
Year 2017	0.2397** (3.08)	0.2356** (3.03)	0.2795*** (3.32)	0.7462*** (9.39)
Year 2018	-0.0895 (-1.14)	-0.0910 (-1.16)	-0.0634 (-0.74)	0.4996*** (6.19)
Year 2019	-0.2718*** (-3.40)	-0.2744*** (-3.44)	-0.2552** (-2.94)	0.3394*** (4.12)
Year 2020	-0.6605*** (-7.62)	-0.6667*** (-7.71)	-0.6226*** (-6.66)	0.0412 (0.46)
Constant	-2.3881*** (-15.43)	-2.2908*** (-17.32)	-2.3366*** (-16.57)	-3.0715*** (-23.84)
Observations	51496	51686	45031	77697
Pseudo R-squared	0.0719	0.0716	0.0750	0.0545

t statistics in parentheses

** p<0.05. ** p<0.01. *** p<0.001.*

ACS = American Community Survey.

Notes: 1 is full model; 2 is with a trimmed set of covariates; 3 adds a rent-change variable; and 4 identifies predictors of speculative debt, taking all buildings (not just those with repeat debt) into account.

Sources: Building Indicator Project, UNHP, drawing on data from New York City; US Census

In many cases, increased debt is supported by a higher valuation of the property by a lender. The more a property is worth, the easier it is to take out a loan corresponding to its higher value. One clear indicator of a property's value is the rent a landlord can collect. When adding changes in neighborhood rents to the model, these changes do have a statistically significant association with a property taking on the highest levels of increased debt. Rising rents, however, did not play as significant a role in predicting increased debt as they did in predicting increased sales prices. The rent-change variable also did not seem to impact the role of other variables, such as poverty and race, meaning that even when taking rising rent levels into account, the net effect of a building's location in lower-income BIPOC communities remained similar.

What are the Consequences of Speculation for Maintenance Quality?

It is important to understand how market forces have impacted BIPOC and lower-income communities—to show that the greatest wealth increases for owners are more likely to have been generated in communities of color and from buildings that likely house some of the city's poorest tenants. It is also important to show the *consequences* of property owners' speculative wealth building on tenants and communities. To do so, this article draws on the fact that, since its inception, the BIP has collected information on housing maintenance violations the city has recorded on rental properties. In New York, maintenance code violations are reported by tenants and verified by inspectors from the city's Department of Housing Preservation and Development (HPD), which issues citations to the landlord for these problems. HPD violations include a wide range of issues, such as fire safety; heat and hot water problems; defective faucets, drains, and pipes; lead-based paint; vermin, such as cockroaches, mice, and rats; broken plaster; or trash accumulation in common areas.[5] Although violations are an imperfect measure of housing quality because they are reactive to tenant complaints, they are the best available data source for maintenance quality across all New York apartment buildings. This analysis of the relationship between speculation and housing maintenance violations starts in 2014, the point at which city databases provided easier-to-access, higher-quality records. This fact limits the time range of the study, but it still provides a recent view of maintenance quality and its association with speculative activity.

Speculative Sales and Housing Quality

One might think that buildings with few maintenance problems would be sold for the highest change in prices, reflecting the value of the property. Looking across New York, this expectation holds somewhat true. During the study period, the highest-reselling quartile of properties has about 17 to 20 percent of all HPD violations when weighted by the number of units in the building. These properties' share of violations is slightly less than their overall share of units but more than one might expect because these properties escalated the most in value (exhibit 5). One of the reasons buildings that sold for higher values do not have higher maintenance quality is that in some communities, particularly Lower Manhattan and Queens, there are years when the top 25 percent of units have *more* than their share of maintenance violations. For example, in 2015, 2016, and 2019, the highest-rising quartile of sales prices carried 31 percent, 27 percent, and 29 percent of HPD violations among resold units, respectively. In other words, in those areas, the highest-rising sales prices appear to be for buildings with relatively *worse* quality.

[5] For descriptions of violations, see New York's Housing Maintenance Code, https://www.nyc.gov/assets/buildings/pdf/HousingMaintenanceCode.pdf.

Exhibit 5

Proportion of HPD Violations Recorded for the Top Quartile of Time-Adjusted Increased Sales Prices, Weighted by Unit

Year	All Boroughs (%)	The Bronx (%)	Brooklyn (%)	Lower Manhattan (%)	Upper Manhattan (%)	Queens (%)
2014	19	25	14	5	18	28
2015	20	17	21	31	16	27
2016	19	18	19	27	16	31
2017	17	17	14	15	25	14
2018	17	23	13	16	21	11
2019	17	19	14	29	21	13
2020	19	18	22	17	20	7

Data: All BBLs (properties identified by borough-block-lot) with a repeated sale in 2016–17 cohort.
Source: Building Indicator Project, UNHP, drawing on data from New York City

Multiple factors may contribute to housing maintenance problems. Rental income streams are a major driver. The higher the rent rolls, the more that can be directed toward repair. Other factors are the properties' age and construction features. To account for these factors, regressions explored the relationship between speculation (in the form of increased prices) and maintenance quality while holding constant factors such as neighborhood poverty and race (exhibit 6). The model also includes borough-level variables to account for geographic patterns of development that might capture a property's age and construction methods. As previously described, an apartment's location in a lower-income BIPOC neighborhood made it more likely to be sold for the highest additional amount. Because these places are also neighborhoods with higher housing maintenance problems, it might be possible that the association between higher sales prices and maintenance problems is driven by community characteristics and not by the speculative event itself. By controlling for poverty, race, and income characteristics, it is therefore possible to examine the impact of speculation on housing maintenance problems over and above these factors.[6]

[6] As described in exhibit 9, the research also examined at temporal relationships—whether housing violations tend to follow or precede a speculative sale. There is evidence that in New York, the same set of distressed apartment buildings are being resold for higher and higher values and have increased debt taken on them. More violations help predict being sold for the highest additional amount, although being sold is more predictive of subsequent violations, reinforcing the potentially causal relationship between speculation and maintenance quality.

Exhibit 6

OLS Regression Results of Speculative Sales 2016–17 on Violations Per Unit, 2018–20

OLS Regressions	Dependent Variable—HPD Violations Per Unit, 2018–20					
	Citywide	Lower Manhattan	Upper Manhattan	Bronx	Brooklyn	Queens
	(1)	(2)	(3)	(4)	(5)	(6)
Total Speculative Sales 2016–17	0.486***	0.418*	1.338***	1.719***	-0.168	-0.038
	(3.74)	(2.48)	(3.60)	(4.72)	(-0.72)	(-0.14)
Percent Poverty (ACS 2019)	0.704***	0.954***	-1.407*	0.546	-0.218	-0.330
	(5.31)	(4.37)	(-2.32)	(1.22)	(-0.77)	(-1.08)
Percent Black/ African-American (ACS 2019)	2.008***	-0.542*	1.101*	1.424***	2.082***	0.896***
	(36.90)	(-2.00)	(2.14)	(3.97)	(24.23)	(5.23)
Percent Hispanic/ Latino (ACS 2019)	1.208***	0.704***	2.203***	0.591	1.400***	0.799***
	(18.65)	(5.13)	(3.54)	(1.40)	(10.95)	(6.98)
Percent Adults w/ College Degree (ACS 2019)	-0.289***	-0.160	-0.815	-2.521***	-0.855***	-0.280
	(-3.89)	(-1.50)	(-1.19)	(-4.04)	(-5.39)	(-1.82)
Population (ACS 2019)	-0.000	0.000	0.000	0.000	0.000	-0.000***
	(-1.17)	(1.44)	(0.70)	(0.47)	(0.00)	(-3.86)
Population Change ACS 2014–ACS 2019	-0.227***	0.062	0.007	-0.106	-0.548***	0.131
	(-3.87)	(0.88)	(0.02)	(-0.74)	(-4.51)	(1.50)
Constant	0.304***	0.266*	0.572	1.186**	0.826***	0.470***
	(4.83)	(2.42)	(0.80)	(2.80)	(6.06)	(3.93)
Observations	75,768	17,617	7,132	9,041	29,899	12,079
R-squared	0.047	0.012	0.021	0.021	0.038	0.011
Adjusted R-squared	0.047	0.012	0.020	0.020	0.038	0.010
RMSE	2.835	1.215	3.024	3.341	3.494	1.960
F	534.215	30.348	21.680	27.149	167.890	18.740

t statistics in parentheses

** p<0.05. ** p<0.01. *** p<0.001.*

ACS = American Community Survey. HPD = Department of Housing Preservation and Development. OLS = ordinary least square. RMSE = root mean square error.

Sources: Building Indicator Project, UNHP, drawing on data from New York City; US Census

Even when taking these geographic and community factors into account, a property's being acquired as part of a speculative purchase in 2016–17 predicts more housing maintenance violations on that building in 2018–20. When examining these dynamics borough by borough, this overall citywide association is driven by speculative purchases in Manhattan and the Bronx, which conforms to claims by tenants and their advocates that these are places that have been hit particularly hard by speculation. When running regressions independently for each borough, in Brooklyn and Queens, no statistically significant association is apparent between speculative purchases and housing maintenance violations after controlling for community characteristics, although the strength of the association in Manhattan and the Bronx resulted in an overall significant effect. Overall, parcels identified by borough, block, and lot (BBLs) with at least one speculative sale in 2016–17 have 1.09 HPD violations per unit in 2018–20, whereas BBLs without a speculative sale have 0.53 violations per unit in the same period—a difference of about 0.56 violations per unit (exhibit 7). This result aligns with the model, in which the coefficient for total speculative sales in 2016–17 is 0.486—that is, holding all neighborhood characteristics constant, each speculative sale in 2016–17 is associated with a 0.486 increase in HPD violations per unit in 2018–20.

Exhibit 7

Speculative Sales and Violations Per Unit

BBL with Speculative Sale 2016–17	Total HPD Violations 2018–20	Total Units	Total Violations Per Unit 2018–20
Yes	13,025	11,940	1.09
No	1,389,549	2,610,533	0.53
Total	1,402,574	2,622,473	0.53

BBL = property identified by borough-block-lot. HPD = Department of Housing Preservation and Development.
Sources: Building Indicator Project, UNHP, drawing on data from New York City

Speculative Debt and Housing Quality

As previously described, another dynamic of speculation involves taking on increasing debt on apartment buildings—a form of financialization that provides low-cost capital that can be used for higher-return investments. Without controls introduced, the steepest increase in sales price, overall, involved buildings with slightly proportionally fewer maintenance violations than their share of all repeat sales. However, the buildings that took on the greatest increase in debt, without controls, have *more* than their share of housing maintenance problems when adjusting for building size. That is, the top 25 percent of buildings acquiring the largest increases in debt account for about 38 percent of maintenance violations from 2014 to 2020, with some variations by borough (exhibit 8).

Exhibit 8

Proportion of HPD Violations Recorded for the Top Quartile of Increased Debt, Weighted by Units

Year	All Boroughs (%)	The Bronx (%)	Brooklyn (%)	Lower Manhattan (%)	Upper Manhattan (%)	Queens (%)
2014	38	35	48	47	29	21
2015	35	35	41	34	29	25
2016	34	31	35	32	36	32
2017	37	35	41	38	36	33
2018	37	37	39	31	42	27
2019	38	40	37	29	39	28
2020	38	42	40	22	36	30

Data: All BBLs (properties identified by borough-block-lot) with a change in debt in 2016–17 cohort.
HPD = Department of Housing Preservation and Development.
Source: Building Indicator Project, UNHP, drawing on data from New York City

Greater debt can possibly be invested back into properties, especially to repair buildings and provide other forms of property maintenance. For this reason, it might also make sense that more debt is taken out on more distressed properties, and there is some evidence that this also occurs. Buildings with more maintenance violations in 2016–17 are more likely to take on the highest additional debt in 2018–20, as described in exhibit 9. This finding also affirms advocates' understanding that it is often the same distressed portfolios that take on more debt over time. Furthermore, as previously described, lower-income neighborhoods of color were most likely to have properties that took on the greatest amount of debt, and these are also the places with the most housing maintenance issues.

Exhibit 9

Temporal Relationships Between Sales, Debt, and Violations

OLS Regressions	Dependent Variable					
	Total Speculative Sales 2016–17	Total HPD Violations 2018–20	Total HPD Violations 2014–15	Total Speculative Debt Events 2016–17	Total HPD Violations 2018–20	Total HPD Violations 2014–15
	(1)	(2)	(3)	(4)	(5)	(6)
Total Speculative Sales 2014–15	0.008** (3.04)					
Total Speculative Sales 2016–17		9.195*** (3.57)	7.363*** (5.19)			
Total Speculative Debt Events 2014–15				0.045*** (14.36)		
Total Speculative Debt Events 2016–17					26.693*** (22.64)	13.648*** (20.99)
Constant	0.006*** (21.15)	18.221*** (89.21)	9.508*** (84.49)	0.027*** (42.79)	17.510*** (85.09)	9.161*** (80.75)
Observations	76,739	76,739	76,739	76,739	76,739	76,739
R-squared	0.00012100	0.00016600	0.00035000	0.00268000	0.00663000	0.00571000
RMSE	0.0790	56.4070	31.0810	0.1720	56.2240	30.9970
F	9.25	12.73	26.89	206.13	512.46	440.77

t statistics in parentheses

*** p<0.01. *** p<0.001.*

HPD = Department of Housing Preservation and Development. OLS = ordinary least square. RMSE = root mean square error.

Source: Building Indicator Project, UNHP, drawing on data from New York City

For all these reasons, it is important to understand the net effect of taking on higher levels of debt on housing violations by factoring community context into the model, and in fact, even controlling for factors like neighborhood poverty and race, speculative debt levels still impact violations. That is, over and above the influence of poverty and race, a building that takes on higher levels of increased debt in 2016–17 is more likely to have increased maintenance problems in 2018–20 (exhibit 10). Overall, properties with at least one speculative debt in 2016–17 have 1.37 HPD violations per unit in 2018–20, whereas properties without speculative debt have 0.51 violations per unit in the same period—a difference of about 0.86 violations per unit (exhibit 11). This calculation aligns with the model presented in exhibit 10, in which the coefficient for total speculative sales in 2016–17 is 0.780, suggesting that with controls introduced, each speculative debt event in 2016–17 is associated with a 0.780 increase in HPD violations per unit in 2018–20.

Exhibit 10

OLS Regression Results of Speculative Debt 2016–17 on Adjusted Violations 2018–20

OLS Regressions	Dependent Variable—HPD Violations Per Unit 2018–20					
	Citywide	Lower Manhattan	Upper Manhattan	Bronx	Brooklyn	Queens
	(1)	(2)	(3)	(4)	(5)	(6)
Total Speculative Debt Events 2016–2017	0.780***	0.197***	1.454***	1.916***	0.376**	0.177
	(13.09)	(3.50)	(7.91)	(12.14)	(3.20)	(1.26)
Percent Poverty (ACS 2019)	0.672***	0.953***	-1.479*	0.544	-0.229	-0.336
	(5.06)	(4.37)	(-2.45)	(1.22)	(-0.81)	(-1.10)
Percent Black/African-American (ACS 2019)	1.980***	-0.544*	1.049*	1.443***	2.058***	0.890***
	(36.41)	(-2.01)	(2.04)	(4.05)	(23.92)	(5.19)
Percent Hispanic/Latino (ACS 2019)	1.180***	0.705***	2.125***	0.505	1.386***	0.797***
	(18.23)	(5.14)	(3.42)	(1.21)	(10.83)	(6.96)
Percent Adults w/ College Degree (ACS 2019)	-0.317***	-0.159	-0.911	-2.447***	-0.873***	-0.284
	(-4.27)	(-1.48)	(-1.33)	(-3.95)	(-5.50)	(-1.84)
Population (ACS 2019)	0.000	0.000	0.000	0.000	0.000	-0.000***
	(-1.33)	(1.46)	(0.45)	(0.30)	(-0.04)	(-3.86)
Population Change ACS 2014–ACS 2019	-0.223***	0.063	-0.001	-0.119	-0.545***	0.132
	(-3.80)	(0.90)	(-0.00)	(-0.83)	(-4.48)	(1.51)
Constant	0.320***	0.260*	0.662	1.164**	0.836***	0.470***
	(5.08)	(2.37)	(0.93)	(2.77)	(6.13)	(3.94)
Observations	75,768	17,617	7,132	9,041	29,899	12,079
R-squared	0.049	0.012	0.028	0.034	0.038	0.011
Adjusted R-squared	0.049	0.012	0.027	0.033	0.038	0.010
RMSE	2.83	1.22	3.01	3.32	3.49	1.96
F	557.79	31.23	28.89	45.36	169.33	18.97

t statistics in parentheses

** p<0.05. ** p<0.01. *** p<0.001.*

ACS = American Community Survey. HPD = Department of Housing Preservation and Development. OLS = ordinary least square. RMSE = root mean square error.

Sources: Building Indicator Project, UNHP, drawing on data from New York City; US Census

Exhibit 11

Speculative Debt and Maintenance Problems

BBL with Speculative Debt 2016-17	Total HPD Violations 2018–20	Total Units	Total Violations Per Unit 2018–20
Yes	95,773	70,001	1.37
No	1,306,801	2,552,472	0.51
Total	1,402,574	2,622,473	0.53

BBL = property identified by borough-block-lot. HPD = Department of Housing Preservation and Development.

Source: Building Indicator Project, UNHP, drawing on data from New York City

Many reasons might explain this association between debt and poorer housing quality. In some instances, greater debt may directly *cause* maintenance problems. Because landlords use rental income to make mortgage payments, greater loan amounts mean that a higher proportion of rent rolls may be directed toward debt service over building expenses, resulting directly in deteriorating building conditions. In some cases, a landlord might also increase rent to meet building payments, although rent increases at the building level are not observable in the data. In other instances, taking on high levels of debt may be associated with a kind of extractive behavior on the part of landlords—a strategy of drawing out equity to be used for other investments. Regardless of the mechanism, this finding has important policy implications in that taking on high amounts of additional debt is a leading signal of problems for tenants—more powerful even than a speculative increase in sales price.

Speculation and Displacement

Displacement may occur in many ways. As Marcuse argued in his classic 1985 paper, it can occur directly, as individuals are forced to leave their homes due to landlord harassment, rent increases, or building conditions that threaten family well-being—in other words, through physical or economic means. Displacement can also occur indirectly and through a form of exclusionary displacement because BIPOC individuals with lower wages or income who might otherwise have occupied a unit in a community of color are unable to do so when a higher-income or White household has moved into that dwelling (Marcuse, 1985). Data are not publicly available to measure all these forms of displacement, although they are important dynamics of neighborhood change.

Among *direct* forms of displacement, eviction is one of the most traumatic—not only an event caused by poverty, but one which itself causes poverty by subjecting individuals and families to trauma, work and educational disruptions, and, in many cases, great expense (Desmond, 2016). In New York City, the lawful eviction process starts with a notice from the landlord requiring rent to be paid or some lease violation to be corrected. At that point, rather than undergo a court process, many households will choose to vacate their apartment for another, move in with friends or relatives, or seek to enter a shelter. If they do not, the landlord may file for eviction in housing court. Around 175,000 to 190,000 such cases have been filed per year in the past decade, with the majority in the Bronx, hovering at around 20 filings per 100 private dwelling units each year. Of these filings, about 60 percent result in some kind of judgment, but only 10 percent then proceed to the execution of an eviction through court warrant, partly because not every judgment goes against the tenant and partly because many tenants will leave or otherwise resolve their case before such a warrant is executed (Furman Center, 2019). In New York, even though eviction courts were not operating during the pandemic, there were over 223,000 filings waiting to be adjudicated by 2021 (Brand, 2021).

For these reasons, even though eviction warrants represent a very small proportion of eviction filings and an even smaller subset of displacement activity, they are an important phenomenon to study. Completed evictions are also available at a level that ties the eviction to a specific property. In New York, officers of the court are known as marshals, and various marshal's offices have recorded the dates and addresses where they were ordered by housing court to execute an eviction and give control of the apartment back to the landlord. Building on New York City marshals' records, New York's Housing Data Coalition created a file of executed eviction warrants. By adding marshals' data on eviction judgments to the analysis file, it is possible to examine whether a speculative event—a building being in the top tier of sales-price or debt increase—increases the likelihood of a landlord filing for eviction. The dependent variable in this case was the number of eviction judgments per unit. Because pandemic-era restrictions changed eviction dynamics (although the restrictions did not stop eviction filings), the analysis ends in 2019.

Findings echo and support advocates' longstanding claims about the impact of speculation on evictions. Overall, properties with at least one speculative event in 2014–16 experienced 0.0273 evictions per unit in 2017–19, compared to 0.0149 evictions per unit in properties without a speculative event—making the eviction rate almost twice as high in properties with a speculative event compared to properties without, as shown in exhibit 12. To show the *net* effect of speculation, over and above neighborhood-level factors, such as race and poverty, regressions introduced community-level controls—an especially important comparison because speculation tends to occur in the same neighborhoods that also see greater levels of poverty and higher levels of eviction. After introducing controls, such as the size of the property, poverty, neighborhood racial demographics, and local rent changes, to account for local market effects (exhibit 13), the coefficient for speculation on evictions is 1.489, suggesting that properties subject to speculative activity evict at 1.5 times the rate of comparable buildings in similar neighborhoods.

Exhibit 12

Speculative Events and Eviction Warrants

Any Speculative Event (Sale or Debt) 2014-16	Total Evictions 2017–19	Total Units	Total Evictions Per Unit 2017–19
Yes	4,355	159,782	0.0273
No	34,661	2,462,691	0.0141
Total	39,016	2,622,473	0.0149

Sources: Building Indicator Project, UNHP, drawing on data from New York City; New York Housing Data Coalition

Exhibit 13

Association Between Speculative Events and Evictions (1 of 2)

	OLS (1) Evictions 2017-20	Poisson (2) Evictions 2017-20	Poisson IRR (3) Evictions 2017-20
Any Speculative Event (Sale or Debt) 2014–16	0.293*** (10.19)	0.404*** (9.61)	1.498*** (9.61)
Units Per Property	0.00957*** (151.45)	0.000514*** (9.07)	1.001*** (9.07)
Percent Poverty (ACS 2019)	0.420*** (4.62)	0.804*** (3.72)	2.233*** (3.72)
Percent Black/African-American (ACS 2019)	1.051*** (27.94)	1.862*** (28.42)	6.440*** (28.42)
Percent Hispanic/Latino (ACS 2019)	0.923*** (20.63)	1.553*** (16.46)	4.726*** (16.46)

Exhibit 13

Association Between Speculative Events and Evictions (2 of 2)

	OLS (1) Evictions 2017-20	Poisson (2) Evictions 2017-20	Poisson IRR (3) Evictions 2017-20
Percent Adults w/ College Degree (ACS 2019)	-0.494*** (-9.62)	-1.361*** (-11.46)	0.256*** (-11.46)
Population (ACS 2019)	0.0000160*** (5.32)	0.0000325* (2.21)	1.000* (2.21)
Population Change ACS 2014–ACS 2019	0.0890* (2.16)	0.326*** (9.03)	1.386*** (9.03)
Constant	-0.169*** (-3.90)	-1.525*** (-17.84)	70174
Observations	70174	70174	
R-squared	0.283		
RMSE	1.865		
F	3461.9		

t statistics in parentheses

** p<0.05. *** p<0.001.*

ACS = American Community Survey. IRR = incidence rate ratio. OLS = ordinary least squares. RMSE = root mean square error.

Sources: Building Indicator Project, UNHP, drawing on data from New York City; US Census; New York Housing Data Coalition

How do Affordable Housing Investments Break Cycles of Speculation and Distress?

The findings in exhibit 13 grimly illustrate challenging realities for lower-income communities of color in New York. Not only is more additional wealth generated (for others) from their homes, but the properties that generate this wealth and capital are more poorly maintained than comparable buildings and evict a higher proportion of their tenants. At the same time, although New York City has a long history of affordability challenges, housing speculation, and predatory ownership in different forms, it also has a long history of activist tenant and affordable housing movements, which have generated public support for relatively high levels of housing investment—approximately $19 billion from the city's own capital budget in the years from 1987 to 2018. Combined with federal and state resources, this support has resulted in approximately 17,000 annual affordable housing units produced or preserved, although at different levels of affordability (Schwartz, 2019).

Building on previous analyses, this research examined whether affordable housing investments were associated with better-quality housing and fewer speculative events.[7] To do so, data about financial and building characteristics were combined with data from New York University's Furman Center, which collected information about various kinds of affordable housing subsidies directed toward apartments. Because the article's primary concern was identifying forms of tenant,

[7] Calculating evictions by subsidy in comparison to other buildings is unfair because, by definition, affordable housing programs generally serve the lowest-income tenants, whereas other buildings, even in low-income areas, may have a range of tenant incomes.

community, or nonprofit ownership, it limits our analysis to certain subsidy streams and excluded other forms of affordability subsidies from the analysis, although for-profit affordable housing owners were also included.[8] Finally, it is worth remembering that only privately owned buildings are in the data, whether they are owned by a for-profit or nonprofit entity. Public housing is not included because it operates through a separate regulatory regime in which a lack of funding has resulted in severe housing maintenance issues and because the overall analysis and the BIP data set were focused on the private market.

When looking across all private rental housing units—including luxury apartment buildings and newly constructed apartments—there are about one-half to two-thirds fewer violations in subsidized apartments than in unsubsidized apartments, as shown in exhibit 14. On the one hand, one might assume that the presence of affordable investments should be associated with better housing quality because these investments were provided with public subsidies for the property's repair or for new construction. On the other hand, many of the buildings designated for such efforts had significant maintenance problems to begin with, and they house people with low incomes at affordable rents, meaning that there is no significant ongoing cash flow to devote to their maintenance. This situation speaks to the power of these investments and/or their community stewardship in maintaining not just affordability but also residential quality of life for tenants.

Exhibit 14

Total Violations Per Unit for Subsidized and Nonsubsidized Properties

Year	Total HPD Violations in Nonsubsidized Properties	Total HPD Violations in Subsidized Properties	Total HPD Violations	Total Units in Nonsubsidized Properties	Total Units in Subsidized Properties	Total Units	HPD Violations Per Nonsubsidized Unit (%)	HPD Violations Per Subsidized Unit (%)
2014	283,957	18,320	302,277	2,058,499	407,586	2,466,085	14	4
2015	357,974	23,517	381,491	2,046,335	419,750	2,466,085	17	6
2016	352,024	27,992	380,016	2,027,315	438,770	2,466,085	17	6
2017	380,879	36,724	417,603	2,014,032	452,053	2,466,085	19	8
2018	427,142	45,213	472,355	1,999,855	466,230	2,466,085	21	10
2019	449,411	52,588	501,999	1,986,910	479,175	2,466,085	23	11
2020	298,294	39,991	338,285	1,971,335	494,750	2,466,085	15	8

Data: BBLs (properties identified by borough-block-lot) with a selected subsidy vs. all other BBLs (removed BBLs with other forms of subsidy).
HPD = Department of Housing Preservation and Development.
Sources: Building Indicator Project, UNHP, drawing on data from New York City; NYU Furman Center Subsidized Housing Information Project

Adding regressions that factor in community conditions, such as race and poverty, demonstrates that these subsidies are associated with significantly fewer violations. That is, when compared to unsubsidized buildings in similar communities, units with affordable housing subsidies still are shown to have significantly fewer violations (exhibit 15). For example, BBLs with at least one subsidy in 2014–15 overall have 0.086 HPD violations per unit, whereas BBLs without a

[8] Programs included in the analysis are Section 202/8, Section 221d(3) and Section 221d(4) Mortgage Insurance, Section 223(f), Article 8A/HRP, LAMP – HDC, LIHTC 4%, LIHTC 9%, Multi-Family Program, Mitchell-Lama, Neighborhood Entrepreneur Program, Neighborhood Redevelopment Program, the Participation Loan Program, the Project Rental Assistance Contract, Project-Based Section 8, Section 8/RAD, TPT, and LIHTC Year 15, as well as those marked "Other HPD, HUD, and HUD Project-Based Rental Assistance."

subsidy have 0.326 violations per unit in the same period, a difference of about -0.24 violations per unit (exhibit 16). This calculation aligns with the model, in which the coefficient for subsidy in 2014–15 is -0.653 (larger than the raw difference). In other words, holding all neighborhood characteristics constant, a property with a subsidy in 2014–15 is associated with a 0.653 decrease in HPD violations per unit. When controlling for neighborhood context and reporting violations per unit, the analysis is not able to account for factors such as unit size, although community-level controls may address these issues, and recent analyses (Duranti-Martínez and Greenberg, 2023) show that properties matched on price and maintenance quality that receive acquisition rehabilitation subsidies have three times fewer maintenance violations versus comparable buildings sold to another owner without a subsidy.

Exhibit 15

Subsidy and HPD Violations, 2014–15

OLS Regressions	Dependent Variable—HPD Violations Per Unit 2014–15		
	(1)	(2)	(3)
Subsidy 2014–15	-0.167*** (-6.39)	-0.653*** (-24.48)	-0.673*** (-23.26)
Bronx		0.181*** (5.48)	0.208*** (5.43)
Brooklyn		0.095*** (4.25)	0.122*** (4.26)
Queens		-0.188*** (-6.76)	-0.163*** (-4.96)
Upper Manhattan		0.072* (2.46)	0.093** (2.78)
Percent Poverty (ACS 2019)		0.309*** (3.52)	0.321*** (3.38)
Percent Black/African-American (ACS 2019)		1.089*** (30.36)	1.095*** (28.48)
Percent Hispanic/Latino (ACS 2019)		0.869*** (19.70)	0.865*** (18.17)
Percent Adults w/ College Degree (ACS 2019)		-0.209*** (-3.90)	-0.218*** (-3.53)
Population (ACS 2019)		0.000** (2.85)	0.000** (3.08)
Population Change ACS 2014–ACS 2019		-0.117** (-3.17)	-0.154*** (-3.36)
Percent Rent Change ACS 2014–ACS 2019			0.034 (0.75)
Constant	0.557*** (86.84)	0.126* (2.24)	0.085 (1.33)
Observations	65,875	65,860	57,241
R-squared	0.0006	0.0668	0.0591
Adjusted R-squared	0.0006	0.0666	0.0589
RMSE	1.596	1.543	1.643
F	40.826	428.201	299.613

t statistics in parentheses

** p<0.05. ** p<0.01. *** p<0.001.*

ACS = American Community Survey. HPD = Department of Housing Preservation and Development. OLS = ordinary least square. RMSE = root mean square error.

Sources: Building Indicator Project, UNHP, drawing on data from New York City; US Census; NYU Furman Center Subsidized Housing Information Project

Exhibit 16

Affordable Subsidy and HPD Violations

Subsidized BBLs 2014–15	Total HPD Violations 2014–15	Total Units	Total Violations Per Unit 2014–15
Yes	43,519	504,008	0.086
No	640,249	1,962,077	0.326
Total	683,768	2,466,085	0.277

BBLs = properties identified by borough-block-lot. HPD = Department of Housing Preservation and Development.

Note: Subsidized BBLs include properties that had an active subsidy in either year.

Sources: Building Indicator Project, UNHP, drawing on data from New York City; NYU Furman Center Subsidized Housing Information Project

Removing Buildings from Cycles of Speculation

Subsidized properties not only have better maintenance quality but are also less likely to experience a debt increase or spike in sales value when compared to all other properties. Overall, about 1.12 percent of units with a subsidy in 2016–17 had a speculative event in the same period. About 3.14 percent of units without a subsidy in 2016–17 had a speculative event in the same period, as shown in exhibit 17. These results show that buildings with affordable housing investments—at least while the subsidy is in place—are, as a whole, removed from the cycles of disinvestment and speculation that so negatively impact tenants and communities. This circumstance appears to be driven by a reduction in speculative sales because owners of affordable housing are less likely to resell for higher amounts, although some may take on additional debt that is channeled directly into property improvements.

Exhibit 17

Speculative Events and Subsidy

Total Units		Any Speculative Event 2016–17			
		Yes	No	Total	%
Subsidized 2016–17	**Yes**	2,908	256,190	259,098	1.12
	No	70,936	2,186,280	2,257,216	3.14
	Total	73,844	2,442,470	2,516,314	

Sources: Building Indicator Project, UNHP, drawing on data from New York City; NYU Furman Center Subsidized Housing Information Project

Discussion

These analyses describe the costs of speculation to tenants and BIPOC communities and the power of affordable housing investments to promote tenant quality of life. In essence, the article finds that the greatest increases in landlord wealth are derived from buildings in the communities of color where tenants receive the lowest incomes and that buildings generating the greatest added wealth also hold the most harm for their tenants. It also finds that affordable housing investments provide far superior living standards and remove buildings from cycles of speculation and disinvestment.

Several implications can be derived from these findings. First, the finding that steeper increases in sales price and higher increases in debt were associated with more evictions speaks to the general need to provide tenant assistance and rental protections, both to aid tenants directly and to reduce the incentive to speculate, making it more difficult to displace longer-term tenants. Policies that can achieve these goals include extensions of effective rental relief funds, good-cause eviction protections, right-to-counsel initiatives, harassment protections, and similar measures.

Second, this research shows how community development investments created better-maintained properties and removed them from cycles of speculation. Broadly, investments at the federal level in affordable housing—which have declined significantly over time—can be used to acquire and rehabilitate rental housing, and advocates have called for their increase (LISC, 2023), including affordable housing programs targeted to provide flexible acquisition resources to mission-based housing organizations, such as the Housing Investment Fund. Tenant, nonprofit, and community ownership, including community land trusts, mutual housing associations, and limited-equity cooperatives, can be particularly beneficial to residents. Tenant Opportunity to Purchase Act (TOPA) and Community Opportunity to Purchase Act (COPA) policies may also be effective vehicles for this goal when paired with significant acquisition funding and support for ongoing organizing, capacity-building for nonprofit developers and technical and legal assistance to help tenants and community partners navigate the purchase and rehabilitation process. TOPA has a 40-year track record of preventing displacement and preserving affordable housing in Washington, D.C. San Francisco passed COPA in 2019, and Massachusetts and New York are considering statewide TOPA legislation, whereas Berkeley, Los Angeles, Oakland, New York City, and Minneapolis are exploring local opportunity-to-purchase policies (Duranti-Martínez and Greenberg, 2023).

Third, both speculative purchases and speculative finance were associated with poor housing maintenance. Advocates have called for increased code enforcement focused on poorly maintained portfolios and owners with histories of neglecting properties, both to improve tenant quality of life and potentially disincentivize speculators from deferring maintenance as a profit-making strategy. Code enforcement can create escalating civil penalties for deferred maintenance and tenant harassment, and it may involve receivership programs to assign property management of highly distressed buildings to a third-party administrator. Such enforcement programs could focus on investor owners and large property owners with the worst impact on communities. In these cases, tenant organizing is a valuable tool that can leverage code enforcement policies and promote tenant self-determination, and they may also require public and private support.

Fourth, the fact that increasing debt was a leading signal of maintenance quality problems and evictions suggests not only that financing is not generally being directed toward property improvements, but also that it may in fact be harming tenants because greater mortgage payments take up revenue streams that might otherwise be used for repairs and maintenance. A policy implication of this finding is to examine mechanisms to ensure that greater debt taken out on rental housing results in improvements for tenants and that lenders should be held accountable, as other investors are, for the quality of the properties on which loans are placed. For example, through the Community Reinvestment Act (CRA), there are incentives for responsible lending to rental housing and regulation of investments in housing that receive CRA credit. Currently, as long as a rental housing mortgage is provided in a low- to moderate-income census tract and to a building with lower-income tenants, that mortgage is often assumed to be a community reinvestment. The findings in this article imply that these investments do not always benefit tenants—suggesting that CRA commitments should incentivize mortgage lending in a manner that does not incentivize displacement or harm for tenants. Mortgage lending should include transparent benchmarking of expense minimums that are consistent with safe housing in all loan underwriting and clear processes for holding landlord borrowers accountable when they fail to responsibly steward the rental housing against which the mortgage was originated. Another mechanism to ensure that multifamily mortgage lending promotes safe, stable, and affordable housing is through Federal Housing Finance Agency (FHFA) regulation of the government-sponsored enterprises (GSEs), such as Freddie Mac and Fannie Mae. Over the past decade, these two GSEs have become major lenders in the rental market, and they have recently come under scrutiny for financing provided to large private equity landlords (Vogell, 2022); on this subject, the Biden Administration recently released a call to understand how FHFA might promote tenant protections on future loans backed by Fannie Mae and Freddie Mac, an exploration that this research would support.

Acknowledgments

The authors would like to thank anonymous *Cityscape* reviewers, internal reviewers within LISC Policy, Housing, and New York State divisions, and external reviewers at the Community Service Society, Banana Kelly Community Improvement Association, New Economy Project, Center for Popular Democracy, TakeRoot Justice, Southside HDFC, and Cooper Square Committee. Jim Buckley, executive director at UNHP, provided essential guidance in the project.

Authors

David M. Greenberg is vice president of community research and impact at the Local Initiatives Support Corporation (LISC), where Julia Duranti-Martínez is a senior program officer for community research and impact and Francisca Winston is a director of data and analytics. Spenser Anderson worked at LISC at the time of research. Jacob Udell is director of research and data at the University Neighborhood Housing Program (UNHP), where Caroline Kirk worked at the time of research. Richard D. Hendra is a part-time assistant professor at the New School. For more information, please contact David Greenberg at dgreenberg@lisc.org.

References

Bocian, Debbie Gruenstein, Wei Li, and Keith S. Ernst. 2010. “Foreclosures by Race and Ethnicity,” *Center for Responsible Lending*.

Brand, David. 2021. “Every NYC Tenant Has Access to Housing Court Lawyer, as Eviction Protections Near End,” *City Limits*, November 18. https://citylimits.org/2021/11/18/every-nyc-tenant-has-right-to-housing-court-lawyer-as-eviction-protections-near-end/.

Desmond, Matthew. 2016. *Evicted: Poverty and Profit in the American City*. New York, NY: Crown.

Duranti-Martínez, Julia, and David M. Greenberg. 2023. *Stable Homes and Resident Empowerment: Implementing Effective Tenant and Community Opportunities to Purchase Programs*. Local Initiatives Support Corporation. https://www.lisc.org/our-resources/resource/stable-homes-and-resident-empowerment/.

Gao, Zhenyu, Michael Sockin, and Wei Xiong. 2020. "Economic Consequences of Housing Speculation," *The Review of Financial Studies* 33 (11): 5248–5287.

Glaeser, Edward L. 2013. "A Nation of Gamblers: Real Estate Speculation and American History," *American Economic Review* 103 (3): 1–42.

Hornbach, Celeste, Oksana Mironova, Samuel Stein, and Jacob Udell. 2020. "Corporate Windfalls or Social Housing Conversion? The Looming Mortgage Crisis and the Choices Facing New York." Press release. New York, NY: Community Service Society. https://www.cssny.org/news/entry/new-css-report-corporate-windfalls-or-social-housing-conversions.

Local Initiatives Support Corporation (LISC). 2023. *LISC's Policy Priorities 2023–2024*. Local Initiatives Support Corporation. https://www.lisc.org/policy/policy-priorities/.

Marcuse, Peter. 1985. "Gentrification, Abandonment, and Displacement: Connections, Causes, and Policy Responses in New York City," *Washington University Journal of Urban and Contemporary Law* 28 (1): 195–240.

Mironova, Oksana, Samuel Stein, Celeste Hornbach, and Jacob Udell. 2022. *Pathways to Social Housing in New York: 20 Policies to Shift from Private Profit to Public Good*. New York, NY: Community Service Society.

New York University, Furman Center. 2019. *Trends in New York City Housing Court Eviction Filings*. New York, NY: NYU Furman Center. https://furmancenter.org/research/publication/trends-in-new-york-city-housing-court-eviction-filings.

Raymond, Elora Lee, Ben Miller, Michaela McKinney, and Jonathan Braun. 2021. "Gentrifying Atlanta: Investor Purchases of Rental Housing, Evictions, and the Displacement of Black Residents," *Housing Policy Debate* 31 (3–5): 818–834.

Schwartz, Alex. 2019. "New York City's Affordable Housing Plans and the Limits of Local Initiative," *Cityscape* 21 (3): 355–388.

Seymour, Eric, and Joshua Akers. 2021. "'Our Customer is America': Housing Insecurity and Eviction in Las Vegas, Nevada's Postcrisis Rental Markets," *Housing Policy Debate* 31 (3–5): 516–539.

Vogell, Heather. 2022. "When Private Equity Becomes Your Landlord," *ProPublica*, February 7. https://www.propublica.org/article/when-private-equity-becomes-your-landlord.

Commentary: Evidence-Based Policymaking to Address the Affordable Housing Crisis: The Potential of Local Data

Karen Chapple
University of Toronto and University of California, Berkeley

Abstract

Many housing market processes remain invisible because of the lack of comprehensive data and systematic research to create an evidence base. The articles in this section shed light on the relationship between investment activity and rental housing markets, opening up new avenues for research via strategic data linkage and providing much-needed evidence to support the preservation of affordable housing stock.

Introduction

For decades, housing research has drawn largely from a rich but limited collection of datasets put out by the federal government, such as the census, American Community Survey, American Housing Survey, Survey of Construction, program participation statistics, and Home Mortgage Disclosure Act (HMDA) data, augmented by local data, such as tax assessor data, zoning data, or building code violations. Researchers have gradually built a robust evidence base from these sources about federal programs, homeownership, housing conditions, housing starts, and related topics.

However, gaps in knowledge abound. Existing research provides very little evidence about rents, renters, and landlords and how these data relate to specific building types. We have only a basic sense of capital flows in and out of housing stock. We can only guess about informal or precarious housing arrangements. Because many of these housing market processes remain invisible, policymakers lack a full picture of the housing affordability crisis.

The lack of comprehensive data and systematic research on these particular housing dynamics has arguably contributed to a dearth of policymaking, particularly at the federal level, that addresses the affordability of rental housing and the role of speculative capital, particularly in older housing stock. Most of the country's affordable rental housing consists of this older unsubsidized housing stock owned and operated by the private sector (Joint Center for Housing Studies, 2020). Thus, understanding the relationship between investment activity and rental housing markets could help set the stage for critical new regulations and affordable housing preservation programs.

In "Housing Speculation, Affordable Investments, and Tenant Outcomes in New York City," Greenberg et al. explore the relationship between sales prices, increases in mortgage debt, maintenance violations, and evictions, showing an association among speculation-like activity, building conditions, and tenant instability. Sales prices, repeat sales, and mortgage debt are not a perfect proxy for speculation because they do not reveal the motives or origins of the buyers. Researchers also know little about the relationship between debt and maintenance violations (which turns out to be positive, despite the common assumption that landlords assume debt to improve property conditions), but the patterns are powerful and convincing, with particularly significant relationships in neighborhoods with concentrations of high poverty and Black and Latino populations. A supplemental analysis reveals that subsidized buildings benefit from higher maintenance quality and less debt. As the authors note, the findings call for not just investing more in nonprofit housing (for example, through large-scale acquisition funds) but also a reexamination of how loans support building improvements and tenant stability, particularly in light of the role that the Community Reinvestment Act and other government-supported loan programs may be playing in these dynamics.

In "Assessing How Gentrification- and Disinvestment-Related Market Pressures Drive the Loss of Small Multi-Unit Housing in Chicago Neighborhoods," Duda, Smith, and Jiao similarly use parcel-level data, examining the loss of two- to four-unit buildings and the role of the local real estate market in their transition to either single-family homes or vacant lots. Given the national movement to support gentle density in the form of missing middle housing, this article makes an essential contribution by demonstrating a trend toward reducing density, at least in Chicago. The analysis is compelling and original due to its tracking of the evolution of existing housing stock (a feat only possible with historic assessor data) and linking to neighborhood market conditions. The fate of a building depends on its context. The findings support reinvestment in areas of population loss and legislation to support preservation and make conversion to single-family homes more difficult.

These excellent contributions help open the door to many new avenues of research, particularly through strategic data linkage. For example, the actual renters are largely absent from both analyses (except evicted households). What if researchers could link buildings to tenants via rental registries; consumer credit panels, such as Equifax; or consumer reference datasets, such as Infutor? Most displacement occurs with no formal notice of eviction but rather through various forms of landlord harassment and disinvestment (Marcuse, 1986). Data on renter-occupied households would enable the examination of residential turnover, both as an intervening variable that may affect debt and building code violations (as seen in the Greenberg et al. analysis) and as an outcome of both speculation and building conversions. It would also allow linking to other types of administrative records—for example, to identify health impacts.

On the owner and investor side, understanding motives and outcomes in more detail is critical to designing more effective legislation to deter profiteering from housing. Cities such as Chicago and Portland are enacting zoning regulations that hinder the conversion of multifamily buildings, but these regulations may present insufficient obstacles given the readily available capital and lure of profit for some actors. Faced with the rapid increase in investor-owned housing, governments need to enact some new guardrails to protect renters; however, policymakers do not yet understand enough about the benefits and costs of investor ownership—or even how widespread it is—to regulate it effectively.

These articles raise questions about how best to prevent the loss of affordable housing before it occurs. Neither analysis attempts to predict where speculation, eviction, or conversion will occur; however, it would not take much more analysis to identify which areas are most at risk for future activity because of the convincing evidence presented. Chicago is where such early warning systems started when the Center for Neighborhood Technology began experimenting with parcel data to track disinvestment in the 1980s. At the very least, such predictions can help validate local perceptions and empower communities to organize for change (Chapple and Zuk, 2016).

The studies in this section and others in this volume provide an opportunity to refocus policymakers' attention on the affordable housing crisis and identify potential solutions. As new data and analyses become available, the critical role of existing unsubsidized but affordable housing is likely to become obvious. It will be important to connect the loss of such stock to the goals of Affirmatively Furthering Fair Housing because much displacement is occurring in urban neighborhoods with the most opportunity. Policymakers will likely provide more funding for affordable housing preservation in the form of acquisition (such as the Small Sites Program in San Francisco) and tenant purchasing (such as the Tenant Opportunity to Purchase Act in Washington, D.C.).

Broader implications arise out of this body of research. As society wrestles with how to address the long-term impacts of systemic racism and dismantle its root causes, analyses such as Greenberg et al. on the connection between speculation and communities of color will prove invaluable. Speculation, building maintenance, community stability, and the loss of affordable housing stock also have extensive fiscal implications that need better understanding to build momentum for policy interventions. In other words, researchers need to mobilize knowledge about the costs of speculation or displacement, just as the Urban Institute has done with the cost of segregation (Acs et al., 2017).

Finally, it is worth noting that these studies depend on the availability of fine-grained local data, with minimal use of data from federal agencies. Many studies on the affordability challenges of U.S. high-cost metropolitan areas provide evidence of the effectiveness of interventions, such as just cause eviction policies; however, such case studies do not translate well in weaker market contexts, even when they experience similar issues, so these cases may not be effective in moving policy at the state or federal level. Creating fine-grained datasets at the national level, such as a national parcel dataset linked to administrative data on households, would help to remedy the growing chasm in housing policy and programs between strong and weak market regions.

Author

Karen Chapple is Professor of Geography and Planning and Director of the School of Cities at the University of Toronto, and Professor Emerita of City and Regional Planning at the University of California, Berkeley.

References

Acs, Gregory, Rolf Pendall, Mark Treskon, and Amy Khare. 2017. *The Cost of Segregation*. Washington, DC: Urban Institute.

Chapple, Karen, and Miriam Zuk. 2016. "Forewarned: The Use of Neighborhood Early Warning Systems for Gentrification and Displacement," *Cityscape* 18 (3): 109–30.

Joint Center for Housing Studies of Harvard University. 2020. *America's Rental Housing 2020*. Cambridge, MA: Harvard University, Joint Center for Housing Studies.

Marcuse, Peter. 1986. "Abandonment, Gentrification, and Displacement: The Linkages in New York City." In *Gentrification of the City*, edited by Neil Smith and Peter Williams. London: Routledge: 153–77.

Commentary: Improving Housing Policy with Neighborhood Data

Leah Hendey
Elizabeth Burton
Kathryn L.S. Pettit
Urban Institute
National Neighborhood Indicators Partnership

Abstract

This volume demonstrates many past and potential applications of administrative data that inform and change housing policy. We identify three areas to enhance the use of local administrative data based on our experiences from the National Neighborhood Indicators Partnership: 1) collaborating with residents and community organizations to inform research questions and findings; 2) improving infrastructure around court records, zoning, and parcel data; and 3) integrating data across sectors, such as health, housing, education, and others. With cross-sector collaboration and investments in building community data capacity, researchers, advocates, foundations, the private sector, and governments at all levels can play a role in improving the availability and use of administrative data to inform housing policy to ensure all neighborhoods are places where people can thrive.

Introduction

In the 1990s, organizations in several communities, including the Center on Poverty and Community Development at Case Western Reserve University in Cleveland, developed an innovative approach to tracking changes at the neighborhood level using a variety of administrative data sources for the purpose of informing community action (Kingsley, Coulton, and Pettit, 2014).[1] These data provided more timely and granular information, allowing users to explore differences in trends across neighborhoods and develop a shared understanding of community conditions across sectors. The organizations came together with the Urban Institute to form the National Neighborhood Indicators Partnership (NNIP), which shares practices to accelerate progress on the ground and spread its approach to new communities. Since its formation, NNIP has specialized in

[1] Coulton (2008) cataloged a number of state and local administrative data sources that can be used for neighborhood-level indicators.

transforming local administrative data sources and helping local actors use the data to support their own priorities so that all neighborhoods are places where people can thrive.

The NNIP network has elevated its collective experience from the local partners to share insights nationally on programs and policy for housing. Among other projects, it has worked with parcel data to support government and nonprofit decisionmaking, layered many data sources to inform neighborhood stabilization, combined data to understand the effects of foreclosures on children and schools, and used mixed methods to explore displacement risk in recovering housing markets (Federal Reserve, 2011; Kingsley and Pettit, 2007; Pettit, Cohen, and Levy, 2019; Pettit and Comey, 2012). This issue of *Cityscape* demonstrates the range of housing policy issues that people can address by analyzing administrative data—including resident health, the impacts of flooding, investor speculation, displacement pressures on unsubsidized housing stock, evictions, and tenant outcomes.

This essay focuses on three areas in which the field can improve access and application of administrative data: (1) collaborating with residents and community organizations to inform research to advance more inclusive policies, (2) improving data infrastructure, and (3) continuing to innovate with data integration for cross-sector insights.

Collaborating with Residents and Community Organizations

A great deal of administrative data exist related to housing, and this volume demonstrates its application to inform housing policy. The field now needs to increase collaboration with residents and community organizations at all stages of the research process. Doing so can improve the research by ensuring that the focus and results are salient for community-driven priorities and presented in the format needed for advocacy. Such efforts can help put information into the hands of residents and communities historically marginalized from decisionmaking and enable them to advocate for the changes they seek.

NNIP's cross-site project, *Turning the Corner*, was formed after community organizations in recovering housing markets expressed concerns about displacement risk. In the resulting report, Pettit, Cohen, and Levy (2019) documented that a mixed methods approach involving analysis of quantitative data (including administrative data), interviews, and focus groups helped lift up a variety of perspectives, including those of renters and homeowners, long-term residents, newcomers, and people of different racial and ethnic backgrounds. Residents provided important insights for understanding the changes seen in quantitative data and helped researchers articulate their implications.

In "Assessing How Gentrification- and Disinvestment-Related Market Pressures Drive the Loss of Small Multi-Unit Housing in Chicago Neighborhoods," the team at DePaul University's Institute for Housing Studies (IHS) documented their analysis of parcel and sales data of two- to four-unit buildings in Chicago (Duda, Smith, and Jiao, 2024). The team worked closely with community organizations to understand their concerns about these properties and helped quantify what has been happening since the early 2010s. Community organizations raised concerns about displacement risk for tenants living in this largely unsubsidized affordable housing stock, first from foreclosures and then later from gentrification pressure around The 606 trail. IHS helped the organizations document the loss of this housing stock and worked with the Chicago Department of Housing as they formulated policies to reduce displacement (Burton, 2021). Based on these efforts, the Chicago City Council passed anti-deconversion ordinances in 2021 for two neighborhoods with increased displacement risk to prevent this stock from being converted into single-family houses. IHS collaborated with a coalition of community organizations, led by Elevated Chicago, to advocate for the Connected Communities Ordinance, which was passed by the Council in 2022 to limit deconversion in two- to four-unit buildings in markets with displacement pressure near transit and bring equitable transit-oriented development policies into the city's zoning code (Burton, 2023).

Although policy researchers are increasingly collaborating with residents and community organizations, more can be done to center residents with lived experience and co-lead research with community members. NNIP articulated several goals for improving the use of data to advance racial equity for its members and to acknowledge and address systemic harms (NNIP, 2021). These goals range from collaborating with residents on research projects to using data to focus on systems and highlighting assets of people and communities. Torres Rodríguez, et al. (2023) also offer guidance for quantitative researchers to incorporate community-engaged methods, which will require researchers to develop new skills and awareness. Doing so will improve the quality of the analyses of housing markets and conditions, which is critical as advocates, practitioners, and policymakers use them to set priorities and plan for action. More examples in a variety of contexts can help to strengthen the methods and result in nuanced findings amongst academics and policymakers.

Improving Data Infrastructure

Although tremendous improvements in the availability of housing data have occurred since NNIP formed in the mid-1990s, and the authors in this issue of *Cityscape* document the utility of these data for policy analysis, certain types of administrative data—namely court records, zoning data, and parcel data—remain very difficult to access and use within a jurisdiction and limit the potential for research and analysis.

Administrative court records are key to understanding the patterns and processes of civil legal actions like eviction and foreclosures, as illustrated by the article from Ellen, Lochhead, and O'Regan on evictions by property type in New York City. However, electronic court records are not universally available, and even when they are, often the most detailed (and relevant) information on causes and outcomes is in scanned images of documents uploaded to the docket records that are not machine readable, making the information unsearchable and hindering analysis. Pioneering efforts like that of Legal Services Corporation's Civil Court Data Initiative have scraped court records to facilitate their use in eviction tracking. Thomas et al. also suggest an approach using natural language processing to identify tenants' names and addresses from the images of court filings and digitizing the text for analysis and action. This could be scaled to other jurisdictions and potentially expanded to extract additional information about the cases from the images.

Improvements in new methods, technology, and collaborations are promising ways to fill in the civil justice information gaps in the short term and should be expanded. As an example, the Civil Justice Data Commons is a repository for civil legal data gathered from courts, legal service providers, and other civil law institutions available to researchers on a secure data platform (Georgetown Law, n.d.). However, legal decisions or explicit legislation may be needed to protect other stopgap data collection measures like scraping court websites (ACLU, 2023). Longer-term progress will require dedicated funding and technical assistance programs from state and federal governments to modernize the court data systems in ways that support transparency, data access, and policy analysis.

Although shapefiles are often available for the land uses and zones in a jurisdiction's zoning code, interpreting the code itself to understand the potential for new development and density has required manual review and coding. Early efforts have been made to pilot automating the collection of zoning data (Axelrod, Lo, and Bronin, 2023), and a group of cross-discipline researchers collaborating on the National Zoning Atlas housed at the Cornell University Legal Constructs Lab are working to digitize roughly 30,000 U.S. zoning codes (Cornell University, n.d.). The Atlas's research collaborators at the Urban Institute have been experimenting to leverage machine learning, along with surveys and manual review, to unlock zoning data for research and policy (Urban Institute, n.d.). The researchers' sector-spanning efforts merit investments in this key local policy lever, which can be used to expand or deter the supply and types of housing.

Despite being used often by NNIP partners and many of the authors in this issue of *Cityscape*, the accessibility of parcel data varies considerably across jurisdictions. In their study of the feasibility of a national parcel database, Abt Associates and Fairview Industries (2013) document that, although most counties' parcel data were publicly available, some counties charged fees or did not have the capacity to engage with the data request. Hopefully, a decade later, more jurisdictions have publicly available data, although we expect standardization across them remain a challenge and know that several NNIP partner cities require fees or the negotiation of data use agreements for parcel data and sales transactions. Small fees, even under $2,000, could be prohibitive for many nonprofits and community-based organizations.

In the case of parcel data, private, proprietary sources have arisen to fill some public-sector gaps—offering data for thousands of counties, making them particularly valuable sources for cross-city analysis. However, several barriers to their use exist, including cost, restrictions on use, and proprietary methodologies used to standardize the data or create indicators. The first barrier limits who can access the data and likely increases existing inequities in access to data. Restrictions on use are problematic because they typically limit the granularity of the information that can be shared publicly or even privately with a third party—which may be less of a concern for policy research but more of a concern for those intending to use data for planning, community investment, and collaborating with residents and community-based organizations. Proprietary methodologies can make it more difficult for the analyst to understand the underlying data quality and the biases of the data or indicators.

The civic sector can play a role in helping make proprietary data sources available in aggregated formats and tools like the Eviction Lab's ground-breaking data release in 2017. However, for jurisdictions that sell their assessor's and/or sales files exclusively to private firms, local researchers and advocates will need to organize to make the case for the public benefit of open data and press for new state and local policies that mandate the data be provided in an accessible format free of charge or at a reasonable cost.

Continuing Innovation with Data Integration

NNIP was founded on the belief that neighborhoods are important for equity because the places where people live affect their health, security, education, and economic success. Housing policy outcomes are affected by all these issues as well. Integrating data across issue areas helps us understand how these issues intersect and creates the potential for new collaborations across sectors for change. Integrating health and housing data might inspire a healthcare network to invest in affordable housing (Kaiser Permanente Insider, 2021) or school districts to understand the impact of housing mobility and affordability on student absenteeism (Deitrick et al., 2015).

In Cleveland, our NNIP partner at the Center on Poverty and Community Development at Case Western Reserve University (CWRU) matched its integrated data system containing 35 administrative data sources, CHILD, with their integrated property data in NEO CANDO to explore the effect living in distressed properties had on school readiness. Coulton, et al. (2016b) found living in poor-quality housing units or units that are tax-delinquent, owned by speculators, or in foreclosure can lead to lower literacy scores for kindergartners, creating achievement gaps before entering school. Living in such homes is associated with a high risk of elevated blood lead levels, child maltreatment, and residential instability, which all influence literacy scores (Coulton, et al., 2016a). The study found that living near distressed housing is also problematic, affecting children's ability to be healthy and ready to succeed. Kindergartners who lived within 500 feet of distressed properties had lower literacy scores than those living farther away. These findings prompted the City of Cleveland to take a proactive approach to addressing lead exposure by developing a rental registry and requiring inspections. The city also prompted education and early childhood officials to take more holistic views of school readiness and child health that incorporate understanding of the housing and neighborhoods where children live (Hendey, 2018). In this issue of *Cityscape*, the CWRU team continues their efforts to help local stakeholders use data to address lead exposure in Cleveland.

Systems like CHILD and NEO CANDO (CWRU, n.d.) are not built overnight, but they serve as a demonstration of the value of data integration to inform housing policy and connect it to outcomes in other domains that could help bring attention and resources to increase the availability of safe, affordable housing and improve outcomes for children and families. Actionable Intelligence for Social Policy is one organization helping states and other entities overcome the legal and political hurdles to building equitable integrated data systems through peer learning, technical assistance, and documentation of the payoff of these systems. State, federal, and local government investments can accelerate the spread of this advanced data infrastructure by enacting enabling legislation where needed and funding the development and maintenance of integrated data systems. Integrated data efforts are also often centered in service and people-focused agencies, so the field needs a continual reminder of the importance of housing quality and location to people's opportunity and wellbeing to justify the effort of collecting regular address information (Hendey, 2016).

Conclusion

A critical component of expanding the use of administrative data to improve housing policy is to ensure that all communities have the capacity to use data—that is, people living there can access and use data to inform efforts to understand and improve outcomes where they live (Hendey, et al. 2020). "People" refers to everyone—residents, philanthropists, and those in governments, nonprofits, or the private sector—and expanding the use of data can be done in a variety of ways (Hendey and Pettit, 2021). This capacity is not held equitably; historically, people of color and those with low incomes have had fewer opportunities to access data and build skills to use data to advocate for the community changes they seek.

Since the founding of NNIP in the mid-1990s, the field has made remarkable gains in community data capacity and in the other areas discussed in this essay—community-engaged research, data infrastructure, and integrated data. With coordinated efforts, researchers, advocates, foundations, the private sector, and governments at all levels can play a role in improving the availability and use of administrative data related to housing to ensure all neighborhoods are places where people can thrive.

Acknowledgments

The authors dedicate this call to action to use local administrative data to support improvements in housing policy and planning in memory of G. Thomas Kingsley, Senior Fellow at the Urban Institute and the founder of NNIP, who passed away in 2018. Tom was instrumental in the success of the network and spreading the philosophy of using administrative data for tracking and addressing neighborhood change, including housing conditions and markets. His mentorship to countless researchers and local practitioners brought many benefits to urban communities across the country. https://www.neighborhoodindicators.org/library/catalog/memoriam-g-thomas-kingsley

Authors

Leah Hendey is a principal research associate in the Metropolitan Housing and Communities Policy Center at the Urban Institute. She is experienced in transforming national and local administrative datasets to create neighborhood indicators and study neighborhood conditions. Since 2007, Hendey has been involved in Urban's work with the National Neighborhood Indicators Partnership and serves as its codirector. Elizabeth Burton is a research analyst in the Metropolitan Housing and Communities Policy Center at the Urban Institute. She also provides technical assistance to Promise Neighborhood grantees and helps manage the National Neighborhood Indicators Partnership. Kathryn L.S. Pettit is a senior fellow in the Metropolitan Housing and Communities Policy Center at the Urban Institute codirects the National Neighborhood Indicators Partnership.

References

Abt Associates, Inc., and Fairview Industries. 2013. *The Feasibility of Developing a National Parcel Database: County Data Records Project Final Report*. Washington, DC: US Department of Housing and Urban Development. https://www.huduser.gov/portal/publications/polleg/feasibility_natl_db.html.

American Civil Liberties Union (ACLU). 2023. "Federal Judge Rules That a Ban on Automated Data Collection Could Violate the First Amendment." Press release. Columbia, SC: American Civil Liberties Union. https://www.aclu.org/press-releases/federal-judge-rules-ban-automated-data-collection-could-violate-first-amendment.

Axelrod, Judah, Lydia Lo, and Sara C. Bronin. 2023. *Automating Zoning Data Collection*. Washington, DC: Urban Institute. https://www.urban.org/research/publication/automating-zoning-data-collection.

Burton, Elizabeth. 2023. "New Policies to Preserve Two-to-Four Unit Housing in Chicago," National Neighborhood Indicators Partnership, June 5. https://www.neighborhoodindicators.org/library/stories/new-policies-preserve-two-four-unit-housing-chicago.

———. 2021. "Institute for Housing Studies Guides Local Policies to Preserve Two-to-Four-Unit Housing," National Neighborhood Indicators Partnership, December 9. https://www.neighborhoodindicators.org/library/stories/institute-housing-studies-guides-local-policies-preserve-two-four-unit.

Case Western Reserve University (CWRU), The Center on Poverty and Community Development. n.d. "Data Systems: Harnessing Big Data for Social Good." https://case.edu/socialwork/povertycenter/data-systems.

Cornell University. n.d. "National Zoning Atlas." https://www.zoningatlas.org/.

Coulton, Claudia J. 2008. *Catalog of Administrative Data Sources for Neighborhood Indicators*. Washington, DC: Urban Institute. https://www.urban.org/research/publication/catalog-administrative-data-sources-neighborhood-indicators.

Coulton, Claudia, Robert L. Fischer, Francisca Garcia-Cobian, Richter Seok-Joo Kim, and Youngmin Cho. 2016a. *Housing Crisis Leaves Lasting Imprint on Children in Cleveland*. Chicago, IL: MacArthur Foundation. https://www.macfound.org/media/files/hhm_brief_-_housing_crisis_children_in_cleveland.pdf.

Coulton, Claudia J., Francisca Richter, Seok-Joo Kim, Robert Fischer, and Youngmin Cho. 2016b. "Temporal Effects of Distressed Housing on Early Childhood Risk Factors and Kindergarten Readiness," *Children and Youth Services Review* 68: 59–72. https://www.sciencedirect.com/science/article/abs/pii/S0190740916301943.

Deitrick, Sabina, Feifei Ye, Joshua Childs, and Caiyan Zhang. 2015. *Connecting People and Place—Improving Communities through Integrated Data Systems: Chronic School Absenteeism in Public Schools in Pittsburgh, PA*. Pittsburgh, PA: University of Pittsburgh. https://www.neighborhoodindicators.org/sites/default/files/publications/POLICY%20BRIEF%20University%20of%20Pittsburgh.pdf.

Duda, Sara, Geoff Smith, and Yiwen Jiao. 2024. "Assessing How Gentrification and Disinvestment-Related Market Pressures Drive the Loss of Small Multiunit Housing in Chicago Neighborhoods," *Cityscape* 26 (1): 129–152.

Federal Reserve. 2011. *Putting Data to Work: Data-Driven Approaches to Strengthening Neighborhoods*. Washington, DC: Board of Governors of the Federal Reserve System. https://www.federalreserve.gov/publications/putting-data-to-work.htm.

Georgetown Law, Institute for Technology Law & Policy. n.d. "Civil Justice Data Commons." https://www.law.georgetown.edu/tech-institute/initiatives/georgetown-justice-lab/civil-justice-data-commons/.

Hendey, Leah. 2018. "Groundbreaking Studies Help Prevent Lead Exposure for Children in Cleveland," National Neighborhood Indicators Partnership, June 1. https://www.neighborhoodindicators.org/library/stories/groundbreaking-studies-help-prevent-lead-exposure-children-cleveland.

———. 2016. *Using Integrated Data to Improve Communities: Lessons from a Cross-site Project*. Washington, DC: Urban Institute. https://www.urban.org/research/publication/using-integrated-data-improve-communities-lessons-cross-site-project.

Hendey, Leah, and Kathryn L.S. Pettit. 2021. *Envisioning a New Future: Building Trust for Data Use*. Washington, DC: Urban Institute. https://www.urban.org/research/publication/envisioning-new-future-building-trust-data-use.

Hendey, Leah, Kathryn L.S. Pettit, Jake Cowan, and Marcus Gaddy. 2020. *Investing in Data Capacity for Community Change*. Washington, DC: Urban Institute. https://www.urban.org/research/publication/investing-data-capacity-community-change.

Kaiser Permanente Insider. 2021. "Kaiser Permanente Invests $5 Million in Loan Fund to Catalyze Affordable Housing Along the Purple Line Corridor." News release. Rockville, MD: Kaiser Permanente. https://insider.kaiserpermanente.org/kaiser-permanente-invests-5-million-in-loan-fund-to-catalyze-affordable-housing-along-the-purple-line-corridor/.

Kingsley, G. Thomas, Claudia J. Coulton, and Kathryn L. S. Pettit. 2014. *Strengthening Communities with Neighborhood Data*. Washington, DC: Urban Institute. https://www.urban.org/strengtheningcommunities.

Kingsley, G. Thomas, and Kathryn L.S. Pettit. 2007. *Data and Decisions: Parcel-Level Information Changing the Way Business Gets Done*. Washington, DC: The Brookings Institution. https://www.neighborhoodindicators.org/sites/default/files/publications/UMINNIPReport112107.pdf.

Legal Services Corporation. n.d. "Civil Court Data Initiative." https://civilcourtdata.lsc.gov/.

National Neighborhood Indicators Partnership (NNIP). 2021. "NNIP's Goals to Improve Use of Data in Advancing Racial Equity." https://www.neighborhoodindicators.org/nnips-goals-improve-use-data-advancing-racial-equity.

Pettit, Kathryn L.S., Mychal Cohen, and Diane K. Levy. 2019. *Turning the Corner: Lessons from Five Cities on Displacement Risk in Changing Neighborhoods*. Washington, DC: Urban Institute. https://www.urban.org/research/publication/turning-corner-lessons-five-cities-displacement-risk-changing-neighborhoods.

Pettit, Kathryn L.S., and Jennifer Comey. 2012. *The Foreclosure Crisis and Children: A Three-City Study*. Washington, DC: Urban Institute. https://www.urban.org/research/publication/foreclosure-crisis-and-children-three-city-study.

Torres Rodríguez, Sonia, Justin W. Morgan, Lynden Bond, Shubhangi Kumari, and Kassandra Martinchek. 2023. *Increasing the Rigor of Quantitative Research with Participatory and Community-Engaged Methods*. Washington, DC: Urban Institute. https://www.urban.org/research/publication/increasing-rigor-quantitative-research-participatory-and-community-engaged.

Urban Institute. n.d. "Land Use." https://www.urban.org/research-area/land-use.

Evictions

Analyzing the Effect of Crime-Free Housing Policies on Completed Evictions Using Spatial First Differences

Toward a National Eviction Data Collection Strategy Using Natural Language Processing

Eviction Practices in Subsidized Housing: Evidence From New York State

Commentary: Using Eviction Court Records to Inform Local Policy

Analyzing the Effect of Crime-Free Housing Policies on Completed Evictions Using Spatial First Differences

Max Griswold
Lawrence Baker
Sarah B. Hunter
Jason Ward
RAND Corporation

Cheng Ren
University at Albany, State University of New York

Abstract

Crime-free housing policies attempt to prevent crime within rental properties by enrolling property owners in a local crime-free housing program, which subsequently permits landlords to use a supplemental lease agreement stating certain activities that could lead to a tenant being evicted. Building on third-party policing strategies, crime-free housing policies are widely prevalent across the United States, with an estimated 2,000 jurisdictions adopting them since 1992. Despite the widespread adoption of such policies, no previous research has identified their effect on evictions.

This article analyzes the effect of crime-free housing policies on evictions in four locations (Fremont, Hayward, Riverside, and San Diego County) in California. The authors obtained geocoded data on evictions through Public Records Act requests submitted to sheriff's departments in California seeking writs of execution, with additional Public Records Act requests submitted to municipalities to obtain policy implementation information, including the location of certified multifamily property units. To identify a causal effect, a spatial first differences design was used to exploit variation between U.S. Census Bureau block groups with and without certified properties.

Abstract (continued)

The results show that block groups with crime-free housing certified rental units have lower per capita income and larger proportions of Black and Latin/Hispanic populations. In each location, model results indicate that crime-free housing policies significantly increase evictions. Considered jointly, the findings suggest that crime-free housing policies increase evictions by 24.9 percent (95-percent confidence interval: 15.1–34.6 percent) within treated block groups. Given the harm that evictions cause and the governmental costs of eviction proceedings, municipalities across the United States should weigh the benefits of crime-free housing policies against increases in evictions. In addition, given the close policy similarities between crime-free housing policies, criminal activity nuisance ordinances, chronic nuisance ordinances, and the one-strike policy in public housing, these results indicate that policymakers should consider revising the existing policies as a potential means to reduce evictions nationally.

Introduction

Evictions represent a growing problem in the United States. Between 2000 and 2018, court filings for evictions increased by 21.5 percent to 3.6 million cases annually (Gromis et al., 2022). Housing displacement is a critical pathway to homelessness, causes physical and mental health problems, and exacerbates food insecurity for children (Collinson et al., 2023; Hatch and Yun, 2020; Leifheit et al., 2020; Vásquez-Vera et al., 2017). Evictions disproportionately affect low-income tenants and minority populations, with Black women at the greatest risk of an eviction (Hartman and Robinson, 2003; Hepburn, Louis, and Desmond, 2020). Those costs are not borne solely by individuals; evictions cause broader community harm, including increased emergency room use, hospitalizations, homelessness, and spending on social services (Collinson et al., 2023). The United States outpaces global peers in terms of the percentage of renters evicted, with 6.1 percent of renters in the United States facing eviction proceedings in 2016, compared with less than 2 percent of renters in other Organisation for Economic Co-Operation and Development (OECD) countries (OECD, 2020).

Some local policies nominally motivated by crime prevention may directly increase evictions, such as chronic nuisance ordinances, criminal activity nuisance ordinances, and crime-free housing policies (CFHPs). Those policies penalize property owners who do not evict tenants engaged in certain activities specified by the ordinance. Such ordinances are widely prevalent across the United States: an estimated 2,000 municipalities have a criminal activity nuisance ordinance or crime-free housing policy (Ramsey Mason, 2018). However, to date, limited research has examined the effect of those policies on evictions.

Crime-free housing policies are particularly important to evaluate, given the enforcement mechanism the policy uses to attempt to prevent crime. CFHPs are municipal programs that certify multifamily housing units as crime-free once property owners attend a training offered by law enforcement agencies, make specific physical modifications to their units, and add a supplemental lease addendum to their standard rental agreement. The supplemental lease addendum is the policy's primary enforcement mechanism, permitting property owners to evict a tenant for engaging in or facilitating any criminal behavior (Archer, 2019).[1]

Although CFHPs use evictions as a tool to enforce the policy, calculating the magnitude of the effect is critical because it could inform municipalities' choice to adopt or maintain CFHPs. If CFHPs substantially increase evictions, municipalities will need to weigh the desired policy outcomes of CFHPs against the subsequent social and governmental costs of additional evictions.

This article estimates the effect of crime-free housing policies on evictions, using a newly constructed database on writs of execution across the state of California between 2017 and 2021. To analyze that effect, this study uses a spatial first differences research design (Druckenmiller and Hsiang, 2018) that allows for the estimation of causal effects using cross-sectional data containing small area observations. The data used in the analysis—writs of execution records and CFHP implementation information—were obtained using Public Records Act requests submitted to municipalities and government agencies throughout the state of California. The present spatial first difference design uses block-level variation in the number of CFHP-certified rental units to identify the effect of crime-free housing policies on evictions in four locations in 2019 (Fremont, Hayward, Riverside, and San Diego County). The findings indicate that neighborhood blocks containing CFHP-certified rental units have a significantly higher number of evictions, increasing the average amount by 24.9 percent (95-percent confidence interval: 15.1–34.6 percent) across studied locations.

Challenges in Obtaining Eviction Data in Municipalities

Evaluating the effect of existing policies on evictions is difficult due to the lack of reliable, systematic data on evictions and housing policies at the local level (Goplerud and Pollack, 2021). Although some data on evictions exist, such as the database maintained by The Eviction Lab at Princeton University, those measures frequently are collected only at the state or county level, limiting the use of statistical methods to evaluate the effect of local policymaking on evictions (Gromis et al., 2022). Further, those eviction databases typically rely on measures obtained from court filings, which have substantial limitations. For instance, court filing data frequently do not contain the outcome of a case (i.e., whether the filing led to an eviction or was the cause of the eviction), may contain substantial duplicate counts due to landlords using serial filings to collect rent, or could be unavailable in a jurisdiction due to records being sealed to protect tenants (Garboden and Rosen, 2019; Goplerud and Pollack, 2021; Porton, Gromis, and Desmond, 2021).

Obtaining records on evictions in cities and localities, particularly on completed evictions, could provide the evidence needed to evaluate the effects of local policymaking. One such measure of evictions that is available across jurisdictions is writs, which are orders issued by courts to landlords following both an unlawful detainer action decided in favor of a landlord and a notice to vacate provided to a tenant. Writs permit landlords to pursue a "lockout" (forcible removal of a

[1] The definition of *criminal behavior* is not explicitly defined in the lease addendum, although the addendum notes that "proof of violation shall not require a criminal conviction, but shall be by a preponderance of the evidence" (ICFA, n.d. b.).

tenant by a county sheriff) if the renter does not voluntarily vacate the unit.[2] To do so, landlords provide the writ to a sheriff's office to schedule a lockout; on the scheduled date, a sheriff will remove the tenant from the rental unit. Those actions—scheduled and completed writs—are recorded by the sheriff's departments and can be obtained through Public Records Act requests.

Writs of execution records contain benefits and limitations as a measure of evictions compared with alternative records, such as eviction notices or court filings for unlawful detainer proceedings. Whereas eviction notices and filing records may or may not have led to a completed eviction, writ records correspond directly to known completed evictions. Conversely, because writs are issued only when a tenant has not voluntarily vacated a unit, writ records will necessarily underestimate the total number of evictions occurring in each location. However, underestimation is an issue with all eviction measures, including notices and filings, because informal or illegal evictions are not recorded in administrative records. In addition, writs of execution records may be the only available measure of evictions within smaller geographies, such as municipalities or neighborhood blocks. For example, most records on eviction filings in California have been sealed due to state law (AB2819),[3] making writs of execution one of the only measures available for evaluating municipal policies.

Policy Background and Components

Crime-free housing policies originated in a program started by the Mesa Arizona Police Department in 1992 with the stated purpose of reducing "spiraling crime rates in the city's numerous apartment communities" (Zehring, 1994). CFHPs attempt to achieve that goal by enrolling local landlords in the program, which entails three primary components: trainings provided to landlords on compliance with the program; requirements for landlords to modify their rental units to comply with crime prevention through environmental design (CPTED) standards; and a supplemental lease agreement for landlords to include as part of their standard lease, stating that tenants can be evicted from their unit if they are suspected of any criminal activities (Archer, 2019; Ramsey Mason, 2018).

CFHPs have extended to other jurisdictions through the efforts of the International Crime Free Association (ICFA), a nonprofit organization that produces model policy documentation for CFHPs and markets the policy to law enforcement agencies (ICFA, n.d.a.). To implement the policy in additional jurisdictions, the ICFA conducts 3-day conferences with law enforcement officers to train them on the implementation of CFHPs in their local jurisdictions. The organization also provides agencies with instruction materials for training landlords, CPTED inspection forms, signs to display outside certified rental units, marketing materials, program logos, and supplemental lease agreement language.

The ICFA describes the program as using a three-phase approach to eliminate crime in multifamily housing units (ICFA, n.d.a.). First, the policy aims to train landlords and property managers on

[2] The specific name of the writ corresponding to an eviction lockout can vary depending on the jurisdiction. For example, Washington State refers to them as "writs of restitution," whereas California uses both "writ of possession" and "writ of execution" to refer to eviction lockouts.

[3] "Unlawful Detainer Proceedings," CA AB2819, 2015–2016 Regular Session (CA, 2016). https://legiscan.com/CA/text/AB2819/id/1429026.

compliance with the program. Trainings to landlords are taught by law enforcement officers and typically consist of 11 modules given over an 8-hour workshop. Modules cover several topics, which detail the CPTED modifications landlords will need to make to their unit; how to screen tenant applications for a history of criminal justice involvement; how to use and enforce the supplemental lease addendum; and how landlords should communicate with law enforcement (Western Regional Chapter of the International Crime Free Association/Crime Free & Partners, 2009). To maintain compliance with the program, landlords and property managers must attend the training, usually biannually.

Second, after landlords and property managers attend a CFHP training, they must make physical modifications to rental properties to meet CPTED standards. The modifications are intended to deter criminal activity and aid law enforcement during emergencies. The CPTED inspection reports include a comprehensive set of changes to rental units, which include adding "structurally-sound fences in good condition and at a prescribed height;" "lighting throughout parking lot, walks, and pathways;" "posted [CFHP] certificates;" "properly trimmed landscaping;" "deadbolts and eye viewers in units;" "lift and slide protection on windows and sliding doors;" and "removal of graffiti and general cleanliness."

Third, certified properties must include a supplemental lease agreement as part of their lease. ICFA describes the lease agreement as the policy's "heart and soul" (ICFA, n.d.a.). The addenda include language indicating that a resident "shall not engage in any act intended to facilitate criminal activity" and that "a single violation shall be good cause for termination of the lease. Unless otherwise provided by law, proof of violation shall not require criminal conviction, but shall be by a preponderance of the evidence" (City of Fremont, n.d.; City of San Diego, n.d.). Possible violations of the supplemental lease agreement are typically broad and undefined, permitting nearly any interaction with the criminal legal system to serve as justification for an eviction (Archer, 2019; Prochaska, 2023). Violations also do not have to occur on the property, and the behavior of guests or others living in the unit can be the basis for an eviction (Werth, 2013).

Following those three steps, law enforcement agencies certify enrolled buildings as a "crime-free property," which allows landlords and property managers to use the supplemental lease agreement, post a CFHP sign on their property, and mention their certification in advertisements.

CFHPs build on existing efforts to control crime using evictions. As noted by ICFA, CFHPs were inspired by the "war on drugs" policies of the 1970s through 1990s, along with the "one-strike policy," which applies to federally funded public housing (ICFA, n.d.a.; Ramsey Mason, 2018). Accordingly, CFHPs are closely connected to chronic nuisance ordinances, criminal activity nuisance ordinances, and third-party policing strategies (Buerger and Mazerolle, 1998; Prochaska, 2023). These efforts attempt to prevent crime by compelling non-offending third parties to create active guardianship over a given property or face civil penalties. For example, criminal activity nuisance ordinances deem certain activities as nuisances within a municipal statute (using similarly broad language as the supplemental lease agreements used in CFHPs), then subsequently require landlords and property managers to abate nuisances that occur on their property (typically, through evicting the tenant) or face fines and potential loss of rental licenses (Prochaska, 2023).

Limited previous research has been conducted on the impact of those policies on either crime or evictions, although an analysis of criminal activity nuisance ordinances in Ohio found that criminal activity nuisance ordinances increased evictions (Kroeger and La Mattina, 2020). Previous research on criminal activity nuisance ordinances and third-party policing strategies generally have also found that these policies increase evictions among victims of domestic violence (Desmond and Valdez, 2013; Golestani, 2021; Moss and Shastry, 2019). However, CFHPs differ from criminal activity nuisance ordinances because they introduce additional policy components, such as landlord trainings, CPTED modifications, surveillance components, and use of the supplemental lease addenda. Identifying the effect of CFHPs on evictions could provide necessary evidence to local jurisdictions on the potential community costs of implementing such policies.

Data Sources and Variables

At the project's start, the authors were unable to find data sources on evictions in municipalities across California. The state limits access to unlawful detainer court filings due to AB2819, which permits access to records only in narrow circumstances. The few existing available sources on evictions in California either detail filing counts at the county-year level, such as reports by the Judicial Council of California or The Eviction Lab, or report filings for specific geographies in Los Angeles or San Francisco (Gromis et al., 2022; Lens et al., 2020; San Francisco Anti-Displacement Coalition, 2015). Given that CFHPs are implemented at the municipal level with specific rental units certified within a city, those datasets would have been unsuitable for analyzing the effect of CFHPs on evictions.

For those reasons, the authors sought records on writs of execution containing the addresses of completed writs through county sheriff's departments. These records pertain to the last step of the eviction process in California, which occurs only when a landlord has been provided a judgment of possession (which provides landlords the right to evict a tenant following a trial decision), the tenant has been provided a notice to vacate, and the tenant has not moved out after 5 days following the notice to vacate.[4] At that point, landlords can obtain a writ of execution, which permits the landlord to request a sheriff to lock the tenant out of the unit. As such, these records are a conservative estimate of total evictions in the state because they do not include any evictions that would have occurred after a tenant was given a notice to vacate and chose to leave the premises voluntarily before a court case, during court proceedings, or after judgment. These records also reflect some of the most severe outcomes of the eviction process because the data pertain to individuals forcibly removed from their units by law enforcement officers.

The authors obtained the writs of execution records by submitting Public Records Act requests to all 58 counties in California between September 2021 and October 2022, requesting records between January 2017 and January 2021. Additional requests were submitted to 30 municipalities and counties in February 2022 seeking information on CFHP implementation. Responses were received from four locations: the City of Hayward, the City of Fremont, the City of Riverside, and San Diego

[4] For additional clarification: A notice to vacate is provided to the tenant following an unlawful detainer proceeding, unlike a notice to quit, which would have been provided to the tenant before the court filing.

County.[5] More details on the request process and data processing can be found in web appendix A. The request language can be found in web appendix B. Sociodemographic indicators were extracted at the block group level from the American Community Survey 5-Year Data Release for the years 2015 through 2019 (U.S. Census Bureau, 2020). Indicators included total population, population proportions by race and ethnicity, number of renting households, and per capita income.

Research Design

This analysis used a recently developed estimator, spatial first differences (SFD), to estimate the effect of CFHPs on evictions (Druckenmiller and Hsiang, 2018).[6] Described briefly, the SFD approach involves organizing relatively small geographic areas into a series of cross-border comparisons between neighbors and then taking the pairwise difference of all included variables across each border to form a series of first-differenced relationships. Those differences are then included within a regression in which the differenced outcome is regressed on a differenced treatment variable and differenced covariates. The approach can be conceptually likened to a difference-in-differences (two-way fixed effects) estimator, which employs fixed effects for both time and group membership (such as a set of states by year), aiming to mitigate significant sources of unobserved confounding through differencing (Angrist and Krueger, 1999). In the case of SFD, a spatial sequence of neighboring geographies is substituted for the time dimension, and (arbitrary) contiguous collections of spatial neighbors are substituted for group membership.

Models were estimated separately for the municipalities of Fremont, Hayward, Riverside, and the County of San Diego. All models were estimated using data on evictions in 2019 based on writs of execution records from sheriff's departments in those locations. Treatment status was parameterized at the block group level using one of two measures: a binary indicator, indicating if a block group had one or more CFHP-certified rental units, or a continuous variable, indicating the number of CFHP-certified rental units within a block group. The coefficient on treatment status is the estimand of interest, indicating the policy's average treatment on the treated (ATT) effect. The outcome variable was counts of evictions within census block groups. Models used eviction counts rather than eviction rates due to heteroskedasticity detected in the estimated residuals based on the number of renting households.[7] To compensate for this fact, models used counts as the outcome variable and included the rate denominator—number of renting households—as an additional covariate, similar to an offset variable included in a Poisson or negative binomial regression.

[5] For Fremont, Hayward, and Riverside, local law enforcement agencies (e.g., the Fremont Police Department) administer the CFHP policy, whereas the county sheriff's department implements CFHPs in San Diego County.

[6] For interested readers, the authors also estimated treatment effects using ordinary least squares (OLS) for all model specifications and outcome measures discussed in this article. Results from those models are consistent with the estimated effects found using the spatial first difference estimator.

However, OLS models do not include the crucial estimation strategy of differencing, which removes confounding due to spatially correlated (with treatment) unobservable variables. For more technical details on this benefit of the estimation strategy, see Druckenmiller and Hsiang (2018), specifically the section on equations 17 and 18, which discusses how the estimator, by construction, removes these confounders.

[7] More specifically, block groups with a small number of renting households have a larger variance in eviction rates than block groups with many renting households.

The key identifying assumption for the SFD estimator is called the *local conditional independence* assumption. This assumption states that units are conditionally independent with respect to local spatial neighbors (which is like the assumption for time-based first differencing approaches such as difference-in-differences, in which sequential observations in a time series are assumed to be conditionally independent).[8] The assumption was tested by estimating spatial first difference models across different angles of rotation over geographic space, constructing arbitrarily different sets of spatial neighbors over a full 360 degrees of rotation. If the estimated effect of treatment is similar across those map rotations, that evidence supports the local conditional independence assumption being valid. The authors also investigated how the ATT effect changes when including additional adjustment variables within model specifications. If the ATT effect is substantially different when adjustment variables are included, that outcome would not support the validity of the local conditional independence assumption.

To check SFD results by map rotation (and to determine spatial indices generally), an algorithm was used for determining neighboring locations (Druckenmiller and Hsiang, 2018; Tanutama, 2019). This algorithm samples neighboring block group polygons in a west-to-east direction, aiming to maximize the length of consecutive neighbors. Once the algorithm is unable to find an additional neighbor, it selects a new “sampling channel” with respect to the next-longest possible series of neighbors (the groups in the analogy to a two-way fixed-effects estimator described previously). Iterative sampling channels are selected until all polygons have been ordered next to a neighbor. Web appendix C demonstrates the results of this algorithm for three sampling channels and two map angles (0 degrees and 90 degrees) using block groups in the City of Riverside.

Once first differences were obtained for a given map rotation and subsequent index of spatial neighbors, linear regression models were used to estimate ATT effects. The spatial first difference models used two specifications: an unadjusted model, which included only treatment status and the number of renting households as covariates; and a regression-adjusted model, which included the following additional covariates: population proportion White, population proportion Black, population proportion Asian, population proportion Native American, population proportion Hispanic, and per capita income. More formally, the SFD model specifications corresponded to the following equations:

Unadjusted: $\Delta_\phi E_i = \alpha + \lambda\Delta_\phi H_i + \beta\Delta_\phi P_i + \Delta_\phi \varepsilon_i$

Adjusted: $\Delta_\phi E_{i,} = \alpha + \tilde{\lambda}\Delta_\phi H_i + \tilde{\beta}\Delta_\phi P_i + \gamma\Delta_\phi X_i + \Delta_\phi u_i$

where E_i is a count of evictions for block group I, α is a model intercept, H_i is a number of renting households, P_i is either a binary indicator that equals one when a block group has one or more CFHP-certified rental units or a count of the number of CFHP-certified rental units in a block group, Δ_ϕ is the result of first-differencing neighboring block groups using a map rotation angle ϕ, and X_i is the set of included adjustment covariates. Standard errors were estimated using procedures in Conley (1999) to account for spatial autocorrelation, using the R package conleyreg 4.0.5. All regression tables are available in web appendix D.

[8] In other words, it is assumed that the differences in unobserved variables between two neighboring block groups are minimal (i.e., ignorable), more so than unobserved differences between block groups far apart from each other.

Ultimately, estimated treatment effects are comparable across map rotations. Web appendix E provides the distribution of treatment effect estimates across map angles by location and model (i.e., estimates of β in the previous equations). The displayed densities were derived by sampling 10,000 draws from normal distribution with a mean equal to the mean estimated treatment effect and standard deviation equal to the standard deviation of the estimated treatment effect. Sampled draws of beta coefficients were used to obtain a unified effect across model angles, with draws collapsed using Rubin's rules (Rubin, 2004); specifically, draws were ordered across all model angles from smallest to largest. Then, draws were summarized using the mean of those draws and the 2.5th and 97.5th percentiles, which correspond to the results displayed in the figure for the row "Overall Effect" in the exhibits in web appendix E. These results display comparable direction and magnitude in the estimated treatment effects across model angles, providing evidence that the assumption of local conditional independence is being met (web appendix E).

To simplify the presentation of results, percentage change in evictions were calculated across treated block groups based on the unified treatment effect estimate across model angles (web appendix F). For each site (Fremont, Hayward, Riverside, and San Diego County), a counterfactual change in evictions was calculated by subtracting the estimated treatment effect (and estimated treatment effect confidence intervals) from the observed mean of evictions across treated block groups in each location. These values were used to calculate the percentage change based on the observed mean in each location.[9] A summary treatment effect across sites was calculated by summarizing the treatment effect draws across both sites and model angles using Rubin's rules. The same procedure was used to calculate a counterfactual change using the average eviction count across all sites.

Results

Exhibits 1.1 through 1.4 report descriptive statistics for each variable by study site. The first column displays the mean and standard deviation across block groups that do not contain a CFHP-certified rental unit (control groups). The second column displays the same information for block groups with a CFHP-certified rental unit (treatment groups). The last column displays the estimated mean difference in each variable between treated and control groups, along with the confidence intervals for the mean difference.

[9] For example, percentage change = ATT / Mean evictions.

Exhibit 1.1

Descriptive Statistics in Fremont for Block Groups With and Without CFHP-Certified Rental Units

	Blocks Without CFHP-Certified Units	Blocks With CFHP-Certified Units	Mean Difference
Eviction Count	0.34 (0.76)	2.00 (2.26)	1.67 (0.86, 2.46)
Black (Pop %)	3 (4)	4 (3)	1 (– 1, 2)
Asian (Pop %)	59 (18)	56 (17)	– 3 (– 10, 4)
White (Pop %)	26 (13)	22 (11)	– 4 (– 8, 10)
American Indian/ Alaskan Native (Pop %)	0 (1)	1 (1)	0 (0, 1)
Latin/Hispanic (Pop %)	11 (10)	17 (12)	6 (1, 11)
Per Capita Income	$54,932 ($12,453)	$48,385 ($12,013)	– $6,547 (– $11,472,– $1,622)
Rental Units	169.8 (170.9)	481 (288.1)	311.3 (205.1, 417.4)
CFHP-Certified Properties		1.6 (1.0)	
N	86	32	

CFHP = crime-free housing policy. Pop = population.

Sources: Writs of execution data obtained from sheriff's departments; American Community Survey 5-year data, 2019

Exhibit 1.2

Descriptive Statistics in Hayward for Block Groups With and Without CFHP-Certified Rental Units

	Blocks Without CFHP-Certified Units	Blocks With CFHP-Certified Units	Mean Difference
Eviction Count	1.19 (1.93)	4.58 (4.06)	3.39 (1.06, 5.72)
Black (Pop %)	9 (7)	14 (9)	5 (0, 10)
Asian (Pop %)	26 (14)	20 (5)	– 6 (– 10,– 2)
White (Pop %)	37 (13)	38 (50)	1 (– 3, 5)
American Indian/ Alaskan Native (Pop %)	1 (2)	1 (1)	0 (– 1, 1)
Latin/Hispanic (Pop %)	39 (19)	43 (13)	4 (– 4, 13)
Per Capita Income	$35,544 ($14,279)	$32,630 ($8,444)	– $2,914 (– $8,547, $2,718)
Rental Units	226.8 (174.1)	519.6 (341.7)	292.8 (96, 489.5)
CFHP-Certified Properties		1.3 (1.2)	
N	88	12	

CFHP = crime-free housing policy. Pop = population.

Sources: Writs of execution data obtained from sheriff's departments; American Community Survey 5-year data, 2019

Exhibit 1.3

Descriptive Statistics in Riverside for Block Groups With and Without CFHP-Certified Rental Units

	Blocks Without CFHP-Certified Units	Blocks With CFHP-Certified Units	Mean Difference
Eviction Count	1.65 (1.87)	3.88 (3.71)	2.23 (1.33, 3.14)
Black (Pop %)	6 (6)	7 (5)	2 (0, 3)
Asian (Pop %)	7 (8)	6 (7)	– 1 (– 4, 1)
White (Pop %)	61 (16)	56 (13)	– 5 (– 9,– 1)
American Indian/ Alaskan Native (Pop %)	1 (1)	1 (2)	0 (0, 1)
Latin/Hispanic (Pop %)	51 (23)	58 (19)	7 (1, 13)
Per Capita Income	$29,908 ($13,275)	$22,664 ($10,146)	– $7,244 (– $10,511,– $3,977)
Rental Units	153.4 (154.4)	364.6 (194.6)	211.2 (159.4, 263)
CFHP-Certified Properties		1.9 (1.6)	
N	127	74	

CFHP = crime-free housing policy. Pop = population.

Sources: Writs of execution data obtained from sheriff's departments; American Community Survey 5-year data, 2019

Exhibit 1.4

Summary Statistics in San Diego County for Block Groups With and Without CFHP-Certified Rental Units

	Blocks Without CFHP-Certified Units	Blocks With CFHP-Certified Units	Mean Difference
Eviction Count	1.79 (3.27)	2.87 (2.97)	1.08 (0.38, 1.77)
Black (Pop %)	5 (7)	3 (5)	– 2 (– 3,– 1)
Asian (Pop %)	11 (12)	6 (7)	– 5 (– 6,– 3)
White (Pop %)	72 (17)	77 (12)	5 (2, 8)
American Indian/ Alaskan Native (Pop %)	1 (3)	1 (2)	0 (0, 1)
Latin/Hispanic (Pop %)	32 (25)	40 (26)	8 (2, 14)
Per Capita Income	$40,000 ($20,940)	$30,902 ($12,319)	– $9,098 (– $12,059,– $6,137)
Rental Units	293.3 (312.9)	380.8 (256)	87.4 (27.6, 147.2)
CFHP-Certified Properties		1.8 (1.2)	
N	1693	75	

CFHP = crime-free housing policy. Pop = population.

Sources: Writs of execution data obtained from sheriff's departments; American Community Survey 5-year data, 2019

The descriptive results show that Fremont, Hayward, and San Diego County have higher eviction rates in block groups with CFHP-certified rental units (treated units) compared with block groups without CFHP-certified rental units (control units). However, the mean difference of eviction rates between treated and control groups is nonsignificant across all sites. Across all locations, treated units have significantly more rental properties than control units. In Fremont, Riverside, and San Diego County, treated units have significantly lower per capita income than control units

(-$6,547, -$7,244, and -$9,098, respectively). Per capita income is also lower for treated units in Hayward (-$2,914), although the mean difference is nonsignificant. Concerning race and ethnicity, treated units have a significantly larger Latin/Hispanic population proportion than controls in Fremont, Riverside, and San Diego County, whereas treated units in Hayward and Riverside have a significantly larger Black population proportion. The American Indian/Alaskan Native population proportion for treated units is also modest but statistically significantly larger in Fremont, Riverside, and San Diego County.

Exhibits 2.1 through 2.4 display maps of each site. Shading indicates the number of evictions in each block group, and triangles display the relative location of each CFHP-certified rental unit. To maintain privacy, the number of evictions was categorized into broader bins, and the location of CFHP-certified rental units was randomized in each block group.

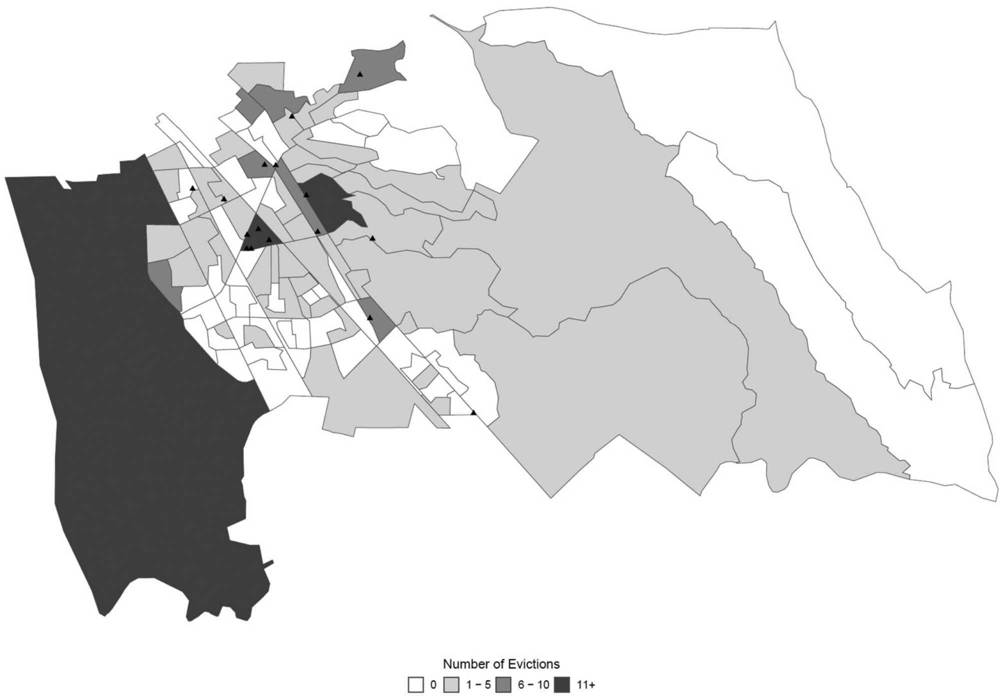
Exhibit 2.1

Number of CFHP-Certified Rental Units and Executed Evictions Within Fremont, California Census Block Groups

Notes: Triangles indicate the relative location of CFHP-certified rental units. The specific location of each rental unit was randomized within block groups.
Source: Public Records Act Requests

Exhibit 2.2

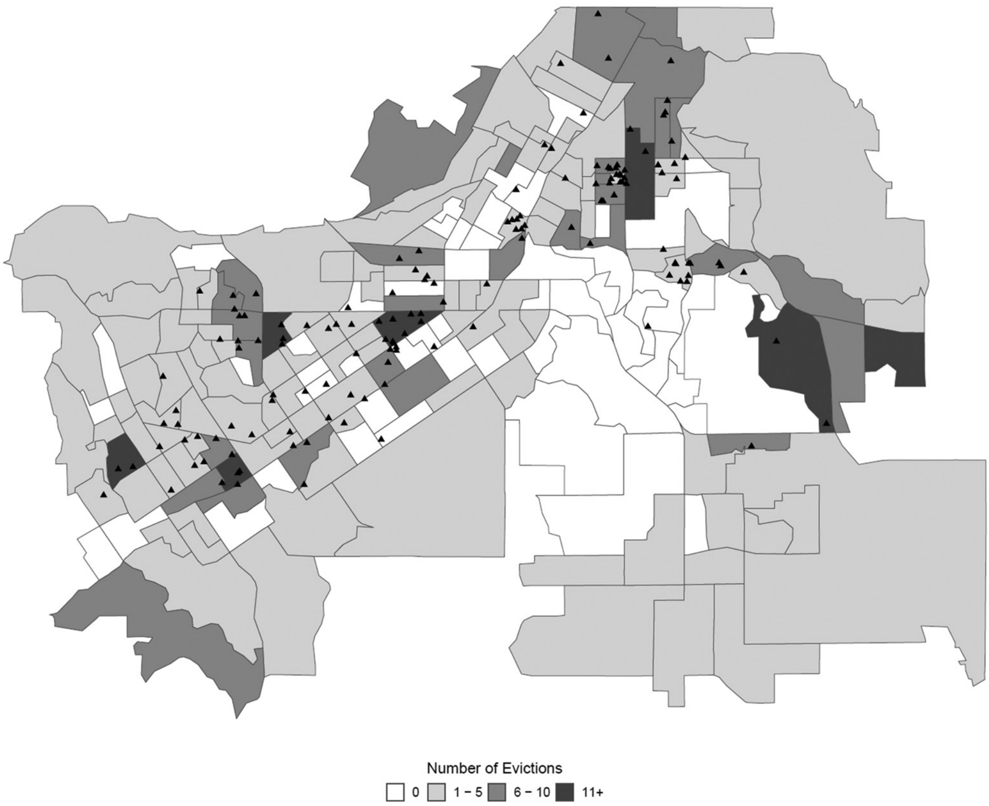
Number of CFHP-Certified Rental Units and Executed Evictions Within Map of Hayward, California Census Block Groups



Notes: Triangles indicate the relative location of CFHP-certified rental units. The specific location of each rental unit was randomized within block groups.
Source: Public Records Act Requests

Exhibit 2.3

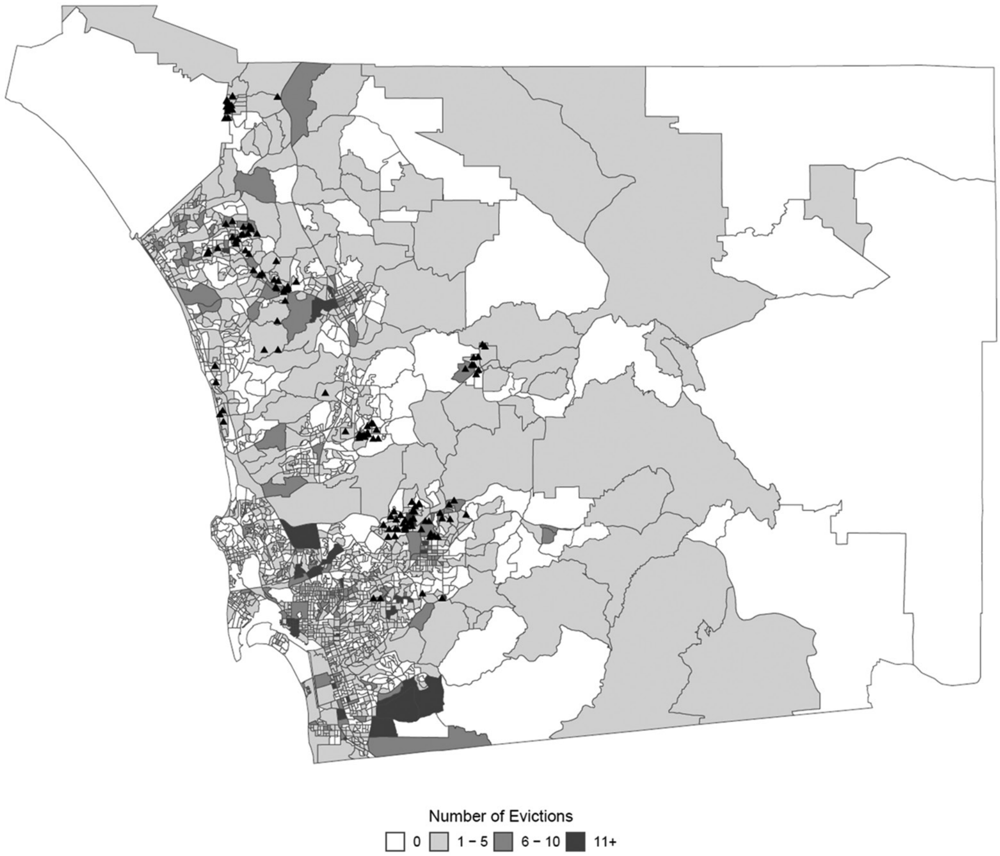
Number of CFHP-Certified Rental Units and Executed Evictions Within Riverside, California Census Block Groups



Notes: Triangles indicate the relative location of CFHP-certified rental units. The specific location of each rental unit was randomized within block groups.
Source: Public Records Act Requests

Exhibit 2.4

Number of CFHP-Certified Rental Units and Executed Evictions Within San Diego County, California Block Groups



Notes: Triangles indicate the relative location of CFHP-certified rental units. The specific location of each rental unit was randomized within block groups.
Source: Public Records Act Requests

The percentage of block groups treated according to the policy varies considerably by location, with 4.2 percent of block groups in San Diego County, 12 percent in Hayward, 27 percent in Fremont, and 37 percent in Riverside. In Fremont and Hayward, treated block groups are clustered within the core of each city. Treated block groups in Riverside are clustered on a west-to-northeast diagonal and within specific cities in San Diego County. The average treated block group contains 1.56 certified properties in Fremont, 1.33 in Hayward, 1.93 in Riverside, and 1.79 in San Diego County.

Empirical Findings

Across all locations, estimated treatment effects are significant, in the same direction, and with comparable magnitude.[10] Regardless of the specific location, model, or treatment measure, treated

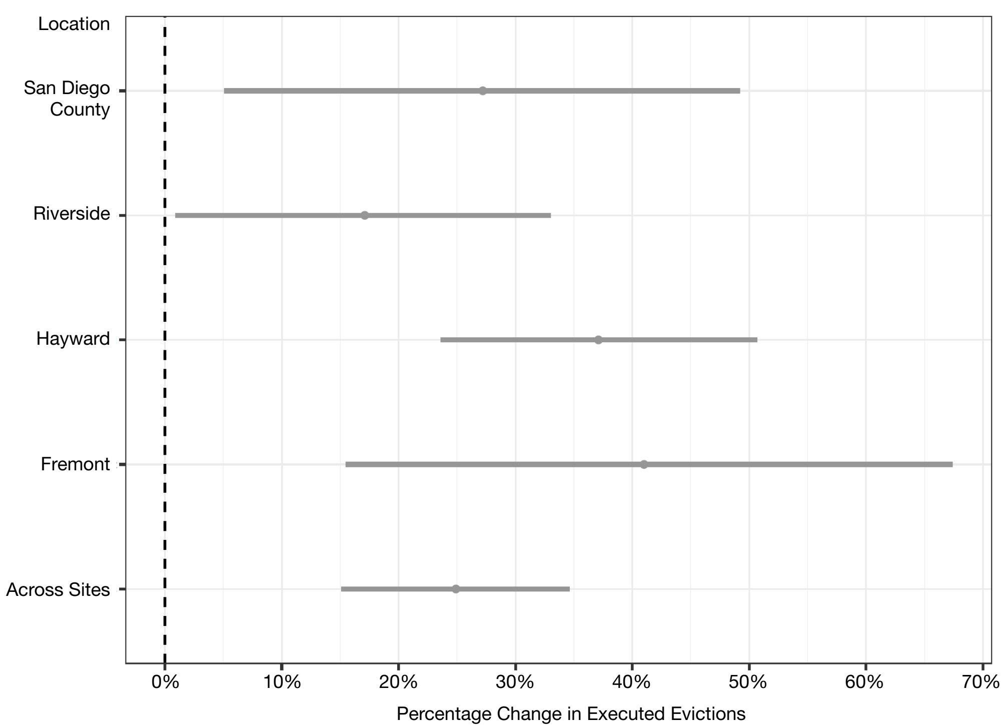
[10] Model regression tables are available in web appendix D, and estimated average treatment-on-the-treated effects by model, location, and site are available in web appendixes E and F.

groups have increased evictions compared with control groups. There is not a large difference in results between the two treatment measures after accounting for the average number of CFHP-certified rental units in treated block groups. In addition, the estimated treatment effects in the adjusted models are similar to the results for unadjusted models, providing additional evidence that the local conditional independence assumption is being met.

Exhibit 3 displays the estimated counterfactual percent change in evictions that occurs in block groups containing one or more CFHP-certified rental properties, using the unadjusted model specification. A significant effect on evictions is found across all locations, with CFHPs increasing evictions within treated block groups by 17.1 percent (0.9 percent, 33 percent) in Riverside; 27.2 percent (5.1 percent, 49.2 percent) in San Diego County; 37.1 percent (23.6 percent, 50.7 percent) in Hayward; and 41 percent (15.5 percent, 67.4 percent) in Fremont. Aggregating the effect across sites, treated block groups experience a 24.9 percent (15.1 percent, 34.6 percent) increase in evictions.

Exhibit 3

Estimated Percentage Change in Evictions in Block Groups Containing CFHP-Certified Rental Units, by Location



Source: Authors calculations using results from Spatial First Difference Models

Conclusion

This analysis demonstrates that crime-free housing policies (CFHPs) increase the number of evictions that occur in neighborhood blocks by an average of 24.9 percent. Evictions are a harmful outcome for individuals and carry a large social cost to governments and communities. Policymakers considering instating policies that rely on evictions to generate an outcome, including third-party policing efforts such as CFHPs, should carefully weigh the costs of additional evictions against the policy's purported benefit. The effect on evictions identified in this study may be similar for other policies—including criminal activity nuisance ordinances, chronic nuisance ordinances, and the one-strike policy—that use evictions based on nuisance actions and contact with the criminal justice system as a strategy to prevent crime. For example, previous research investigating the effect of criminal nuisance ordinances in Ohio found that such policies increased eviction filings by 16 percent (Kroeger and La Mattina, 2020). The close similarity in enforcement strategies across these policies, which differ mainly in the source of statutory language enabling the enforcement (i.e., as either a municipal ordinance or as a supplemental lease agreement), suggests that these policy efforts may have similar effects on evictions.

U.S. Department of Housing and Urban Development memos have raised additional concerns about the implementation of CFHPs (Kanovsky, 2016; McCain, 2022). Those memos focused on how CFHPs could cause a disproportionate number of evictions for victims of domestic violence and cautioned municipalities that CFHPs may cause discrimination in housing access because they prevent formerly incarcerated individuals from tenancy. However, fewer concerns have been raised about the effect of CFHPs on evictions generally. Given the population harm of evictions, municipalities must weigh the costs and benefits of maintaining CFHPs because, even if the policy achieves the stated aim of crime reduction, it may introduce additional community harms. However, recent evidence has found that CFHPs likely do not decrease crime rates, and analysis on nuisance ordinances has shown that they may, in fact, increase crime (Falcone, 2023; Griswold et al., 2023). Evictions may also lead to additional crime, undermining the purpose of CFHPs (Semenza et al., 2022). Further, as the descriptive statistics demonstrate, lower-income populations are disproportionately affected by CFHPs. Eviction events can create disproportionate harm for low-income individuals because evictions can lead to increased financial losses, additional housing instability due to the eviction record, challenges in obtaining subsidized housing, and increased risk of homelessness events (Collinson et al., 2023; Desmond, Gershenson, and Kiviat, 2015; Desmond and Shollenberger, 2015).

One justification offered for the use of third-party policing strategies, such as CFHPs, is that they are cost-effective, reducing the need to use law enforcement resources to prevent crime by promoting landlords to engage in active guardianship over their rental units (Buerger and Mazerolle, 1998; Mazerolle and Roehl, 1998). However, this analysis shows that CFHPs increase the number of completed writs of execution, which would subsequently increase the cost of administering the policy. Each completed writ requires law enforcement resources to execute it, suggesting that CFHPs could increase overall net resource use with little benefit on crime. CFHPs may also lead to other costs to the public as well. Previous evidence suggests that each eviction has a lower bound cost of $8,000 per person evicted, not including court or law enforcement costs associated with executing an eviction (Collinson and Reed, 2018). As such, CFHPs may carry a large cost burden for municipalities that implement the policy, given the magnitude of the identified effect on evictions.

The results also indicate that blocks with CFHP-certified rental units contain more renters, have a lower per capita income, and, depending on the exact location, a larger proportion of Black and Latin/Hispanic populations than blocks without CFHP-certified rental units. By increasing the number of evictions in those blocks, CFHPs may further marginalize low-income populations and people of color and may increase housing instability, homelessness, and the use of social services among those populations.

CFHPs' targets for enforcement are renters in multifamily housing units and individuals with a history of criminal justice involvement. Those populations tend to be more non-White and have lower income than the general population, indicating that CFHPs may have a further disproportionate effect on low-income Black and Latin/Hispanic populations, in addition to the demographic difference of the affected neighborhoods displayed in exhibits 1.1 through 1.4 (DeSilver, 2021; Zeng, 2022). Eviction events also disproportionately affect Black women and children, which further increases the potential risk of discrimination occurring from the use of CFHPs (Graetz et al., 2023; Hepburn, Louis, and Desmond, 2020). Previous research has also noted that the populations enforcing CFHPs—law enforcement officers and landlords—may make racially discriminatory choices when provided additional discretion in their decisionmaking (Archer, 2019; Christensen, Sarmiento-Barbieri, and Timmins, 2021; Goff et al., 2016; Hanson and Hawley, 2011; Lofstrom et al., 2022).

In addition to the populations affected by CFHPs, policymakers and government prosecutors have noted that CFHPs may have a discriminatory impact due to the policy's enforcement. For example, the Department of Justice pursued a lawsuit against the City of Hesperia, California, alleging that the city adopted a CFHP to discriminate against Black and Latin/Hispanic individuals in the municipality (U.S. Department of Justice, 2022). In addition, California recently passed a new law, AB1418, to curtail the use of CFHPs in California municipalities.[11] As part of the bill's committee summary, legislators noted that the introduction of the bill was motivated by the potential of CFHPs to produce racially segregative effects and discriminatory impacts.[12]

Legal researchers have noted additional harms that may result from continued use of CFHPs and related policies—beyond the harms caused by additional evictions and potential discriminatory effects. For instance, legal researchers have argued that the application of CFHPs could lead to violations of the Fair Housing Act; First Amendment rights, such as freedom of association; and Fourteenth Amendment rights, such as procedural due process and equal protection (Jarwala and Singh, 2019; Katach, 2015; Prochaska, 2023; Ramsey Mason, 2018; Smith, 2018; Werth, 2013). Policymakers should consider those additional possible harms—in addition to the results concerning CFHPs' effect on evictions and the demographics of affected populations—when considering continued use or adoption of crime-free housing policies.

[11] CA AB1418, 2023–2024 Regular Session, Amended April 12, 2023. https://legiscan.com/CA/text/AB1418/id/2778119.

[12] Assembly Committee on Judiciary. "Tenancy: Local Regulations: Contact With Law Enforcement or Criminal Convictions." Date of Hearing: April 11, 2023. https://trackbill.com/s3/bills/CA/2023/AB/1418/analyses/assembly-judiciary.pdf.

Limitations

This study carries limitations. First, the data collected on evictions is based on writs of execution records, which are issued only to tenants who have lost an unlawful detainer case and have not vacated their unit. Accordingly, the estimated effect using these data may not hold for informal, illegal, or eviction filings if the pattern in those measures differs considerably between treated and control units compared with the writs data. For example, if landlords of CFHP-certified properties use the serial threat of evictions to remove tenants from their housing more often than landlords without CFHP certification and more often than using writs, then the estimates could understate the true effect on evictions. Writs of execution records are also an underestimate of the total number of evictions that occur in neighborhood blocks because the records apply only to tenants forcibly removed from their rental unit. Accordingly, the estimated effect of CFHPs on completed evictions (as measured by writs of execution) is likely an underestimate of the effect of CFHPs on evictions generally.

Second, while processing the writs of execution records, multiple observations had to be removed from the dataset due to incomplete address and date information, retractions, and implausible geocodes. If the records removed from the analysis are not randomly distributed across treated units and control units, this fact may bias the estimated effect.

Third, if the local conditional independence assumption for spatial first differences does not hold, then the estimated average treatment effects might be biased. This assumption was tested using available evidence from map rotations and adjusted models, with results suggesting the assumption may be valid. However, if a confounding variable exists that is not removed from the analysis through the spatial differencing strategy, then the assumption may not hold, and the estimates may not reflect an unbiased causal relationship.

Summary

Crime-free housing policies (CFHPs) significantly increase the number of evictions that occur in neighborhood blocks containing CFHP-certified rental units. Affected neighborhood blocks tend to have lower per capita income, a larger proportion of Black and Latin/Hispanic populations, and more rental units than the broader municipality. Given the substantial harm of evictions and the cost of evictions for local governments, municipalities should weigh the cost and benefits of maintaining or adopting policies that use eviction to achieve a policy outcome.

In addition, CFHPs are closely related to criminal activity nuisance ordinances, chronic nuisance ordinances, and the one-strike policy, which are widely prevalent across the United States and employ a similar enforcement strategy: using evictions as a crime prevention tool. Those policies also likely increase the number of evictions that occur in the United States. Emerging evidence also suggests that those policies likely do not lead to a reduction in crime, indicating that no concomitant benefit stems from increased evictions. On the basis of those findings, federal and state legislators should closely evaluate the efficacy of using evictions to prevent crime and determine if new legislation is needed to curtail the harmful effects of those policies on communities.

Appendix A. Additional Method Details

Data Processing

Data were received from 55 sheriff's departments in California as either physical documents, Excel files, or PDFs. In total, 14,082 pages of records were received and six Excel files. Appendix G provides a synthetic example of a record page. Physical documents were scanned, converted into PDFs using optical character recognition, and post-processed to make records uniform before data extraction. To do so, hand-scanned documents were rotated so that text was aligned horizontally, identified retractions in documents and replaced them with white blocks, and replaced all other colors with black. Post-processing was performed in Python 3.8.

For files received as PDFs or converted into PDFs, the Azure-AI-Form Recognizer 3.2.0 was used to train eight custom template extraction models to generate tabular data from the PDF files. To train the models, 40 pages of records were hand coded, corresponding to the eight main templating formats received across jurisdictions. Those codes indicated the position of rows and columns within each templating format. The accuracy of the extracted tables was validated by calculating the Levenshtein distance ratio between text in PDFs and extracted tables, finding that the distance ratio exceeded 0.98 across all template formats, indicating a high degree of alignment between extracted text and underlying documents. The final generated dataset consisted of rows for completed writs of execution, along with the event's data and address.

Before post-processing, the dataset consisted of 244,298 records. The following rows were removed from the analysis dataset: rows that did not contain date or location information; contained NA values due to a retraction (departments confirmed that retractions pertained to canceled writs); contained a malformed date due to how the document had been scanned by the sheriff's office; was a duplicate record; or corresponded to a scheduled or canceled writ (rather than a completed writ). Addresses were geocoded to GPS coordinates using the Tidygeocoder package in R 4.2.2. (Cambon et al., 2021). Coordinates were validated by comparing imputed ZIP Codes from geocoding to existing ZIP Codes in the original address text. Across all locations, ZIP Codes were successfully matched for 98 percent of imputed addresses. Rows that did not have imputed ZIP Codes that matched address text were inspected, where these rows contained either informal address text (e.g., "Apartment behind the McDonald's on 96th St.") or address text with incomplete information due to hand scanning; these rows were subsequently removed from the analysis dataset. Finally, geocoded addresses were merged with 2019 Census TIGER files and aggregated eviction counts to the block group level. The final analysis dataset contained 216,412 records.

Descriptive Statistics

The mean and standard deviation was calculated for each included study variable within each location in the spatial first difference models (Fremont, Hayward, Riverside, and San Diego County). Means and standard deviations were stratified by treatment status between block groups containing CFHP-certified rental units and block groups without CFHP-certified rental units (exhibits 1.1–1.3). The mean difference between "Blocks With CFHP-Certified Units" and "Blocks Without CFHP-Certified Units," was calculated, along with the confidence interval for the mean difference, based on an unpaired t-test with unequal variance (Welch's t-test). The 95-percent confidence interval of the t-test for the mean difference is provided to categorize the uncertainty. A confidence interval of the mean difference that crosses zero indicates that the mean difference is nonsignificant at the 5-percent threshold.

Appendix B. Public Records Act Requests

Exhibit B.1

Public Records Act Request: Writs of Execution

Hello,

This is a request under the California Public Records Act (pursuant to California Government Code Section 6253(c)) for records in the possession of the [sheriff department name] pertaining to notices of restoration (e.g. completed evictions) given between Jan 1st, 2017 through Jan 1st, 2021. This information should include:

Records or lists of evictions showing served Notice of Restoration, including the following pieces of information:

- The date the notice of restoration was served, including month & year.
- The city in which the notice of restoration was served.

Thanks to the department for the work on responding to this request.

Kind regards,

Exhibit B.2

Public Records Act Request: CFHP Information

Hello,

This is a request under the California Public Records Act (pursuant to California Government Code Section 6253(c)). We are seeking records on [city's name] "Crime-Free Housing Program". Specifically, we are looking for the following pieces of information:

- In what month/year was the program adopted by the city?
- If the program was in place during 2019, what properties were certified under the program (specifically, we are seeking a list with the addresses for these properties)?
- If the program had been implemented in 2019, could the city provide the training documentation and lease addendum used by the program?
- If the program had been implemented in 2019, could the city provide any electronic databases or text databases related to the enforcement of this program?

Thanks to the city for the work on responding to this request.

Kind regards,

Exhibit B.3

CPRA Request: CFHP Information with Additional Details

Hello,

This is a request under the California Public Records Act (pursuant to California Government Code Section 6253(c)). We are seeking records on [city's name] "Crime-Free Housing Program". Specifically, we are looking for the following pieces of information:

- In what month/year was the program adopted by the city?

We additionally are requesting materials related to the operation of the county's crime-free housing program, specifically:

- Documents concerning properties eligible for crime-free housing program enforcement or violations, including the address of properties participating in the program.
- All documents concerning the crime-free housing program, including enforcement of it against residential properties, landlords, or tenants, including copies of all violations, letters, notices, files, and any other external or internal communication, including emails, related thereto since the program's formation to the present.
- Documents that describe policies or procedures for the writing of police and/or incident reports by the sheriff's department when there is a violation of the crime-free housing program.
- Electronic copy of any database or databases containing information regarding the crime-free housing program's enforcement.
- Police and/or incident reports corresponding to violations of the crime-free housing program.
- All training or informational materials regarding the crime-free housing program provided to landlords, tenants, police, or others, including any electronic, video, or audio recordings of trainings.
- All documents concerning mandatory or suggested lease terms or crime-free lease addendum promoted, adopted, or created by the department.
- Any internal or external communications with elected officials or city employees regarding the crime-free housing program since the program's adoption.

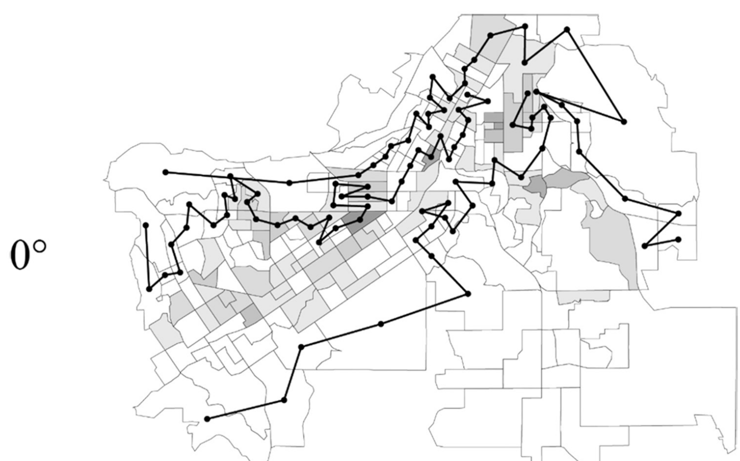
Thanks to the city for the work on responding to this request.

Kind regards,

Appendix C. Example of Sampling Algorithm

Exhibit C.1

Example of Sampling Algorithm Used to Determine Neighboring Block Groups in the City of Riverside When the Map is Rotated Zero and 90 Degrees (1 of 2)

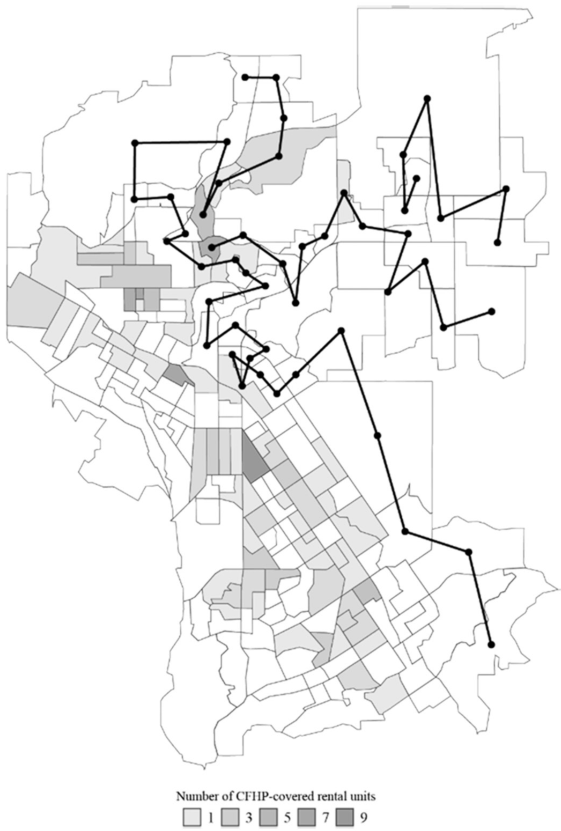


Source: Authors

Exhibit C.1

Example of Sampling Algorithm Used to Determine Neighboring Block Groups in the City of Riverside When the Map is Rotated Zero and 90 Degrees (2 of 2)

90°



Source: Authors

Appendix D. Regression Results for Spatial First Differences Models

Exhibit D.1

Regression Results by Map Rotation for Fremont, Unadjusted Spatial First Differences Model Using Binary Treatment

	Angle=0	Angle=30	Angle=60	Angle=90	Angle=120	Angle=150	Angle=180	Angle=210	Angle=240	Angle=270	Angle=300
One or More CFHP Units	1.09* (0.07)	1.11* (0.07)	0.86* (0.06)	0.31* (0.03)	0.78* (0.06)	1.14* (0.07)	0.55* (0.05)	0.81* (0.06)	0.86* (0.07)	0.95* (0.06)	1.16* (0.07)
Rental Units (in hundreds)	0.18* (0.01)	0.17* (0.01)	0.11* (0.01)	0.19* (0.01)	0.19* (0.01)	0.11* (0.01)	0.18* (0.01)	0.24* (0.01)	0.14* (0.01)	0.14* (0.01)	0.15* (0.01)
Treated Units	32	32	32	32	32	32	32	32	32	32	32
Total Units	103	102	106	103	106	105	106	105	105	104	102
Adjusted R-Squared	0.26	0.26	0.14	0.16	0.19	0.18	0.16	0.23	0.16	0.19	0.26

** p < 0.05.*

CFHP = crime-free housing policy.

Exhibit D.2

Regression Results by Map Rotation for Hayward, Unadjusted Spatial First Differences Model Using Binary Treatment

	Angle=0	Angle=30	Angle=60	Angle=90	Angle=120	Angle=150	Angle=180	Angle=210	Angle=240	Angle=270	Angle=300
One or More CFHP Units	1.64* (0.09)	1.65* (0.09)	1.65* (0.11)	2.01* (0.09)	1.88* (0.10)	1.89* (0.10)	1.87* (0.10)	1.82* (0.10)	1.19* (0.09)	1.26* (0.11)	2.01* (0.09)
Rental Units (in hundreds)	0.52* (0.02)	0.51* (0.02)	0.55* (0.01)	0.51* (0.02)	0.53* (0.02)	0.53* (0.02)	0.54* (0.02)	0.62* (0.01)	0.55* (0.02)	0.56* (0.02)	0.58* (0.01)
Treated Units	12	12	12	12	12	12	12	12	12	12	12
Total Units	88	88	89	87	92	92	92	90	92	92	87
Adjusted R-Squared	0.35	0.34	0.36	0.36	0.3	0.3	0.3	0.36	0.29	0.3	0.41

** p < 0.05.*

CFHP = crime-free housing policy.

Source: Authors

Exhibit D.3

Regression Results by Map Rotation for Riverside, Unadjusted Spatial First Differences Model Using Binary Treatment

	Angle=0	Angle=30	Angle=60	Angle=90	Angle=120	Angle=150	Angle=180	Angle=210	Angle=240	Angle=270	Angle=300
One or More CFHP Units	0.29* (0.09)	0.30* (0.10)	0.18 (0.10)	0.52* (0.11)	0.28* (0.10)	0.72* (0.09)	0.69* (0.07)	0.49* (0.09)	1.13* (0.13)	0.97* (0.14)	0.47* (0.11)
Rental Units (in hundreds)	0.84* (0.02)	0.82* (0.02)	0.85* (0.03)	0.80* (0.03)	0.78* (0.02)	0.69* (0.02)	0.64* (0.02)	0.64* (0.02)	0.64* (0.02)	0.70* (0.03)	0.74* (0.03)
Treated Units	74	74	74	74	74	74	74	74	74	74	74
Total Units	178	177	175	179	185	185	183	182	181	181	182
Adjusted R-Squared	0.38	0.37	0.38	0.37	0.34	0.31	0.27	0.27	0.3	0.34	0.33

* $p < 0.05$.

CFHP = crime-free housing policy.

Exhibit D.4

Regression Results by Map Rotation for Fremont, Unadjusted Spatial First Differences Model Using Binary Treatment

	Angle=0	Angle=30	Angle=60	Angle=90	Angle=120	Angle=150	Angle=180	Angle=210	Angle=240	Angle=270	Angle=300
One or More CFHP Units	0.51* (0.17)	0.67* (0.19)	1.01* (0.25)	0.72* (0.21)	1.45* (0.38)	0.54* (0.17)	0.68* (0.25)	0.69* (0.23)	0.93* (0.28)	0.80* (0.20)	1.00* (0.13)
Rental Units (in hundreds)	0.43* (0.07)	0.50* (0.09)	0.40* (0.10)	0.40* (0.10)	0.39* (0.10)	0.41* (0.11)	0.50* (0.09)	0.40* (0.11)	0.37* (0.10)	0.36* (0.10)	0.32* (0.09)
Treated Units	75	75	75	75	75	75	75	75	75	75	75
Total Units	1,590	1,582	1,580	1,579	1,592	1,616	1,611	1,612	1,614	1,593	1,589
Adjusted R-Squared	0.18	0.22	0.17	0.17	0.17	0.17	0.22	0.18	0.17	0.14	0.14

* $p < 0.05$.

CFHP = crime-free housing policy.

Source: Authors

Exhibit D.5

Regression Results by Map Rotation for Fremont, Adjusted Spatial First Differences Model Using Binary Treatment

	Angle=0	Angle=30	Angle=60	Angle=90	Angle=120	Angle=150	Angle=180	Angle=210	Angle=240	Angle=270	Angle=300
One or More CFHP Units	1.00* (0.07)	1.08* (0.06)	0.79* (0.06)	0.20* (0.03)	0.76* (0.06)	0.98* (0.06)	0.47* (0.04)	0.79* (0.06)	0.83* (0.06)	0.85* (0.06)	1.18* (0.07)
Rental Units (in hundreds)	0.18* (0.01)	0.16* (0.01)	0.12* (0.01)	0.21* (0.01)	0.18* (0.01)	0.08* (0.01)	0.20* (0.01)	0.23* (0.01)	0.14* (0.01)	0.12* (0.01)	0.14* (0.01)
Per Capita Income (in $10,000)	– 0.10 (0.02)	– 0.08* (0.02)	0.01 (0.02)	0.00 (0.02)	– 0.10* (0.02)	– 0.23* (0.04)	– 0.05 (0.03)	– 0.08* (0.03)	– 0.10 (0.02)	– 0.10* (0.02)	– 0.08* (0.02)
Asian (Pop %)	0.27 (0.31)	0.23 (0.28)	– 2.30* (0.37)	– 0.15 (0.17)	– 0.95* (0.45)	0.58 (0.32)	– 1.02* (0.32)	– 2.05* (0.43)	– 0.17 (0.31)	– 0.49 (0.28)	0.23 (0.27)
White (Pop %)	5.09* (0.68)	4.84* (0.66)	0.41 (0.71)	3.68* (0.45)	1.51* (0.54)	4.33* (0.55)	– 0.02 (0.35)	1.63* (0.81)	3.82* (0.61)	5.49* (0.53)	4.93* (0.58)
Black (Pop %)	1.07* (0.27)	0.95* (0.24)	– 1.48* (0.37)	1.98* (0.19)	– 0.36 (0.37)	0.73* (0.25)	– 0.06 (0.24)	– 1.33* (0.33)	0.59* (0.24)	– 0.34 (0.25)	1.24* (0.25)
American Indian/ Alaska Native (Pop %)	24.38* (1.97)	27.49* (1.67)	13.47* (1.76)	27.34* (2.33)	21.56* (1.67)	27.21* (1.68)	25.04* (1.84)	28.10* (3.00)	27.07* (2.44)	27.44* (2.11)	21.49* (1.69)
Latin/Hispanic (Pop %)	– 1.58* (0.39)	– 1.78* (0.38)	– 2.63* (0.43)	– 2.54* (0.19)	– 3.98* (0.41)	– 0.68* (0.32)	– 1.99* (0.29)	– 5.65* (0.59)	– 1.89* (0.37)	– 2.29* (0.32)	– 3.02* (0.32)
Treated Units	32	32	32	32	32	32	32	32	32	32	32
Total Units	103	102	106	103	106	105	106	105	105	104	102
Adjusted R-Squared	0.26	0.26	0.1	0.19	0.2	0.18	0.15	0.26	0.16	0.2	0.27

** $p < 0.05$.*

CFHP = crime-free housing policy. Pop = population.

Source: Authors

Exhibit D.6

Regression Results by Map Rotation for Hayward, Adjusted Spatial First Differences Model Using Binary Treatment

	Angle=0	Angle=30	Angle=60	Angle=90	Angle=120	Angle=150	Angle=180	Angle=210	Angle=240	Angle=270	Angle=300
One or More CFHP Units	1.53* (0.09)	1.55* (0.09)	1.70* (0.12)	1.99* (0.11)	2.04* (0.11)	2.05* (0.11)	2.02* (0.12)	1.65* (0.12)	1.22* (0.09)	1.20* (0.11)	1.97* (0.10)
Rental Units (in hundreds)	0.63* (0.02)	0.62* (0.02)	0.62* (0.02)	0.54* (0.03)	0.57* (0.02)	0.57* (0.02)	0.60* (0.02)	0.71* (0.02)	0.63* (0.02)	0.64* (0.02)	0.68* (0.02)
Per Capita Income (in $10,000)	0.21* (0.03)	0.28* (0.03)	0.18* (0.03)	0.17* (0.03)	0.44* (0.07)	0.45* (0.07)	0.50* (0.07)	0.78* (0.07)	0.39* (0.06)	0.35* (0.06)	0.13* (0.04)
Asian (Pop %)	– 4.97* (0.40)	– 4.58* (0.45)	– 3.80* (0.46)	– 3.40* (0.52)	– 4.94* (0.62)	– 5.05* (0.63)	– 4.38* (0.65)	– 4.08* (0.59)	– 2.56* (0.64)	– 4.19* (0.51)	– 2.26* (0.50)
White (Pop %)	– 5.83* (0.66)	– 5.12* (0.58)	– 7.88* (0.65)	– 4.00* (0.49)	– 7.02* (0.39)	– 7.19* (0.40)	– 5.64* (0.39)	– 6.40* (0.42)	– 9.08* (0.46)	– 8.99* (0.56)	– 5.75* (0.58)
Black (Pop %)	– 5.15* (0.41)	– 4.75* (0.39)	– 5.15* (0.47)	– 4.03* (0.43)	– 5.47* (0.45)	– 5.49* (0.46)	– 4.37* (0.43)	– 4.63* (0.43)	– 5.47* (0.36)	– 6.50* (0.42)	– 4.60* (0.42)
American Indian/ Alaskan Native (Pop %)	15.32* (1.38)	15.33* (1.38)	5.50* (0.95)	– 4.59* (1.13)	– 11.82* (2.03)	– 12.14* (2.00)	– 8.01* (2.05)	– 16.31* (3.23)	6.18* (1.99)	6.49* (1.99)	5.75* (1.33)
Latin/Hispanic (Pop %)	– 5.01* (0.35)	– 4.55* (0.33)	– 4.89* (0.41)	– 3.68* (0.41)	– 3.87* (0.42)	– 3.95* (0.43)	– 3.77* (0.45)	– 3.48* (0.35)	– 1.92* (0.31)	– 3.33* (0.29)	– 3.96* (0.39)
Treated Units	12	12	12	12	12	12	12	12	12	12	12
Total Units	88	88	89	87	92	92	92	90	92	92	87
Adjusted R-Squared	0.38	0.39	0.4	0.36	0.32	0.33	0.32	0.42	0.33	0.34	0.44

** $p < 0.05$.*

CFHP = crime-free housing policy. Pop = population.

Source: Authors

Exhibit D.7

Regression Results by Map Rotation for Riverside, Adjusted Spatial First Differences Model Using Binary Treatment

	Angle=0	Angle=30	Angle=60	Angle=90	Angle=120	Angle=150	Angle=180	Angle=210	Angle=240	Angle=270	Angle=300
One or More CFHP units	0.35* (0.09)	0.34* (0.11)	0.21* (0.11)	0.67* (0.12)	0.39* (0.10)	0.80* (0.09)	0.88* (0.06)	0.65* (0.09)	1.21* (0.12)	1.16* (0.13)	0.65* (0.11)
Rental Units (in hundreds)	0.88* (0.02)	0.88* (0.03)	0.90* (0.03)	0.83* (0.03)	0.83* (0.02)	0.72* (0.02)	0.61* (0.02)	0.62* (0.03)	0.66* (0.02)	0.67* (0.03)	0.73* (0.03)
Per Capita Income (in $10,000)	0.19* (0.02)	0.26* (0.02)	0.26* (0.02)	0.24* (0.02)	0.36* (0.02)	0.23* (0.03)	0.18* (0.02)	0.13* (0.03)	0.14* (0.04)	0.24* (0.03)	0.11* (0.03)
Asian (Pop %)	0.50 (0.57)	– 5.30* (0.89)	– 4.83* (0.93)	– 3.33* (1.08)	1.82 (1.20)	2.21 (1.20)	4.75* (1.16)	6.89* (1.01)	6.24* (1.11)	5.93* (0.63)	2.18* (0.51)
White (Pop %)	– 6.37* (0.36)	– 8.84* (0.59)	– 8.18* (0.65)	– 8.50* (0.60)	– 6.26* (0.70)	– 4.35* (0.85)	0.42 (0.61)	– 1.09 (0.81)	– 2.79* (0.88)	– 2.08* (0.59)	– 2.77* (0.54)
Black (Pop %)	– 3.19* (0.20)	– 3.50* (0.27)	– 3.13* (0.29)	– 3.88* (0.26)	– 1.73* (0.29)	– 2.85* (0.31)	– 2.17* (0.26)	– 1.54* (0.24)	– 2.02* (0.32)	– 2.48* (0.24)	– 2.36* (0.17)
American Indian/ Alaskan Native (Pop %)	9.39* (2.01)	9.21* (2.25)	8.26* (2.30)	5.07* (2.31)	4.43* (1.85)	0.88 (1.71)	14.68* (1.74)	9.93* (1.73)	8.42* (2.08)	4.67* (1.99)	7.38* (2.40)
Latin/Hispanic (Pop %)	1.44* (0.31)	– 0.42* (0.18)	– 0.11 (0.20)	0.69* (0.22)	1.07* (0.22)	0.50* (0.24)	2.31* (0.19)	1.95* (0.25)	1.10* (0.34)	2.18* (0.27)	2.37* (0.25)
Treated Units	74	74	74	74	74	74	74	74	74	74	74
Total Units	178	177	175	179	185	185	183	182	181	181	182
Adjusted R-Squared	0.42	0.39	0.39	0.39	0.36	0.32	0.31	0.3	0.32	0.36	0.37

* $p < 0.05$.

CFHP = crime-free housing policy. Pop = population.

Source: Authors

Exhibit D.8

Regression Results by Map Rotation for San Diego County, Adjusted Spatial First Differences Model Using Binary Treatment

	Angle=0	Angle=30	Angle=60	Angle=90	Angle=120	Angle=150	Angle=180	Angle=210	Angle=240	Angle=270	Angle=300
One or More CFHP Units	0.42* (0.16)	0.57* (0.18)	1.03* (0.24)	0.73* (0.19)	1.41* (0.35)	0.57* (0.17)	0.76* (0.23)	0.70* (0.21)	0.85* (0.25)	0.71* (0.17)	0.84* (0.11)
Rental Units (in hundreds)	0.41* (0.07)	0.48* (0.09)	0.39* (0.10)	0.39* (0.10)	0.38* (0.10)	0.40* (0.11)	0.50* (0.09)	0.39* (0.11)	0.36* (0.11)	0.34* (0.11)	0.30* (0.09)
Per Capita Income (in $10,000)	0.01 (0.03)	0.02 (0.03)	0.03 (0.05)	0.07 (0.05)	0.00 (0.03)	0.05 (0.04)	0.18* (0.04)	0.07 (0.04)	0.03 (0.03)	– 0.04 (0.03)	– 0.12* (0.03)
Asian (Pop %)	1.64* (0.37)	1.06* (0.29)	1.22* (0.58)	1.74* (0.67)	0.61 (0.58)	1.04 (0.94)	0.35 (0.69)	0.99 (0.84)	– 0.18 (0.79)	0.26 (0.89)	1.68* (0.40)
White (Pop %)	3.24* (0.54)	2.89* (0.41)	4.53* (0.85)	5.28* (1.16)	1.62 (1.32)	2.60* (1.20)	2.75* (1.02)	3.89* (1.22)	1.48 (1.08)	2.14* (1.07)	2.21* (0.67)
Black (Pop %)	– 0.18 (0.35)	– 0.43* (0.21)	– 0.40 (0.45)	– 0.09 (0.47)	– 1.13* (0.50)	– 1.59* (0.59)	– 0.80 (0.59)	0.12 (0.71)	– 1.13 (0.61)	– 0.87 (0.65)	0.29 (0.25)
American Indian/ Alaskan Native (Pop %)	0.69 (1.01)	– 1.13 (0.77)	– 2.61* (0.82)	– 2.29* (0.66)	– 4.20* (0.79)	– 4.54* (0.94)	– 1.03 (0.78)	1.57 (1.09)	– 1.01 (0.74)	– 1.24 (0.99)	– 0.44 (0.84)
Latin/Hispanic (Pop %)	1.13* (0.18)	1.39* (0.18)	0.20 (0.24)	0.64* (0.32)	0.47 (0.42)	– 0.34 (0.37)	– 0.14 (0.39)	0.35 (0.38)	0.66 (0.44)	0.63 (0.49)	1.12* (0.22)
Treated Units	75	75	75	75	75	75	75	75	75	75	75
Total Units	1,590	1,582	1,580	1,579	1,592	1,616	1,611	1,612	1,614	1,593	1,589
Adjusted R-Squared	0.18	0.22	0.18	0.18	0.17	0.18	0.23	0.18	0.17	0.15	0.15

** p < 0.05.*

CFHP = crime-free housing policy. Pop = population.

Source: Authors

Exhibit D.9

Regression Results by Map Rotation for Fremont, Unadjusted Spatial First Differences Model Using Continuous Treatment

	Angle=0	Angle=30	Angle=60	Angle=90	Angle=120	Angle=150	Angle=180	Angle=210	Angle=240	Angle=270	Angle=300
Number of CFHP Units	0.61* (0.03)	0.62* (0.03)	0.53* (0.03)	0.18* (0.02)	0.38* (0.03)	0.48* (0.04)	0.19* (0.02)	0.31* (0.04)	0.30* (0.03)	0.35* (0.03)	0.59* (0.03)
Rental Units (in hundreds)	0.18* (0.01)	0.17* (0.01)	0.11* (0.01)	0.19* (0.01)	0.21* (0.01)	0.13* (0.01)	0.19* (0.01)	0.26* (0.01)	0.17* (0.01)	0.17* (0.01)	0.16* (0.01)
Treated Units	32	32	32	32	32	32	32	32	32	32	32
Total Units	103	102	106	103	106	105	106	105	105	104	102
Adjusted R-Squared	0.28	0.28	0.16	0.16	0.18	0.14	0.15	0.21	0.13	0.16	0.26

** $p < 0.05$.*

CFHP = crime-free housing policy.

Exhibit D.10

Regression Results by Map Rotation for Hayward, Unadjusted Spatial First Differences Model Using Continuous Treatment

	Angle=0	Angle=30	Angle=60	Angle=90	Angle=120	Angle=150	Angle=180	Angle=210	Angle=240	Angle=270	Angle=300
Number of CFHP Units	1.20* (0.04)	1.21* (0.04)	0.83* (0.04)	1.16* (0.05)	0.96* (0.05)	0.96* (0.05)	0.95* (0.06)	0.78* (0.05)	0.77* (0.03)	1.08* (0.04)	1.21* (0.05)
Rental Units (in hundreds)	0.40* (0.02)	0.39* (0.02)	0.51* (0.01)	0.45* (0.02)	0.46* (0.02)	0.47* (0.02)	0.48* (0.02)	0.60* (0.02)	0.50* (0.01)	0.47* (0.01)	0.48* (0.01)
Treated Units	12	12	12	12	12	12	12	12	12	12	12
Total Units	88	88	89	87	92	92	92	90	92	92	87
Adjusted R-Squared	0.38	0.37	0.35	0.36	0.28	0.28	0.28	0.34	0.29	0.33	0.41

** $p < 0.05$.*

CFHP = crime-free housing policy.

Source: Authors

Exhibit D.11

Regression Results by Map Rotation for Riverside, Unadjusted Spatial First Differences Model Using Continuous Treatment

	Angle=0	Angle=30	Angle=60	Angle=90	Angle=120	Angle=150	Angle=180	Angle=210	Angle=240	Angle=270	Angle=300
Number of CFHP Units	0.21* (0.03)	0.47* (0.03)	0.42* (0.03)	0.52* (0.04)	0.41* (0.03)	0.38* (0.03)	0.26* (0.03)	0.17* (0.04)	0.26* (0.05)	0.17* (0.06)	0.20* (0.04)
Rental Units (in hundreds)	0.79* (0.02)	0.66* (0.02)	0.68* (0.03)	0.64* (0.03)	0.64* (0.02)	0.61* (0.03)	0.61* (0.02)	0.63* (0.03)	0.66* (0.03)	0.75* (0.03)	0.72* (0.03)
Treated Units	74	74	74	74	74	74	74	74	74	74	74
Total Units	178	177	175	179	185	185	183	182	181	181	182
Adjusted R-Squared	0.38	0.4	0.41	0.4	0.37	0.32	0.27	0.27	0.28	0.32	0.34

* $p < 0.05$.

CFHP = crime-free housing policy.

Exhibit D.12

Regression Results by Map Rotation for San Diego County, Unadjusted Spatial First Differences Model Using Continuous Treatment

	Angle=0	Angle=30	Angle=60	Angle=90	Angle=120	Angle=150	Angle=180	Angle=210	Angle=240	Angle=270	Angle=300
Number of CFHP Units	0.14* (0.05)	0.24* (0.06)	0.44* (0.08)	0.40* (0.07)	0.77* (0.18)	0.24* (0.06)	0.13 (0.07)	0.21* (0.06)	0.31* (0.06)	0.37* (0.06)	0.51* (0.06)
Rental Units (in hundreds)	0.43* (0.07)	0.50* (0.09)	0.40* (0.10)	0.40* (0.10)	0.40* (0.09)	0.41* (0.11)	0.50* (0.09)	0.40* (0.11)	0.37* (0.10)	0.36* (0.10)	0.32* (0.09)
Treated Units	75	75	75	75	75	75	75	75	75	75	75
Total Units	1,590	1,582	1,580	1,579	1,592	1,616	1,611	1,612	1,614	1,593	1,589
Adjusted R-Squared	0.18	0.21	0.17	0.17	0.18	0.17	0.22	0.18	0.17	0.14	0.14

* $p < 0.05$.

CFHP = crime-free housing policy.

Source: Authors

Exhibit D.13

Regression Results by Map Rotation for Fremont, Adjusted Spatial First Differences Model Using Continuous Treatment

	Angle=0	Angle=30	Angle=60	Angle=90	Angle=120	Angle=150	Angle=180	Angle=210	Angle=240	Angle=270	Angle=300
Number of CFHP Units	0.56* (0.03)	0.59* (0.03)	0.51* (0.03)	0.16* (0.02)	0.37* (0.03)	0.37* (0.03)	0.14* (0.02)	0.23* (0.04)	0.24* (0.03)	0.25* (0.04)	0.56* (0.03)
Rental Units (in hundreds)	0.18* (0.01)	0.17* (0.01)	0.12* (0.01)	0.21* (0.01)	0.20* (0.01)	0.11* (0.01)	0.22* (0.01)	0.25* (0.01)	0.17* (0.01)	0.16* (0.01)	0.16* (0.01)
Per Capita Income (in $10,000)	– 0.05* (0.02)	– 0.03 (0.02)	0.01 (0.02)	0.00 (0.02)	– 0.08* (0.02)	– 0.19* (0.03)	– 0.03 (0.02)	– 0.07* (0.02)	– 0.08* (0.02)	– 0.08* (0.02)	– 0.03 (0.02)
Asian (Pop %)	0.36 (0.27)	0.35 (0.26)	– 1.74* (0.35)	– 0.11 (0.16)	– 0.88* (0.44)	0.98* (0.35)	– 0.73* (0.32)	– 1.43* (0.44)	0.23 (0.35)	– 0.31 (0.30)	0.52 (0.27)
White (Pop %)	5.14* (0.67)	4.95* (0.64)	0.72 (0.70)	3.62* (0.44)	1.12* (0.53)	4.20* (0.58)	0.19 (0.38)	1.75* (0.84)	3.95* (0.66)	5.62* (0.57)	4.78* (0.56)
Black (Pop %)	1.37* (0.23)	1.40* (0.23)	– 0.69* (0.33)	2.12* (0.18)	– 0.32 (0.34)	1.07* (0.25)	0.08 (0.22)	– 1.31* (0.33)	0.52* (0.24)	– 0.61* (0.22)	1.55* (0.23)
American Indian/ Alaska Native (Pop %)	27.96* (2.62)	29.70* (2.52)	17.60* (1.95)	27.43* (2.32)	25.20* (2.03)	33.97* (2.68)	28.91* (2.41)	34.33* (3.62)	32.55* (3.07)	33.57* (2.70)	27.10* (2.50)
Latin/Hispanic (Pop %)	– 0.85* (0.33)	– 0.93* (0.33)	– 2.05* (0.45)	– 2.53* (0.19)	– 3.66* (0.41)	0.32 (0.40)	– 1.35* (0.33)	– 4.19* (0.60)	– 0.56 (0.43)	– 1.28* (0.40)	– 1.68* (0.37)
Treated Units	32	32	32	32	32	32	32	32	32	32	32
Total Units	103	102	106	103	106	105	106	105	105	104	102
Adjusted R-Squared	0.28	0.28	0.13	0.2	0.19	0.15	0.14	0.24	0.13	0.17	0.27

** $p < 0.05$.*

CFHP = crime-free housing policy. Pop = population.

Source: Authors

Exhibit D.14

Regression Results by Map Rotation for Hayward, Adjusted Spatial First Differences Model Using Continuous Treatment

	Angle=0	Angle=30	Angle=60	Angle=90	Angle=120	Angle=150	Angle=180	Angle=210	Angle=240	Angle=270	Angle=300
Number of CFHP Units	1.06* (0.06)	1.07* (0.06)	0.90* (0.06)	1.13* (0.06)	1.03* (0.07)	1.03* (0.07)	0.98* (0.07)	0.65* (0.06)	0.87* (0.03)	1.08* (0.04)	1.14* (0.06)
Rental Units (in hundreds)	0.50* (0.02)	0.50* (0.02)	0.57* (0.02)	0.47* (0.02)	0.49* (0.03)	0.50* (0.03)	0.53* (0.03)	0.71* (0.02)	0.56* (0.02)	0.53* (0.02)	0.57* (0.02)
Per Capita Income (in $10,000)	0.06* (0.03)	0.13* (0.03)	0.08* (0.04)	0.06 (0.03)	0.33* (0.07)	0.33* (0.07)	0.40* (0.08)	0.76* (0.07)	0.36* (0.06)	0.29* (0.06)	0.01 (0.04)
Asian (Pop %)	– 3.73* (0.42)	– 3.31* (0.46)	– 3.05* (0.43)	– 2.68* (0.56)	– 4.14* (0.62)	– 4.24* (0.63)	– 3.52* (0.64)	– 3.92* (0.60)	– 2.89* (0.60)	– 3.93* (0.48)	– 1.12* (0.56)
White (Pop %)	– 5.90* (0.64)	– 5.20* (0.56)	– 8.80* (0.63)	– 4.81* (0.46)	– 7.76* (0.38)	– 7.91* (0.39)	– 6.19* (0.38)	– 6.85* (0.43)	– 9.78* (0.46)	– 9.83* (0.55)	– 5.93* (0.56)
Black (Pop %)	– 4.84* (0.44)	– 4.50* (0.41)	– 4.88* (0.47)	– 3.99* (0.45)	– 5.32* (0.48)	– 5.34* (0.48)	– 4.13* (0.45)	– 4.39* (0.45)	– 6.01* (0.36)	– 6.55* (0.45)	– 4.19* (0.46)
American Indian/ Alaska Native (Pop %)	12.92* (1.57)	13.00* (1.56)	4.92* (1.04)	– 5.35* (1.03)	– 12.81* (2.00)	– 13.11* (1.97)	– 8.48* (2.04)	– 11.59* (2.92)	4.31* (1.94)	4.41* (1.93)	7.36* (1.44)
Latin/Hispanic (Pop %)	– 4.30* (0.39)	– 3.85* (0.35)	– 4.87* (0.42)	– 3.33* (0.43)	– 3.67* (0.42)	– 3.74* (0.42)	– 3.54* (0.44)	– 3.80* (0.33)	– 2.03* (0.30)	– 3.03* (0.29)	– 3.10* (0.43)
Treated Units	12	12	12	12	12	12	12	12	12	12	12
Total Units	88	88	89	87	92	92	92	90	92	92	87
Adjusted R-Squared	0.4	0.4	0.39	0.35	0.3	0.3	0.29	0.39	0.34	0.37	0.42

** $p < 0.05$.*

CFHP = crime-free housing policy. Pop = population.

Source: Authors

Exhibit D.15

Regression Results by Map Rotation for Riverside, Adjusted Spatial First Differences Model Using Continuous Treatment

	Angle=0	Angle=30	Angle=60	Angle=90	Angle=120	Angle=150	Angle=180	Angle=210	Angle=240	Angle=270	Angle=300
Number of CFHP Units	0.31* (0.03)	0.46* (0.03)	0.43* (0.04)	0.59* (0.04)	0.47* (0.04)	0.47* (0.04)	0.37* (0.04)	0.26* (0.05)	0.39* (0.05)	0.28* (0.06)	0.31* (0.05)
Rental Units (in hundreds)	0.78* (0.02)	0.70* (0.03)	0.72* (0.03)	0.64* (0.02)	0.66* (0.03)	0.59* (0.03)	0.53* (0.03)	0.59* (0.04)	0.64* (0.04)	0.69* (0.04)	0.67* (0.03)
Per Capita Income (in $10,000)	0.21* (0.02)	0.26* (0.02)	0.27* (0.02)	0.24* (0.02)	0.40* (0.02)	0.24* (0.03)	0.17* (0.02)	0.16* (0.02)	0.22* (0.04)	0.25* (0.03)	0.15* (0.02)
Asian (Pop %)	0.45 (0.58)	– 4.75* (0.92)	– 4.31* (0.97)	– 3.03* (1.07)	2.04 (1.21)	2.34 (1.23)	5.30* (1.15)	7.20* (1.13)	6.51* (1.23)	6.52* (0.76)	2.25* (0.51)
White (Pop %)	– 5.62* (0.36)	– 7.10* (0.55)	– 6.49* (0.63)	– 6.21* (0.54)	– 4.86* (0.63)	– 2.37* (0.92)	2.86* (0.74)	– 0.05 (0.98)	– 1.59 (1.08)	– 0.44 (0.85)	– 1.93* (0.60)
Black (Pop %)	– 3.79* (0.19)	– 4.10* (0.34)	– 3.87* (0.34)	– 4.55* (0.28)	– 2.40* (0.24)	– 3.44* (0.32)	– 2.19* (0.24)	– 1.75* (0.23)	– 2.72* (0.32)	– 2.40* (0.20)	– 2.69* (0.17)
American Indian/ Alaska Native (Pop %)	6.67* (2.12)	5.88* (2.24)	4.84* (2.30)	2.02 (2.29)	1.09 (1.91)	– 0.79 (1.80)	12.53* (1.85)	9.54* (1.86)	7.92* (2.29)	3.33 (2.34)	4.91 (2.73)
Latin/Hispanic (Pop %)	1.70* (0.31)	– 0.23 (0.19)	0.14 (0.21)	1.00* (0.25)	1.41* (0.22)	0.88* (0.27)	2.60* (0.23)	2.01* (0.30)	1.31* (0.41)	2.48* (0.33)	2.62* (0.27)
Treated Units	74	74	74	74	74	74	74	74	74	74	74
Total Units	178	177	175	179	185	185	183	182	181	181	182
Adjusted R-Squared	0.43	0.42	0.42	0.43	0.39	0.34	0.32	0.3	0.31	0.35	0.38

** $p < 0.05$.*

CFHP = crime-free housing policy. Pop = population.

Source: Authors

Exhibit D.16

Regression Results by Map Rotation for San Diego County, Adjusted Spatial First Differences Model Using Continuous Treatment

	Angle=0	Angle=30	Angle=60	Angle=90	Angle=120	Angle=150	Angle=180	Angle=210	Angle=240	Angle=270	Angle=300
Number of CFHP Units	0.10 (0.05)	0.20* (0.07)	0.45* (0.07)	0.40* (0.06)	0.76* (0.16)	0.26* (0.06)	0.16* (0.07)	0.21* (0.05)	0.28* (0.05)	0.33* (0.05)	0.44* (0.06)
Rental Units (in hundreds)	0.41* (0.07)	0.48* (0.09)	0.39* (0.10)	0.39* (0.10)	0.39* (0.09)	0.40* (0.11)	0.50* (0.09)	0.39* (0.11)	0.37* (0.11)	0.34* (0.11)	0.30* (0.09)
Per Capita Income (in $10,000)	0.01 (0.03)	0.02 (0.03)	0.03 (0.05)	0.07 (0.05)	0.01 (0.03)	0.05 (0.04)	0.18* (0.04)	0.07 (0.04)	0.03 (0.03)	– 0.04 (0.03)	– 0.12* (0.03)
Asian (Pop %)	1.63* (0.37)	1.03* (0.29)	1.22* (0.60)	1.71* (0.66)	0.53 (0.55)	1.01 (0.93)	0.33 (0.69)	0.99 (0.85)	– 0.18 (0.81)	0.22 (0.86)	1.68* (0.40)
White (Pop %)	3.23* (0.54)	2.86* (0.40)	4.51* (0.84)	5.24* (1.14)	1.51 (1.26)	2.55* (1.19)	2.74* (1.03)	3.86* (1.21)	1.46 (1.09)	2.09* (1.05)	2.19* (0.67)
Black (Pop %)	– 0.20 (0.35)	– 0.48* (0.22)	– 0.43 (0.44)	– 0.12 (0.46)	– 1.18* (0.48)	– 1.63* (0.57)	– 0.83 (0.58)	0.08 (0.70)	– 1.17 (0.61)	– 0.93 (0.63)	0.26 (0.25)
American Indian/ Alaska Native (Pop %)	0.73 (1.01)	– 1.11 (0.77)	– 2.60* (0.81)	– 2.36* (0.66)	– 4.25* (0.78)	– 4.57* (0.92)	– 0.89 (0.77)	1.63 (1.13)	– 0.97 (0.74)	– 1.32 (1.01)	– 0.59 (0.86)
Latin/Hispanic (Pop %)	1.14* (0.18)	1.40* (0.19)	0.24 (0.25)	0.62 (0.32)	0.42 (0.41)	– 0.34 (0.38)	– 0.10 (0.41)	0.40 (0.41)	0.71 (0.48)	0.61 (0.50)	1.12* (0.23)
Treated Units	75	75	75	75	75	75	75	75	75	75	75
Total Units	1,590	1,582	1,580	1,579	1,592	1,616	1,611	1,612	1,614	1,593	1,589
Adjusted R-Squared	0.18	0.22	0.18	0.18	0.18	0.18	0.23	0.18	0.17	0.15	0.15

** $p < 0.05$.*

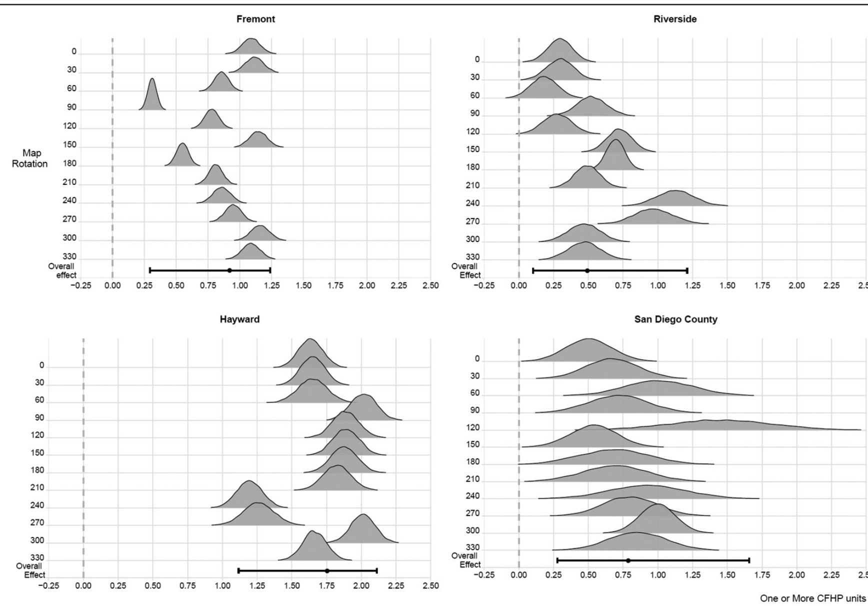
CFHP = crime-free housing policy. Pop = population.

Source: Authors

Appendix E. Estimated Average Treatment on the Treated Effects by Map Rotation, Model, and Location

Exhibit E.1

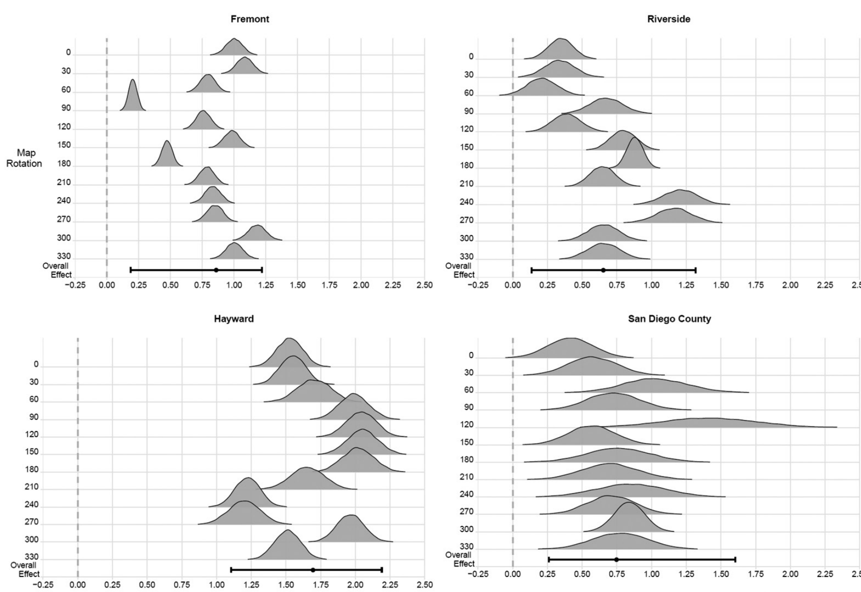
Unadjusted Models with Binary Treatment



Source: Authors

Exhibit E.2

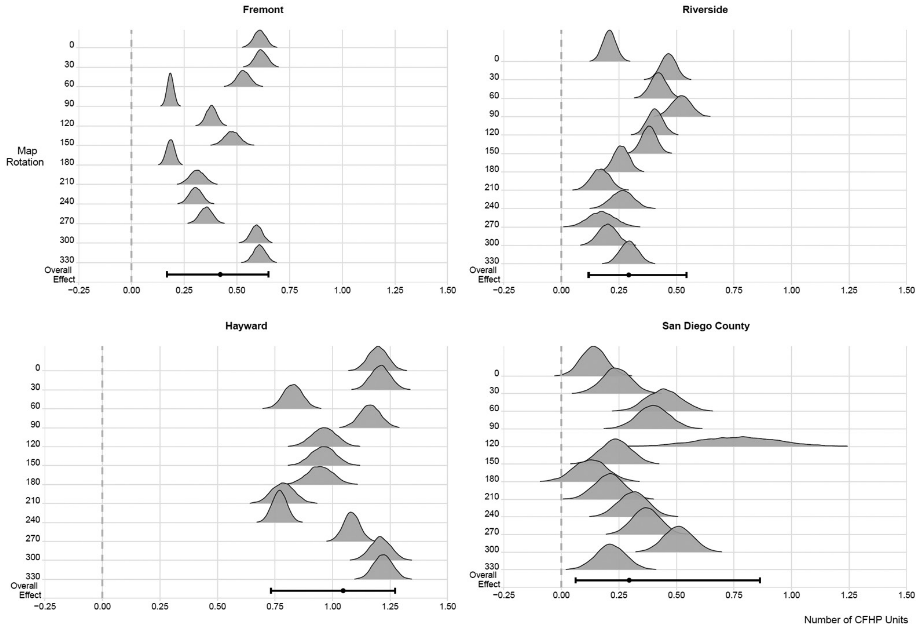
Adjusted Models with Binary Treatment



Source: Authors

Exhibit E.3

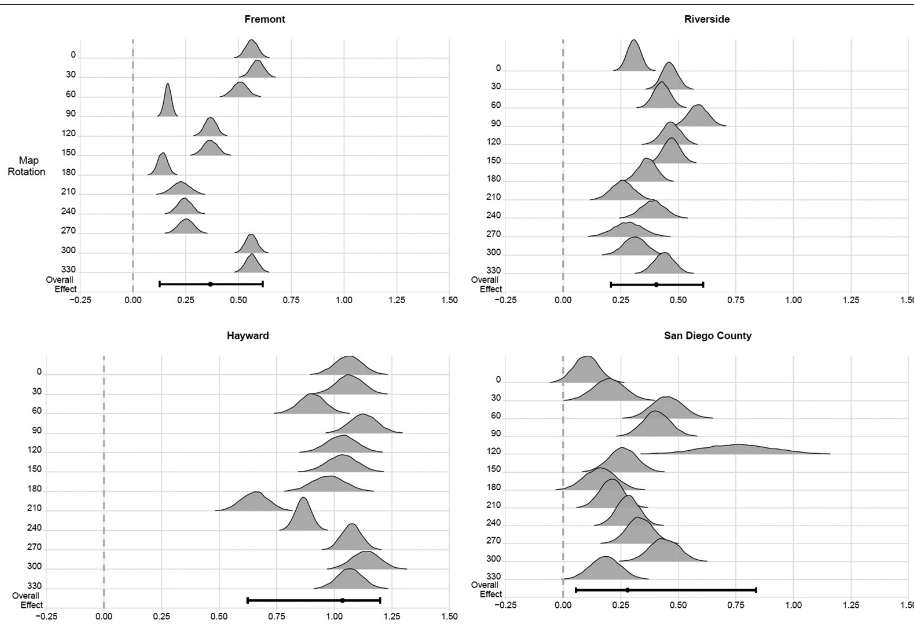
Unadjusted Models with Continuous Treatment



Source: Authors

Exhibit E.4

Adjusted Models with Continuous Treatment

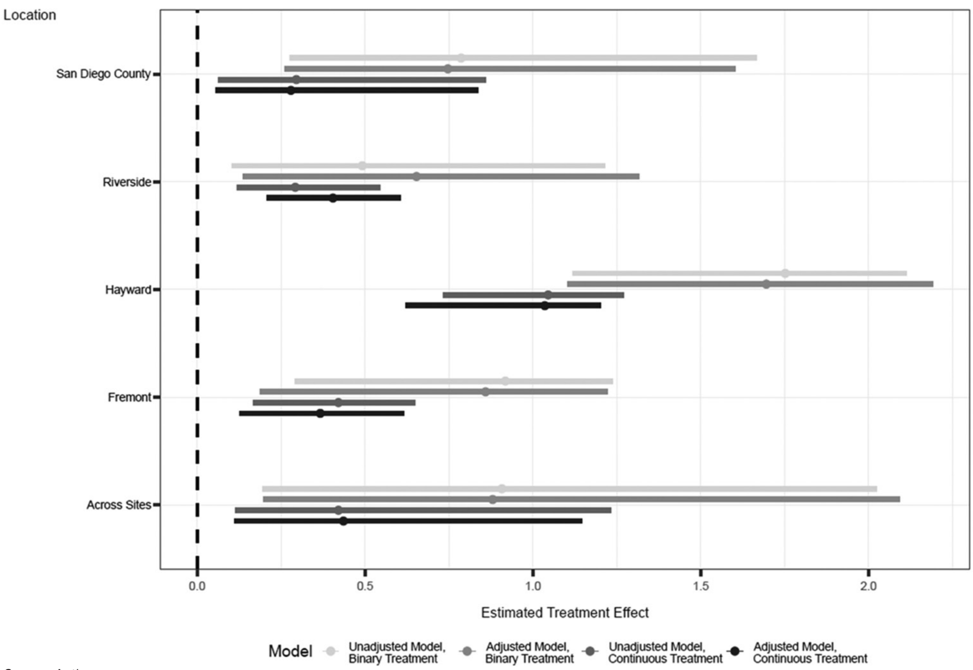


Source: Authors

Appendix F. Estimated Average Treatment on the Treated Effects by Model and Location

Exhibit F.1

Plot of Estimated Average Treatment on the Treated Effects by Model and Location



Source: Authors

Exhibit F.2

Table of Estimated Average Treatment on the Treated Effects by Model and Location (1 of 2)

Location	Specification	Treatment Measure	Estimated ATT
San Diego County	Unadjusted	Binary	0.786 (0.127, 1.683)
San Diego County	Adjusted	Binary	0.74 (0.077, 1.625)
San Diego County	Unadjusted	Continuous	0.3 (0.029, 0.887)
San Diego County	Adjusted	Continuous	0.279 (0.015, 0.858)
Riverside	Unadjusted	Binary	0.493 (0.104, 1.211)
Riverside	Adjusted	Binary	0.623 (0.128, 1.325)
Riverside	Unadjusted	Continuous	0.292 (0.119, 0.545)
Riverside	Adjusted	Continuous	0.419 (0.203, 0.592)
Hayward	Unadjusted	Binary	1.753 (1.117, 2.112)
Hayward	Adjusted	Binary	1.622 (1.078, 2.117)
Hayward	Unadjusted	Continuous	1.046 (0.733, 1.272)
Hayward	Adjusted	Continuous	0.986 (0.577, 1.171)
Fremont	Unadjusted	Binary	0.918 (0.293, 1.239)
Fremont	Adjusted	Binary	0.816 (0.259, 1.189)

Exhibit F.2

Table of Estimated Average Treatment on the Treated Effects by Model and Location (2 of 2)

Location	Specification	Treatment Measure	Estimated ATT
Fremont	Unadjusted	Continuous	0.421 (0.167, 0.65)
Fremont	Adjusted	Continuous	0.297 (0.066, 0.587)
Across Sites	Unadjusted	Binary	0.907 (0.17, 2.024)
Across Sites	Adjusted	Binary	0.869 (0.189, 2.024)
Across Sites	Unadjusted	Continuous	0.421 (0.095, 1.233)
Across Sites	Adjusted	Continuous	0.436 (0.07, 1.121)

ATT = average treatment on the treated.
Source: Authors

Appendix G. Example of a Writ of Execution Record

Exhibit G.1

Synthetic Data Replicating a Writ of Execution Record

A COUNTY SHERIFF'S OFFICE
Eviction List
12/01/2021 TO 12/31/2021

File Number	Occupants	Address	Restoration Date	Time	Status
0000000001	John Doe	2 Mayflower Ave. Marion, CA 28752	01/01/2020	12:00AM	SERVED
[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	CANCELLED
[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	CANCELLED
0000000002	Joe Bloggs	755 Campfire Ave. Hyde Park, CA 02136	01/02/2020	12:00AM	SERVED
0000000003	Jane Smith	936 Brickell Ave. East Brunswick, CA 08816	01/03/2020	12:00AM	SERVED

Source: Authors

Appendix H. Ordinary Least Square Models

Exhibit H.1

Unadjusted Models Using Binary Treatment

	Fremont	Hayward	Riverside	San Diego County
Intercept	0.01 (0.17)	0.01 (0.32)	0.25 (0.24)	0.31* (0.09)
One or More CFHP Units	1.06* (0.32)	1.86* (0.69)	0.32 (0.38)	0.63 (0.34)
Rental Units (in hundreds)	0.20* (0.06)	0.52* (0.10)	0.91* (0.09)	0.51* (0.02)
Treated Units	32	12	74	75
Total Units	118	100	201	1,754
R-Squared	0.31	0.36	0.42	0.24

** $p < 0.05$.*

CFHP = crime-free housing policy.

Source: Authors

Exhibit H.2

Adjusted Models Using Binary Treatment

	Fremont	Hayward	Riverside	San Diego County
Intercept	– 2.94 (2.77)	0.36 (3.10)	0.74 (1.94)	0.93 (1.01)
One or More CFHP Units	0.95* (0.35)	1.93* (0.70)	0.30 (0.38)	0.61 (0.34)
Rental Units (in hundreds)	0.19* (0.06)	0.62* (0.11)	0.88* (0.10)	0.48* (0.02)
Per Capita Income (in $10,000)	– 0.04 (0.13)	0.17 (0.21)	0.21 (0.18)	0.02 (0.05)
Asian (Pop %)	3.27 (2.89)	0.29 (3.27)	1.09 (2.53)	– 1.98 (1.05)
White (Pop %)	6.07 (3.64)	– 3.78 (3.63)	– 0.94 (3.24)	3.38* (1.28)
Black (Pop %)	2.07 (2.34)	– 1.07 (2.40)	2.91* (1.47)	– 1.20 (0.96)
American Indian/Alaska Native (Pop %)	10.17 (8.97)	12.72 (9.95)	1.78 (5.60)	1.82 (2.14)
Latin/Hispanic (Pop %)	1.95 (2.82)	– 1.58 (2.48)	1.49 (1.41)	0.98* (0.40)
Treated Units	32	12	74	75
Total Units	118	100	201	1,754
R-Squared	0.33	0.41	0.45	0.26

** $p < 0.05$.*

CFHP = crime-free housing policy. Pop = population.

Source: Authors

Exhibit H.3

Unadjusted Models Using Continuous Treatment

	Fremont	Hayward	Riverside	San Diego County
Intercept	0.04 (0.17)	0.15 (0.33)	0.32 (0.24)	0.32* (0.09)
Number of CFHP Units	0.48* (0.16)	1.07* (0.42)	0.23 (0.14)	0.18 (0.16)
Rental Units (in hundreds)	0.21* (0.06)	0.49* (0.11)	0.86* (0.09)	0.51* (0.02)
Treated Units	32	12	74	75
Total Units	118	100	201	1,754
R-Squared	0.3	0.36	0.43	0.24

** $p < 0.05$.*

CFHP = crime-free housing policy.

Source: Authors

Exhibit H.4

Adjusted Models Using Continuous Treatment (1 of 2)

	Fremont	Hayward	Riverside	San Diego County
Intercept	– 2.99 (2.80)	0.25 (3.12)	0.77 (1.92)	0.93 (1.01)
Number of CFHP Units	0.42* (0.17)	1.09* (0.42)	0.30* (0.14)	0.16 (0.16)
Rental Units (in hundreds)	0.22* (0.06)	0.58* (0.12)	0.79* (0.10)	0.48* (0.02)
Per Capita Income (in $10,000)	– 0.01 (0.13)	0.15 (0.21)	0.23 (0.18)	0.02 (0.05)
Asian (Pop %)	3.15 (2.95)	0.43 (3.29)	1.57 (2.51)	– 1.99 (1.05)
White (Pop %)	5.46 (3.70)	– 3.21 (3.64)	– 0.31 (3.22)	3.34* (1.28)
Black (Pop %)	1.88 (2.38)	– 0.58 (2.41)	– 4.5844	– 1.16 (0.96)

Source: Authors

Exhibit H.4

Adjusted Models Using Continuous Treatment (2 of 2)

	Fremont	Hayward	Riverside	San Diego County
American Indian/Alaska Native (Pop %)	12.70 (8.91)	14.59 (10.02)	1.03 (5.56)	1.92 (2.15)
Latin/Hispanic (Pop %)	2.30 (2.83)	– 1.56 (2.49)	1.63 (1.40)	0.97* (0.40)
Treated Units	32	12	74	75
Total Units	118	100	201	1,754
R-Squared	0.32	0.4	0.46	0.26

* $p < 0.05$.
CFHP = crime-free housing policy. Pop = population.
Source: Authors

Acknowledgments

The authors thank the anonymous reviewers and editors for improving the quality of this article. They also offer thanks to the Urban Displacement Project at the University of California, Berkeley for assistance in processing the data used within this study and for researchers at RAND Corporation who provided critical feedback on the study's methods.

Authors

Max Griswold is an assistant policy researcher at RAND Corporation. Lawrence Baker is a Ph.D. candidate at RAND Corporation. Sarah B. Hunter is a senior behavioral scientist at RAND Corporation. Jason Ward is an economist at RAND Corporation. Cheng Ren is a lecturer at the University at Albany, State University of New York.

References

Angrist, Joshua D., and Alan B. Krueger. 1999. "Empirical Strategies in Labor Economics." In *Handbook of Labor Economics*, Vol. 3, edited by Orley C. Ashenfelter and David Card. Amsterdam, The Netherlands: Elsevier: 1277–1366. https://doi.org/10.1016/S1573-4463(99)03004-7.

Archer, Deborah. 2019. "The New Housing Segregation: The Jim Crow Effects of Crime-Free Housing Ordinances," *Michigan Law Review* 118 (2): 173–232. https://doi.org/10.36644/mlr.118.2.new.

Buerger, Michael E., and Lorraine Green Mazerolle. 1998. "Third-Party Policing: A Theoretical Analysis of an Emerging Trend," *Justice Quarterly* 15 (2): 301–327. https://doi.org/10.1080/07418829800093761.

Cambon, Jesse, Diego Hernangómez, Christopher Belanger, and Daniel Possenriede. 2021. "Tidygeocoder: An R Package for Geocoding," *Journal of Open Source Software* 6 (65): 3544. https://joss.theoj.org/papers/10.21105/joss.03544.pdf.

Christensen, Peter, Ignacio Sarmiento-Barbieri, and Christopher Timmins. 2021. Racial Discrimination and Housing Outcomes in the United States Rental Market. NBER Working Paper No. 29516. Cambridge, MA: National Bureau of Economic Research. https://doi.org/10.3386/w29516.

City of Fremont. n.d. "Crime Free Lease Addendum." https://www.fremontpolice.gov/home/showpublisheddocument/176/637092421542900000.

City of San Diego. n.d. "Crime Free Lease Addendum." https://www.sandiego.gov/sites/default/files/crime_free_lease_addendum.pdf.

Collinson, Robert, John Eric Humphries, Nicholas Mader, Davin Reed, Daniel I. Tannenbaum, and Winnie van Dijk. 2023. "Eviction and Poverty in American Cities," *The Quarterly Journal of Economics*, September 18. https://doi.org/10.1093/qje/qjad042.

Collinson, Robert, and Davin Reed. 2018. The Effects of Evictions on Low-Income Households. Working paper. https://economics.nd.edu/assets/303258/jmp_rcollinson_1_.pdf.

Conley, Timothy G. 1999. "GMM Estimation with Cross Sectional Dependence," *Journal of Econometrics* 92 (1): 1–45. https://doi.org/10.1016/S0304-4076(98)00084-0.

DeSilver, Drew. 2021. "As National Eviction Ban Expires, a Look at Who Rents and Who Owns in the U.S." Pew Research Center. https://www.pewresearch.org/fact-tank/2021/08/02/as-national-eviction-ban-expires-a-look-at-who-rents-and-who-owns-in-the-u-s/.

Desmond, Matthew, Carl Gershenson, and Barbara Kiviat. 2015. "Forced Relocation and Residential Instability Among Urban Renters," *Social Service Review* 89 (2): 227–262. https://www.journals.uchicago.edu/doi/full/10.1086/681091.

Desmond, Matthew, and Tracey Shollenberger. 2015. "Forced Displacement From Rental Housing: Prevalence and Neighborhood Consequences," *Demography* 52 (5): 1751–1772. https://doi.org/10.1007/s13524-015-0419-9.

Desmond, Matthew, and Nicol Valdez. 2013. "Unpolicing the Urban Poor: Consequences of Third-Party Policing for Inner-City Women," *American Sociological Review* 78 (1): 117–141. https://doi.org/10.1177/0003122412470829.

Druckenmiller, Hannah, and Solomon Hsiang. 2018. Accounting for Unobservable Heterogeneity in Cross Section Using Spatial First Differences. NBER Working Paper No. 25177. Cambridge, MA: National Bureau of Economic Research. https://doi.org/10.3386/w25177.

Falcone, Stefano. 2023. Nuisance Ordinances, Homelessness, and Crimes of Desperation. Working paper. https://www.dropbox.com/scl/fi/z0n3y8bf6820tan7zdcrs/Falcone_Crimes%20OfDesperation.pdf?rlkey=trv2tq7q3re7ev162xtyu4cvy&dl=0.

Garboden, Philip M.E., and Eva Rosen. 2019. "Serial Filing: How Landlords Use the Threat of Eviction," *City & Community* 18 (2): 638–661. https://doi.org/10.1111/cico.12387.

Goff, Phillip Atiba, Tracey Lloyd, Amanda Geller, Steven Raphael, and Jack Glaser. 2016. *The Science of Justice: Race, Arrests, and Police Use of Force*. Los Angeles, CA: Center for Policing Equity. https://policingequity.org/images/pdfs-doc/CPE_SoJ_Race-Arrests-UoF_2016-07-08-1130.pdf.

Golestani, Aria. 2021. "Silenced: Consequences of the Nuisance Property Ordinances." https://ariagolestani.io/wp-content/uploads/2022/03/AriaGolestani_JMP_03102022.pdf.

Goplerud, Dana, and Craig Pollack. 2021. "Prevalence and Impact of Evictions," *Evidence Matters*, Summer. https://www.huduser.gov/portal/periodicals/em/Summer21/highlight2.html.

Graetz, Nick, Carl Gershenson, Peter Hepburn, Sonya R. Porter, Danielle H. Sandler, and Matthew Desmond. 2023. "A Comprehensive Demographic Profile of the US Evicted Population," *Proceedings of the National Academy of Sciences* 120 (41): e2305860120. https://doi.org/10.1073/pnas.2305860120.

Griswold, Max, Stephanie Brooks Holliday, Alex Sizemore, Cheng Ren, Lawrence Baker, Khadesia Howell, Osande A. Osoba, Jhacova Williams, Jason M. Ward, and Sarah B. Hunter. 2023. *An Evaluation of Crime-Free Housing Policies*. Santa Monica, CA: RAND Corporation. https://doi.org/10.7249/RRA2689-1.

Gromis, Ashley, Ian Fellows, James R. Hendrickson, Lavar Edmonds, Lillian Leung, Adam Porton, and Matthew Desmond. 2022. "Estimating Eviction Prevalence Across the United States," *Proceedings of the National Academy of Sciences* 119 (21): e2116169119. https://doi.org/10.1073/pnas.2116169119.

Hanson, Andrew, and Zackary Hawley. 2011. "Do Landlords Discriminate in the Rental Housing Market? Evidence from an Internet Field Experiment in US Cities," *Journal of Urban Economics* 70 (2–3): 99–114. https://doi.org/10.1016/j.jue.2011.02.003.

Hartman, Chester, and David Robinson. 2003. "Evictions: The Hidden Housing Problem," *Housing Policy Debate* 14 (4): 461–501. https://doi.org/10.1080/10511482.2003.9521483.

Hatch, Megan E., and Jinhee Yun. 2020. "Losing Your Home Is Bad for Your Health: Short- and Medium-Term Health Effects of Eviction on Young Adults," *Housing Policy Debate* 31 (3–5): 469–489. https://doi.org/10.1080/10511482.2020.1812690.

Hepburn, Peter, Renee Louis, and Matthew Desmond. 2020. "Racial and Gender Disparities Among Evicted Americans," *Sociological Science* 7: 649–662. https://doi.org/10.15195/v7.a27.

International Crime Free Association (ICFA). n.d.a. "About Crime Free." http://www.crime-free-association.org/about_crime_free.htm.

———. n.d.b. "Crime Free Lease Addendums." http://www.crime-free-association.org/lease_addendums.htm.

Jarwala, Alisha, and Sejal Singh. 2019. "When Disability Is a 'Nuisance': How Chronic Nuisance Ordinances Push Residents with Disabilities out of Their Homes," *Harvard Civil Rights—Civil Liberties Law Review* 54: 875–915.

Kanovsky, Helen R. 2016. "Office of General Counsel Guidance on Application of Fair Housing Act Standards to the Enforcement of Local Nuisance and Crime-Free Housing Ordinances Against Victims of Domestic Violence, Other Crime Victims, and Others Who Require Police or Emergency Services." https://www.hud.gov/sites/documents/FINALNUISANCEORDGDNCE.PDF.

Katach, Salim. 2015. "A Tenant's Procedural Due Process Right in Chronic Nuisance Ordinance Jurisdictions," *Hofstra Law Review* 43 (3): 875–908. https://scholarlycommons.law.hofstra.edu/cgi/viewcontent.cgi?article=2825&context=hlr.

Kroeger, Sarah, and Giulia La Mattina. 2020. "Do Nuisance Ordinances Increase Eviction Risk?" *AEA Papers and Proceedings* 110 (May): 452–456. https://doi.org/10.1257/pandp.20201119.

Leifheit, Kathryn M., Gabriel L. Schwartz, Craig E. Pollack, Maureen M. Black, Kathryn J. Edin, Keri N. Althoff, and Jacky M. Jennings. 2020. "Eviction in Early Childhood and Neighborhood Poverty, Food Security, and Obesity in Later Childhood and Adolescence: Evidence from a Longitudinal Birth Cohort." *SSM [Social Science & Medicine]– Population Health* 11 (August): 100575. https://doi.org/10.1016/j.ssmph.2020.100575.

Lens, Michael C., Kyle Nelson, Ashley Gromis, and Yiwen Kuai. 2020. "The Neighborhood Context of Eviction in Southern California," *City & Community* 19 (4): 912–932. https://doi.org/10.1111/cico.12487.

Lofstrom, Magnus, Joseph Hayes, Brandon Martin, and Deepak Premkumar. 2022. *Racial Disparities in Traffic Stops*. San Francisco, CA: Public Policy Institute of California. https://www.ppic.org/?show-pdf=true&docraptor=true&url=https%3A%2F%2Fwww.ppic.org%2Fpublication%2Fracial-disparities-in-traffic-stops%2F.

Mazerolle, Lorraine Green, and Jan Roehl. 1998. "Civil Remedies and Crime Prevention: An Introduction," *Crime Prevention Studies* 9: 1–18. https://live-cpop.ws.asu.edu/sites/default/files/library/crimeprevention/volume_09/0b_editor_introduction.pdf.

McCain, Demetria L. 2022. "Implementation of the Office of General Counsel's Guidance on Application of Fair Housing Act Standards to the Use of Criminal Records by Providers of Housing and Real Estate-Related Transactions." https://www.hud.gov/sites/dfiles/FHEO/documents/Implementation%20of%20OGC%20Guidance%20on%20Application%20of%20FHA%20Standards%20to%20the%20Use%20of%20Criminal%20Records%20-%20June%2010%202022.pdf.

Moss, Emily, and Gauri Kartini Shastry. 2019. Why She Didn't Just Leave: The Effect of Nuisance Ordinances on Domestic Violence. Honors thesis. Wellesley, MA: Wellesley College. https://repository.wellesley.edu/object/ir909.

Organisation for Economic Co-Operation and Development (OECD). 2020. *OECD Affordable Housing Database Section HC 3.3 Evictions*. Social Policy Division Technical Report. http://www.oecd.org/els/family/HC3-3-Evictions.pdf.

Porton, Adam, Ashley Gromis, and Matthew Desmond. 2021. "Inaccuracies in Eviction Records: Implications for Renters and Researchers," *Housing Policy Debate* 31 (3–5): 377–394. https://doi.org/10.1080/10511482.2020.1748084.

Prochaska, Jenna. 2023. "Breaking Free from 'Crime-Free': State-Level Responses to Harmful Housing Ordinances," *Lewis & Clark Law Review* 27 (1): 259–326. https://heinonline.org/hol-cgi-bin/get_pdf.cgi?handle=hein.journals/lewclr27&section=9.

Ramsey Mason, Kathryn. 2018. "One-Strike 2.0: How Local Governments Are Distorting a Flawed Federal Eviction Law," *UCLA Law Review* 65 (5): 1146–1199. https://ssrn.com/abstract=3949663.

Rubin Donald B. 2004. *Multiple Imputation for Nonresponse in Surveys*. New York: John Wiley and Sons.

San Francisco Anti-Displacement Coalition. 2015. *Eviction Crisis 2015: Trends, Impacts, Real Stories*. San Francisco, CA: San Francisco Anti-Displacement Coalition. http://sfadc.org/2015/04/21/eviction-crisis-2015-trends-impacts-real-stories/.

Semenza, Daniel C., Richard Stansfield, Jessica M. Grosholz, and Nathan W. Link. 2022. "Eviction and Crime: A Neighborhood Analysis in Philadelphia," *Crime & Delinquency* 68 (4): 707–732. https://journals.sagepub.com/doi/pdf/10.1177/00111287211035989.

Smith, Rachel. 2018. "Policing Black Residents as Nuisances: Why Selective Nuisance Law Enforcement Violates the Fair Housing Act," *Harvard Journal on Racial & Ethnic Justice* 34: 87–116. https://heinonline.org/hol-cgi-bin/get_pdf.cgi?handle=hein.journals/hblj34&section=5.

Tanutama, Vincent. 2019. "Sfd-algo [R]." https://github.com/vincentanutama/sfd-algo.

U.S. Census Bureau. 2020. "2019 American Community Survey 2015–2019 5-Year Data Release." https://www.census.gov/newsroom/press-kits/2020/acs-5-year.html.

U.S. Department of Justice. 2022. "Justice Department Secures Landmark Agreement with City and Police Department Ending 'Crime-Free' Rental Housing Program in Hesperia, California." Press release, December 14. https://www.justice.gov/opa/pr/justice-department-secures-landmark-agreement-city-and-police-department-ending-crime-free.

Vásquez-Vera, Hugo, Laia Palència, Ingrid Magna, Carlos Mena, Jaime Neira, and Carme Borrell. 2017. "The Threat of Home Eviction and Its Effects on Health Through the Equity Lens: A Systematic Review," *Social Science & Medicine* 175 (February): 199–208. https://doi.org/10.1016/j.socscimed.2017.01.010.

Werth, Emily. 2013. *The Cost of Being "Crime Free": Legal and Practical Consequences of Crime Free Rental Housing and Nuisance Property Ordinances*. Chicago, IL: Shriver Center on Poverty Law. https://www.povertylaw.org/wp-content/uploads/2019/09/cost-of-being-crime-free.pdf.

Western Regional Chapter of the International Crime Free Association/Crime Free & Partners. 2009. "Crime-Free Multi-Housing Program Manual." https://www.sandiego.gov/sites/default/files/cfmhmanual.pdf.

Zehring, Timothy L. 1994. "The Mesa Crime-Free Multi-Housing Program," *FBI Law Enforcement Bulletin* 63: 8. https://heinonline.org/hol-cgi-bin/get_pdf.cgi?handle=hein.journals/fbileb63&section=39.

Zeng, Zhen. 2022. *Jail Inmates in 2021—Statistical Tables*. Washington, DC: Bureau of Justice Statistics. https://bjs.ojp.gov/library/publications/jail-inmates-2021-statistical-tables.

Toward a National Eviction Data Collection Strategy Using Natural Language Processing

Tim Thomas
Alex Ramiller
University of California, Berkeley

Cheng Ren
University at Albany, State University of New York

Ott Toomet
University of Washington

Abstract

During the past decade, eviction research has relied heavily on preprepared (structured) data from third parties and state agencies who have taken the effort to create readable and accessible filing data. However, massive data gaps across the country exist because third parties may not provide a complete count of filings and many states do not have a formalized process to digitize, enumerate, analyze, or release information on evictions. In some states, the bulk of eviction filings are buried in court filings.

To address this issue, the Eviction Research Network developed a natural language processing (NLP) approach to mine court record images to enumerate and map eviction filing counts at the neighborhood level and help researchers identify disparities by location, race, and gender. This approach involved downloading eviction court record images from online county court systems, digitizing the text, isolating and geocoding addresses, and estimating demographics based on names and location.

In a case study for the State of Washington, millions of pages in more than 110,000 eviction filings from 2004 to 2017 were processed to demonstrate this approach. The research shows massive racial and gender disparities, where up to one in five African-American/Black female-headed households were named in eviction filings. Eviction rates peak in areas with the lowest rent and in the most diverse neighborhoods when analyzing neighborhood dynamics related to eviction. This research helped pass several tenant protection policies in the state and informed other strategies on how to address housing precarity. A suggested strategy for collecting eviction data across the country concludes the article.

Introduction

For renters, there are absolutely no benefits that come from an eviction. At best, the mark of an eviction impedes access to preferable housing for years to come (Franzese, 2018), even if a case had a favorable resolution for the tenant.[1] At worst, it forces vulnerable households to move within an average of 3 weeks or fewer (Davidson, 2019), often to lower-income neighborhoods with higher crime rates (Desmond and Shollenberger, 2015), and increases the odds of homelessness to 1 in 5—even without accounting for related housing precarity risk factors (Shinn et al., 2013). These consequences induce severe economic, mental, physical, and social harms, including higher debt, arears debts, court fees, and security deposits; declining credit scores (Parker and Smith, 2021); food insecurity; lower school performance (Fowler, Henry, and Marcal, 2015); disrupted social ties (Desmond and Shollenberger, 2015); longer commutes to work; unexpected time off or job loss (Desmond and Gershenson, 2017); depression; greater suicide risk; and exposure to infectious diseases (Benfer et al., 2021; Fowler et al., 2015; Hatch and Yun, 2021). Evictions feature severe racial and gender disparities, where the highest eviction rates fall overwhelmingly on the backs of African-American/Black renters, particularly female-headed households (Hepburn, Louis, and Desmond, 2020). As the United States exits pandemic-era renter assistance and policy protections, 19 million U.S. renters burdened with housing costs (U.S. Census Bureau, 2022) face new challenges such as higher inflation, higher costs for food and basic necessities, and even record-breaking eviction rates in some states (Legal Services Corporation, 2023).

Although the harms of eviction are well documented, data for measuring current coverage and trends remain largely fragmented and incomplete (Pan, Zainulbhai, and Robustelli, 2021). These information gaps leave many policymakers woefully uneducated about the trends and extent of the problem in their jurisdiction and, consequently, ill-prepared to act swiftly or advocate for better housing policies. The reason for these gaps is that eviction records are generally processed in jurisdictional court systems that vary widely in recording and storage protocols. Several scholars and institutions have been able to collect structured data from some of the better-organized court systems, which include names, addresses, judgment amounts, and resolutions. Some states only provide county-level counts, which conceal very important details about neighborhood-level processes, such as housing market effects and concentrations in marginalized communities. For the rest of the country, eviction details are buried deep within court filing texts that are either in PDF images or in physical paper form—an almost impossible data source to mine until now.

The focus of this article is to (1) describe how data science tools can be used to extract records from jurisdictions with opaque eviction recordkeeping and (2) illustrate how these tools can supplement existing data collection practices to build a comprehensive national dataset. This type of dataset would allow scholars, policymakers, and the public to analyze in and between jurisdictional trends, measure the severity of the eviction problem, and identify solutions. This supplemental approach uses a natural language processing (NLP) technique to mine court records and fill gaps in missing data for underrepresented and underresourced states and counties. This article applies the NLP method to a case study in Washington State and demonstrates its practical application, hurdles, best practices, and findings. It also describes the political impact

[1] *Smith v. Wasatch Property Management, Inc., et al.*

of enumerating unknown populations, which motivated the adoption of several tenant protection policies across the state, including just cause and extending the state's pay-or-vacate notice period from 3 to 14 days. The article concludes with a recommended outline of steps and tools to apply this approach on a broader scale.

Background

Evictions happen to the most vulnerable citizens in the country when the primary causes at the household level are inadequate minimum wages and insufficient welfare support competing with rising rents (Desmond, 2012). At the neighborhood and county level, rental markets and race are two dominate themes. More specifically, regions that have the lowest median rent, volatile[2] (or gentrifying) housing markets, and higher proportion of African-American/Black tracts in Whiter counties have higher eviction rates (National Academies of Sciences, Engineering, and Medicine, 2023). These analyses draw from a subset of eviction data in the country and require a larger dataset to confirm generalizability and to further explore the nuances among different regions. The primary goal of this study is to collect and structure the most difficult to obtain data to improve coverage and further scholarship; however, before discussing this approach, a general understanding of the eviction process and how these data are structured is required.

The Eviction Process and Data Points

An eviction is not a single event, but rather a process that plays out over time and is documented in varying ways with varying outcomes at each stage. There are five primary stages: (1) prenotice, (2) notice, (3) court filing, (4) writ of restitution, and (5) physical removal. This process can play out in less than 5 days or up to 53 days, depending on varying state rules about how each event is executed (Davidson, 2019).[3] The prenotice—also referred to as an informal or illegal eviction—is not documented and therefore impossible to analyze. Notices are rarely collected, despite being the starting point of the legal process. The most commonly available and studied data are eviction filings, followed by sheriff lockout data. An important caveat about eviction research is that the actual number of renters who face eviction is likely severely undercounted because the sum of notices and informal evictions that precede a court filing is unknown. Estimates suggest that there may be anywhere between two and five-and one-half informal evictions for every formal eviction recorded by the courts (Desmond, Gershenson, and Kiviat, 2015). In addition, outcomes and mobility patterns are difficult to determine because tenants may either move at some point within the five stages or strike a deal with the landlord to stay. On rare occasions, filing data will include judgment resolutions, which allows researchers to make a few assumptions about the outcome (e.g., default judgment means the tenant did not show up to court and the landlord's demands were likely favored in the case).

[2] Volatility is measured as the median rent gap in a county: The degree to which tract median rents are lower or higher than nearby tract rents. In counties where nearly all tracts share similar median rents, little volatility appears in the rental market, and vice versa.

[3] Georgia, Indiana, Maryland, Minnesota, Missouri, New Jersey, North Carolina, Oklahoma, Pennsylvania, and West Virginia combine the notice and court filing period into one event.

Filings may be available in one of three different primary forms: structured data, data within digital images of court records, and data only available on paper records. Structured data generally refer to easily accessible tabular formats that record the attributes of each eviction record, facilitating easy analysis. However, these types of data are only available in certain jurisdictions where court systems or sheriff's offices have decided to record data in that manner. Other jurisdictions may provide aggregate counts of evictions within specific timeframes. More commonly, courts will only have scanned images of civil court records that require computational language tools to mine the text or paper copies, which require the extra step of being scanned.

One of the most important, but often missing, data points is the evictee's address. Addresses allow researchers to get a closer look at eviction trends otherwise concealed by aggregated county counts. Addresses allow for examining neighborhood dynamics and where rates may concentrate within a city and provide a deeper understanding about demographic stratification of eviction through racial and gender estimation. Structured data occasionally provide addresses, but language models are necessary to extract them from court texts.

Current State of National Eviction Data Collection

The current landscape of eviction data collection in the United States consists of several independent organizations and research institutes that each contribute to a more complete national picture of evictions. The data collection strategies of these organizations fall into several main categories. First are organizations such as the Eviction Lab[4] and Legal Services Corporation (LSC)[5] that aim for a truly national data collection strategy, gathering structured data from as many sources as possible, even if those data are only available at higher geographic scales. Second, several local organizations have opted for a more targeted approach, collecting comprehensive data in a single county, region, or state and supplementing this analysis with local knowledge about the legal and technical specificities of the eviction process.[6] A final emergent approach, exemplified by groups such as the Anti-Eviction Mapping Project (AEMP)[7] and the Eviction Research Network[8] involves a combination of these approaches, combining local knowledge with sustained data collection efforts across multiple different jurisdictions.

The most comprehensive data currently available on evictions in the United States are currently held by the Eviction Lab, which uses a combination of data collection techniques to construct an eviction dataset with a national scope. For structured case-level datasets, Eviction Lab makes bulk data requests to state court systems throughout the United States, which has yielded records from 16 states and the District of Columbia. For county-level eviction data, Eviction Lab submits annual requests to state and county court systems, receiving data from 2,204 counties across 46 states.

[4] See Eviction Lab at https://evictionlab.org.

[5] See Legal Services Corporation's Eviction Tracker at https://civilcourtdata.lsc.gov.

[6] See the Atlanta Regional Eviction Tracker at https://metroatlhousing.org/atlanta-region-eviction-tracker/; Root Case Research at https://www.rootcauseresearch.org/lel; The University of Michigan's Eviction page at https://poverty.umich.edu/research-funding-opportunities/data-tools/michigan-evictions/#:~:text=Key%20Findings,6%20rental%20units%20(17%25); and the Richmond Virginia Eviction Lab at https://rampages.us/rvaevictionlab/.

[7] See the Anti-Eviction Mapping Project at https://antievictionmap.com.

[8] See The Eviction Research Network at https://evictionresearch.net/.

Finally, in an effort to fill in the gaps left by these methods, Eviction Lab purchases proprietary data from groups like LexisNexis Risk Solutions that document some cases from the local level but also include manually written, paper-based case management systems (Gromis et al., 2022). Like Eviction Lab, LSC aims for comprehensiveness. Eviction Lab either receives datasets from collaborators or purchases them from third-party data vendors, whereas LSC scrapes digital data from county court systems to create its own structured datasets. LSC has collected data from 32 states, including 18 with complete state coverage. Data from 13 of those 32 states contain name and address information, including 5 with complete state coverage.

As an alternative to this comprehensive approach, which is wide-ranging but less able to capture hard-to-reach records or compare legal differences between jurisdictions, hyper-local approaches to eviction data collection involve collecting data in a single city or region and leveraging those data to explore specific local research questions and challenges. Researchers at Georgia Tech (Raymond et al., 2020), for example, have leveraged the collection of eviction data in the Atlanta region from local county court systems to explore the relationship between evictions and the growing trend toward investor purchases of single-family rental properties, which is a particularly acute problem in Atlanta and other Sun Belt cities. The Evicted in Oregon team at Portland State University (Bates, 2023) has similarly focused on evictions, primarily within the state of Oregon, and gathered knowledge about the local legal landscape through a mixed-methods approach, including analyzing data, conducting interviews, and observing courtrooms. Some organizations have branched out from this local approach by collecting data from multiple different jurisdictions across the country, combining local specificity with an increasingly nationwide scope. This work includes the AEMP, which began by collecting eviction data in the San Francisco area in partnership with local tenant-advocacy organizations and has since expanded its data collection to Los Angeles and New York. Finally, the Eviction Research Network applies data science tools to process, analyze, and visualize collaborator data and to collect data from more difficult areas, such as Washington, California, Chicago, and Baltimore.

The current approaches to eviction data collection reveal an important gap in the landscape of eviction research. Although national-scale efforts by Eviction Lab and LSC have been successful in collecting high-level data for many jurisdictions and detailed records for a smaller number, these organizations are not able to capture less accessible records or data that only appear in unstructured formats. These forms of data may be usable to researchers working on data collection for a single city or region, but a significant amount of resources would be required to collect, clean, and process a larger set of unstructured eviction data, such as digitized PDFs of court records and physical files. This challenge calls for data science tools like NLP that can dramatically streamline the collection and cleaning process for unstructured data.

Data

To demonstrate the NLP approach, court records from Washington State were examined, where counts are only available at the county level. The goal in this project was to fulfill a request by the Washington State legislature to estimate racial and gender disparities, which required both names and addresses of eviction defendants. The data collection process began with a public records request to the state, in collaboration with the Washington Office of Civil Legal Aid, for case numbers that included names, resolution, and judgment amounts for every county in the entire state from 2004 to 2017. Both resolution and amounts were incomplete, and addresses were not provided, requiring access to the court records to mine these elements.

Each state court system operates and holds records differently. For Washington, these records were stored on three different web portals, where access was granted at the discretion of the elected county court clerk. Permission was requested from most of the 39 court clerks in the state, targeting the more populated counties around Puget Sound, near the Oregon border, and Spokane County in eastern Washington. In addition to county clerks, local legal aid providers who operated in these counties were contacted to see if they would advocate to the clerk for access.

The reception for requesting online access was mixed. One county clerk enthusiastically granted permission, and two others willingly provided access through local legal aid contacts; the fourth county was more difficult because the clerk was unwilling to waive the 25 cents per downloaded or printed page charge—an estimated $350,000 cost for all the records from 2004 to 2017. Luckily, a subset of records was provided by a local legal aid provider. The most common reasons for unsuccessful attempts were in part due to clerk unresponsiveness and a lack of support resources. One county clerk said that the state of Washington had the second least funded court system in the country, which might explain the unwillingness among most counties to extend any resources toward a project not mandated by state law.

In the end, full access to Pierce, Snohomish, and Whatcom's online portal was provided for the study period of 2004 to 2017 and limited access to King County's portal, where records were pulled from 2006, 2010, 2011, 2013, and 2017. Using the case numbers provided by the state, an HTML scraper script was built that input the case number in the respective portal and pulled the record at a delayed interval in order to avoid overloading the servers. The scraper downloaded 111,740 PDF or TIFF court record images. Each record held several parts of the eviction proceeding, the most important of which was the eviction summons,[9] which typically listed both the defendant's name and address of residence.

Methods

Extracting Geographic Data from Court Documents

Text extraction proceeded broadly in the following fashion: First, the original document image files were converted to text; the defendant's address was then extracted from the converted text; and finally, the address was geocoded in a census-tract and estimated demographics. The first step of converting images to text was completed using *Tesseract* optical character recognition (OCR) software (Tesseract-OCR, 2023). Because Tesseract cannot handle PDF and TIFF images, a preprocessing step converted these files to PNG raster images at 200 dots per inch (dpi) resolution using the *ImageMagick*® software suite ("ImageMagick," 2023).

[9] An eviction summons is an official notice issued by a court to renters stating that an eviction process has been initiated against them. It typically lists name(s) of the lessees and the landlord(s), address of the premises, and attorneys name(s) if any are involved. The notice also gives lessees a response deadline.

The extracted text quality relied heavily on image quality. Poor quality images can lead to a host of problems, such as causing errors in the text interpretation. For example, the numeral "1" or capital letter "I" turning into the pipe "|" symbol, superscript letters like "th" used in street names transforming into a quote mark """, or in other cases, misspellings due to misidentification of similar-looking characters (e.g., "Seattle" may turn into "Seatile" or "Bellingham" to "Bellmgham"). Another error comes from how court documents are formatted with numerals on the left side of the page, which leads to misclassification of columns (e.g., an address spanning row number 17 and 18 incorporates "17" and "18" in the middle of an address, such as "123 4th street S 17 18 Seattle, WA," where "17" and "18" in the middle of the address are the row numbers). An additional challenge comes from handwritten documents, which Tesseract cannot recognize. Finally, letters and numbers may be concatenated if a separator, such as a space, is missing (e.g., "123 N Jackson Street" may turn into "123N Jackson Street"). Despite these issues, most addresses were extracted correctly.

In the second step, text is taken from the first step where all addresses are extracted using two approaches: *rule-based recognition* using regular expressions and *named entity recognition* (NER) using the neural network–based *spaCy* library ("spaCy", 2023). The rule-based address recognition approach uses a complex set of rules to recognize addresses in the text by looking for house numbers, street type patterns (e.g., *"street," "st," "str," "ave,"* etc.), ZIP Code patterns, and so on, in a given order. Although most addresses follow a clear distinct pattern, several less-common cases are harder to handle, including addresses ending in a county name; missing ZIP Codes or states; names that look like something else, such as "Federal Way" being a city, not a road, and "street court" being a street type, not a separate street and court; long multiword names like "Martin Luther King Jr," which is a single four-word street name; and false positives, such as "Superior Court of the State of Washington," which tends to be misclassified as an address in Washington. The algorithm starts by looking for the address on the first page of the summons file, where they are normally located. If it cannot find an address on that page, the algorithm looks at all other documents in the respective case and extracts all the addresses it can find. Any additional addresses are parsed in a string of components, such as house number, direction, street name, and street type. This information is used afterward to convert the written addresses to a normalized form in which all acronyms are replaced with expanded lowercase words.

As an alternative to the rule-based method, a NER approach using neural networks trains the *spaCy* library to recognize addresses in documents. The training data are generated through multiple steps that include fake addresses; a negative sample of other kinds of similar objects, such as phone numbers and dates; and a set of actual documents with addresses labeled correctly. Based on these training data, the model outputs a large number of addresses (and a few false positive entities). Next, these addresses are parsed into components like in the rule-based method using the *spaCy* neural networks, and this trains the algorithm to distinguish components using 600 manually parsed addresses (including negative examples). Both steps work rather well, but the parser is better at skipping spurious words and symbols related to the problem of numbers in the first column.

The main issues include missing separators between address components (e.g., missing spaces between a street number and prefix). Sometimes, the address components are misidentified or combined in a wrong order. For instance, "123 N Alder street" may be read as "123 n.alder street," and the resulting street name will be "n.alder." Other issues include mistaken conversions between letters and numbers (e.g., prefixes like "s" can be mistaken as a house number like "5"). In this case, the house number will either be wrong or completely missing in the result. Finally, multiword street or city names may confuse the parser, causing it to skip part of the name, mistakenly considering it being a spurious word from another column.

To distinguish defendant addresses from other addresses (e.g., attorney's office), a simple Naive Bayes algorithm is used to look at words surrounding the address and computes the probability that this result is the address of interest. Thereafter, the algorithm picks the address with the largest such probability.

Geocoding

Next, structured addresses are geocoded for spatial analyses and demographic estimation. Several geocoding platforms were tested, each using slightly different databases and methods for geocoding, including Open Street Map (OSM); Census Bureau, Google, Azure, and Esri Business Analyst (BA) data; and Master Address File (MAF) for King County. Free Open Street Map and Census Bureau geocoding platforms struggle with interpreting ambiguous address strings. Pay-as-you-go services Google and Azure operate with a much greater degree of effectiveness, each of which using proprietary fuzzy string-matching algorithms to find matches—even for more ambiguous address strings. BA geocoding data operated through Esri's ArcMap program are purchased up front for 1 year. Finally, the King County MAF data are free to the public and provide a record of every single address location within a local jurisdiction for administrative recordkeeping and emergency services.

Each geocoding platform offers different strengths. OSM and Census Bureau geocoders are optimal in terms of accessibility, although both Google and BA offer various levels of address accuracy from the administrative unit to the rooftop. Crucially, Census Bureau and BA approaches share the advantage of reproducibility because they are based on static benchmarks that can be referenced at any point in the future (although the Census Bureau's dataset is regularly updated, it saves benchmarks at each point in time). Other data sources, such as OSM, Google, and Azure, meanwhile, continually update without retaining previous records, which makes it impossible to reproduce the exact same results at a later point.

Each geocoder was tested on approximately 5,000 manually extracted 2013 King County addresses. These addresses had a success rate of 85.1 percent for OSM, 90.6 percent for the Census Bureau geocoder, 92.1 percent for the King County MAF, 97.8 percent for Esri BA data, and 99.9 percent for Azure. For the NLP addresses, success rates were consistently lower: 75.6 percent for OSM, 77.2 percent for the Census Bureau geocoder, 84.5 percent for the King County MAF, 90 percent for Esri BA data, and 99.9 percent for Azure. The high return rate for Azure includes geocodes at the center of some places: 81.6 percent for "address point," 12.3 percent for "address range," 5.5

percent for “street,” 0.3 percent for “geography,” and 0.3 percent for “cross street.”[10] Azure was the most successful in geocoding the provided addresses; however, in terms of geocoding tools with a high degree of replicability, Esri BA data had the highest rate of success. Azure’s spatial accuracy will range from rooftop coordinate to the centroid of a ZIP Code or county. The final dataset uses a subset of addresses with rooftop to block accuracy, which is about 93 percent of the NLP addresses. Addresses with accuracies at the “street” centroid or “geography” (e.g., county centroid) were omitted because they likely fell within tracts far from their true location.

Validation

Address Validation. Did the correct defendant address get extracted? A Naive Bayes probability approach alone determined that the correct address was obtained with a confidence of 98 percent. Repeated addresses from a large number of cases were also manually analyzed. In most cases, these addresses were apartment complexes, although at least one incorrect place slipped through (e.g., Pierce County courthouse). However, the Naive Bayes probability was relatively low in cases with the 90th percentile being less than 90 percent.

NLP address components were also compared to the manually extracted addresses from 2013, using the average ratio of the Levenshtein distance as a metric of accuracy—a method to measure the differences between two strings. The ratio is calculated by the python-Levenshtein package (version:0.12.2) with the formula:

$$ratio\ (a, b) = \frac{(lensum - ldist)}{lensum} = \frac{((len(a) = len(b)) - dist\ (a, b)}{(len(a) + len(b))}$$

For example, the distance between “apple” and “appl” is 1, and the ratio is 0.89. After comparing the manually extracted addresses and algorithmically extracted addresses, the average ratio was 0.82. In this situation, if the ratio was more than 0.7, the two pairs were usually on the same street or a close street, although when the ratio was lower, the two pairs may be in the different cities. A common error is from algorithmically extracted addresses; the method extracts the address of the court or attorney’s office, rather than the address of the defendant. The final result for this comparison yielded an accuracy of 93 percent.

Geographic validation. Address accuracy was evaluated by comparing the distance between the coordinates of the 5,000 manually extracted addresses and NLP coordinates and determining whether the two addresses landed in the same census tract. When using Azure’s geocoder and isolating locations within the county of the case, the average distance was 3.2 miles. When joining these locations to census tracts, 83.1 percent of the pairs shared the same census tracts on Azure. For OSM, the accuracy was 63.3 percent, which is much lower because OSM was unable to geocode as many addresses as Azure.

[10] Geocoders provide various accuracies for a given address. “Address point” is the highest level of accuracy where the latitude and longitude mark the exact point where the address is located. “Address range” means the latitude and longitude falls within a range of addresses between cross streets, but not at the exact address location. “Cross street” means the latitude and longitude marks the nearest cross street. “Street” means the address number was difficult to find, so the latitude and longitude marks the centroid of the street. This can be problematic when the street is long and crosses several geographies such as census tracts. Finally, “geography” marks the centroid of a much larger geography such as a city, county, or state.

Race and Gender Estimation

With addresses and names, the race of the defendant is estimated by running a first batch of surnames through the R packages *wru* (Khanna et al., 2023) and then *rethnicity* (Xie, 2022) for the few hundred names that could not be estimated in the first round. The ecological inference package *wru* uses the Bayes' Rule to examine the racial probability of a surname compared to the racial composition for each neighborhood (census tracts) where evicted defendants lived.[11,12] This estimates a racial category of White, Black, Latinx, Asian, or other for everyone with geocoded addresses. The few names that were not estimated by the *wru* package are passed through the *rethnicity* package—a Bidirectional Long Short-Term Memory (Bi-LSTM) model based on a recurrent neural network architecture commonly used for natural language processing. This package also uses voter registry records for training data. Although these methods have been widely used in various studies, including eviction research (Hepburn, Louis, and Desmond, 2020; Hepburn, Rutan, and Desmond, 2022; Thomas, 2017; Thomas et al., 2020), they struggle with correctly assigning race and ethnicity in states outside of its training data, which were voter files based in Southern states. Sampled observations of the outputs showed that *wru* did well in estimating Asian, Black, Latinx, and White surnames, whereas *rethnicity* did well with Black and Latinx names.[13] Gender is inferred using the R package *gender* (Mullen, Blevins, and Schmidt, 2023) which cross-validates the first name of an individual with the Social Security Administration (SSA) Name Registry from 1932 to 2012 and the U.S. Census Bureau Integrated Public Use Microdata Series (IPUMS). To see if estimates were close to real-world conditions, demographic estimates of the manually extracted addresses from 2013 were compared to aggregated intake data from a King County tenant attorney's office for that year, and it was found that the compositions of each group between the two datasets were within several percentage points. Although this comparison improves confidence in the demographic estimates, there should still be some caution regarding the reliability and interpretation of these values.

The final dataset for King, Pierce, Snohomish, and Whatcom Counties held 111,740 cases, 91.7 percent of which were geocoded and had racial estimation, 95.7 percent had gender estimation, and 87.7 percent had both race and gender[14] estimation. The data was then subset to 2013–17 to compare the 2017 5-year PUMS data for renter households by race and gender. Racial demographic counts for King County's missing years of data between 2014–16 were interpolated using a linear regression controlling for the data's county, counts, and year.

[11] The most recent version of Khanna et al.'s fully Bayesian Improved Surname Geocoding (fBISG) method is used for the estimates.

[12] For example, a person with the last name Jackson, a common Black surname, living in a high-Black neighborhood would have a greater likelihood of being Black. Whereas the same name found in a high-White neighborhood would have a lesser probability of being Black.

[13] Asian and White estimates in *rethnicity* included more Middle Eastern and non-European names, which *wru* includes in the "Other" category. The final racial estimate consisted primarily of *wru's* estimates where *wru's* "Other" identification was replaced for any names that received a Black or Latinx designation from *rethnicity*. This helped improve a few hundred estimates from "Other" to Black or Latinx.

[14] Race and gender estimates were calculated by multiplying the race and gender probabilities together.

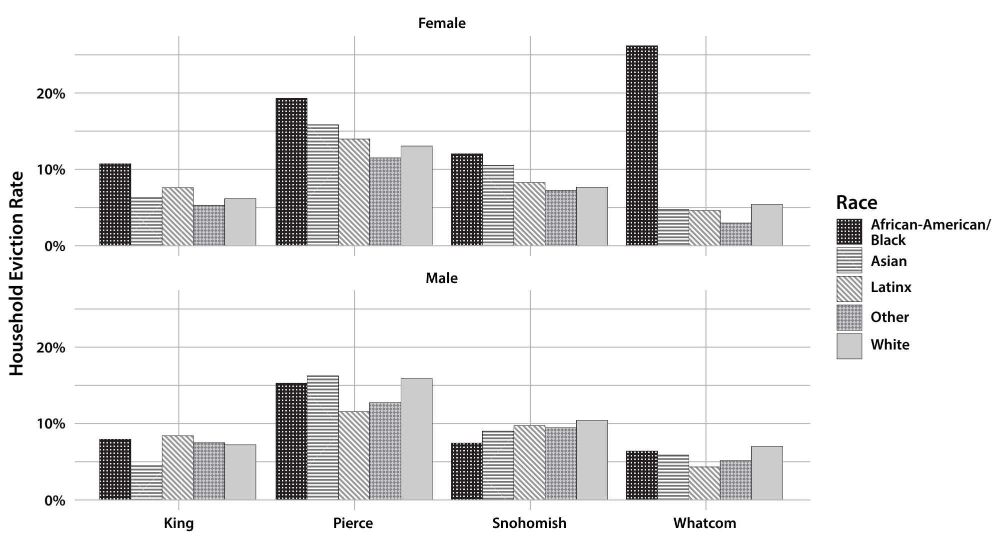
Findings

The NLP address extraction and demographic estimation process revealed sizable racial disparities in the data that would have otherwise been unknown (see exhibit 1). Particularly, one in five (19 percent) of Black female-headed households in Pierce County were named in the process, and one in nine (11 percent) were named in King County. What is important about this finding is that Pierce County was one of the largest recipients of new movers from across the country and from neighboring King County, home of gentrified Seattle (Balk, 2023; 2017). This movement increased demand in the area and applied market pressure on more affordable, and often more racially diverse, communities. King County is unique in that the majority of evictions were located on the south side of the county, which is a common destination for Black families displaced from Seattle (Thomas, 2017). Whatcom County had the smallest number of Black female householder cases (n=15) but the highest rate of 26 percent (one in four households).

King County recorded judgment amounts in less than one-half of all cases, with approximately 41 percent of these cases having estimated race (see exhibit 2, truncated to $10,000). Not much variation in the medians by race exists, but Black households had the lowest median judgment amounts at $2,840.

Exhibit 1

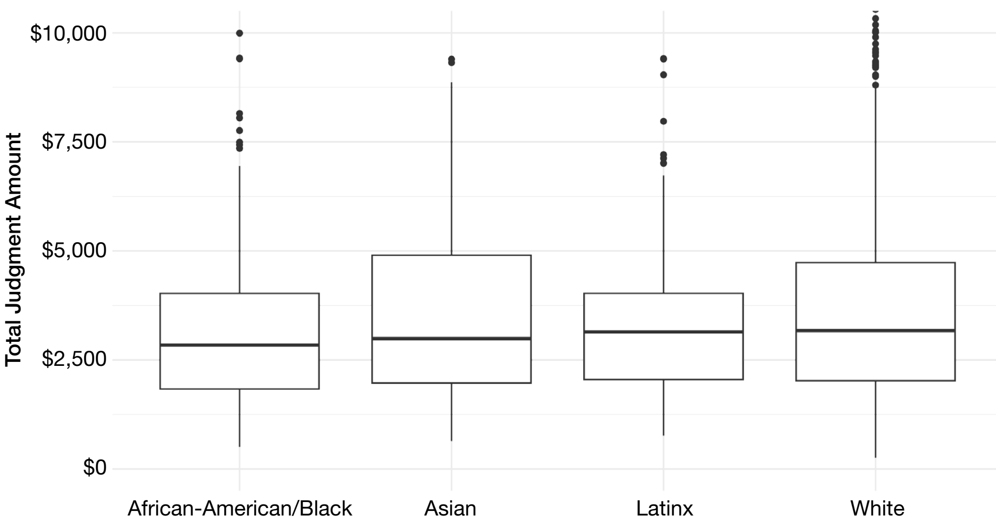
Five Year Eviction Rates by Race and Gender in Puget Sound, 2013–17



Source: Estimated demographics of heads of household with eviction filings over race and sex of renter heads of household (U.S. Census Bureau 2017 5-year American Community Survey)

Exhibit 2

Judgment Amount Totals by Race in King County, 2017



Source: Unlawful detainer judgment amounts from the Washington State Courts and estimated race and ethnicity

Judgment resolution by race shows that default judgments in which the tenant did not show up in court were the most common outcome, with the highest share being among Latinx householders (see exhibit 3). Next was dismissal without trial, which can mean a failure to appear for trial for both parties or that the final documents were not filed. Between 9 percent and 15 percent of cases ended in settlement or agreed judgment, which most likely resulted in some sort of payment plan or agreement to resolve the unlawful detainer.

Exhibit 3

Judgment Resolution by Race in King County, 2017

Resolution	Asian (%)	Black (%)	Latinx (%)	White (%)
Default Judgment	48.3	46.3	**51.6**	48.5
Dismissal Without Trial	**25.4**	21.9	25.0	22.0
Settled by Parties and/or Agreed Judgment	12.7	**14.9**	9.4	12.3
Closed by Court Order After a Hearing	7.6	10.8	**10.9**	10.7
Dismissal by Clerk	2.1	2.3	1.6	**3.4**
Uncontested Resolution	1.7	**1.8**	1.0	1.3
Court Decision After Trial	0.8	0.3	NA	0.2
Summary Judgment	0.8	**1.8**	0.5	1.4
Transferred to Federal Bankruptcy Court	**0.4**	NA	NA	0.2

NA = data not available.

Source: Unlawful detainer judgment resolutions from the Washington State Courts and estimated race and ethnicity

Neighborhood analysis shows that evictions occurred in communities with the most racial diversity and lowest rents in the region. Temporal spatial analysis shows that case frequency was greater in South Seattle's racially diverse neighborhoods in earlier years and then increased outside of the city in South King County, both of which are known displacement destinations for Black, Indigenous, Latinx, and Asian households. In Pierce and Snohomish Counties, evictions concentrated largely in segregated urban spaces.

Future Strategies

Several alternative approaches stand out for future consideration. The first is document layout analysis, which employs computer vision and NLP to extract contents of interest by their location in the image and would help reduce errors such as misidentifying the defendant address. Document layout analysis has its limitations, such as the high cost compared with the technique introduced in this article.[15]

Another promising technique involves the use of large language models like ChatGPT, where researchers can input unstructured court file content and request outputs like the defendant address. This method was tested on the 2013 training data and highlighted several concerns, such as how current models collect data, which can lead to disclosing sensitive information, and the quality of the returned content can sometimes fall short of expectations.[16] Additional efforts may thus be required to transform this output into a structured dataset. Despite these drawbacks, these two methods should lessen barriers for researchers with limited computational skills in text analysis and even help extract more information, such as the cause of eviction or amounts.

Conclusion and Discussion

Combining data science tools with existing eviction data collection efforts can fill in massive gaps. The NLP approach can extract addresses from court record images, creating the opportunity to analyze demographic disparities in the State of Washington and create a clearer picture of who was being evicted and from where. This work also provided evidence-based research for several statewide policy changes, including extending the pay-or-vacate notice period from 3 to 14 days (Senate Committee on Housing Stability & Affordability, 2019) and just cause.

A recommendation to build a more comprehensive national eviction dataset consists of continuing the practice of collecting structured data where available and then gathering court records for text mining in regions with lesser-known trends. This process will more than likely involve public record requests or direct contact with the respective jurisdiction's datakeepers to gain access, possibly requiring document scanning if records are not already in image formats. With records in hand, apply the following NLP process to these texts: (1) conduct OCR to convert images to text using tools like Tesseract or paid services like Azure or Amazon Web Services (AWS); (2) extract content, such as addresses, using rule-based recognition or NER through tools like spaCy;

[15] During the model customization and training process, researchers must manually label a small diverse sample and then employ a graphics processing unit (GPU) to train the model for application. This process necessitates a careful tradeoff between resource expenditure and data accuracy.

[16] For instance, a defendant's address may be requested, but the address is intermingled with the defendant's name.

(3) geocode addresses using tools from the Census Bureau or Azure; (4) validate the address; and optionally, (5) apply demographic estimations using packages like wru, rethnicity, and gender.

Several factors should be considered before scaling this approach nationwide. First, the sociological, demographic, and technical nature of this research requires an interdisciplinary team of scholars to ensure that ethical and accurate data integrity practices and considerations are implemented throughout the process of collecting and engineering these data. Second, engaging with court officers around collection can be difficult, especially if no motivation or resources are available on their side to facilitate cooperation with collectors. This process may require mandates from legislatures, grant support, or other considerations not discussed in this article. The upside is that a comprehensive database now seems like a realistic goal.

Appendix

Exhibit A.1

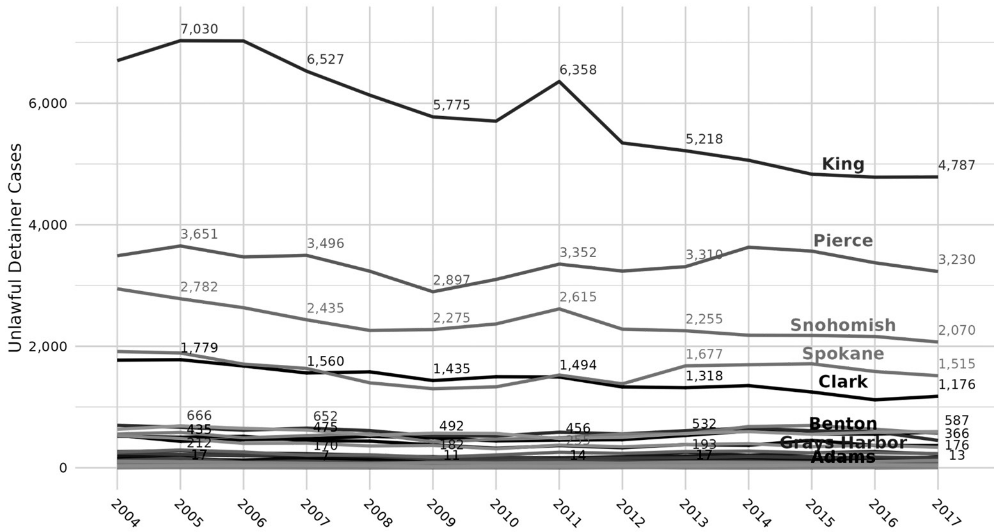
Five Year Eviction Rate by Race and Gender in King, Pierce, Snohomish, and Whatcom Counties, 2013–17

	King	Pierce	Snohomish	Whatcom
Total Cases	24,467	17,042	10,817	1,816
Asian Female Householders	6% (1,233 of 19,663)	16% (460 of 2,905)	10% (337 of 3,215)	5% (27 of 568)
Asian Male Householders	4% (1,420 of 31,607)	16% (453 of 2,777)	9% (390 of 4,321)	6% (29 of 483)
Black Female Householders	11% (1,811 of 16,909)	19% (1,392 of 7,233)	12% (240 of 1,993)	26% (15 of 57)
Black Male Householders	8% (1,330 of 16,727)	15% (1,059 of 6,937)	7% (241 of 3,221)	6% (14 of 215)
Latinx Female Householders	8% (1,315 of 17,431)	14% (754 of 5,389)	8% (547 of 6,618)	5% (69 of 1,480)
Latinx Male Householders	8% (1,654 of 19,654)	12% (751 of 6,502)	10% (749 of 7,675)	4% (62 of 1,411)
Other Race/ Ethnicity Female Householders	5% (698 of 13,040)	12% (674 of 5,847)	7% (260 of 3,563)	3% (42 of 1,401)
Other Race/ Ethnicity Male Householders	8% (774 of 10,274)	13% (653 of 5,109)	10% (309 of 3,247)	5% (45 of 868)
White Female Householders	6% (6,460 of 103,700)	13% (5,011 of 38,386)	8% (3,484 of 45,496)	5% (703 of 12,897)
White Male Householders	7% (7,772 of 106,571)	16% (5,835 of 36,630)	10% (4,259 of 40,777)	7% (812 of 11,518)

Source: Demographic estimates of unlawful detainer heads of household and renter demographics from the U.S. Census Bureau Private Use Microdata Sample

Exhibit A.2

State Unlawful Detainer Raw Count by County

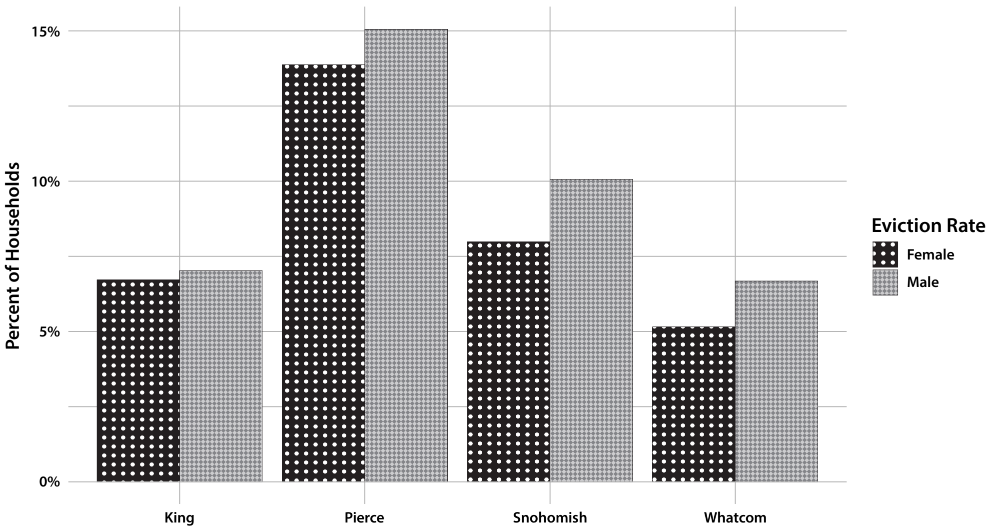


Source: Unlawful detainer filings from the Washington State Courts

County trends show that 5 of Washington's 39 counties had more than 1,000 cases per year, with King County in the lead with less than 5,000 cases in 2017 (see appendix exhibit A.1). Immediately before the Great Recession, state case counts peaked at approximately 22,500 unlawful detainers in 2005 and then declined to approximately 17,500 in 2017. During this time, the Puget Sound region experienced several changes in population growth, housing costs, and demographics. Big technology companies brought in higher-earning workers, spurring gentrification and displacement of Seattle's more vulnerable populations, particularly for Black and brown households. Between 2013 and 2019, King County saw a rapid rent increase from a median of $1,400 to $2,200, which requires a take-home income of approximately $90,000 to avoid rent burden. Black and Latinx median household incomes rested near 50 percent of the area median income of $40,000 and $55,000, which provides little room to afford the cost of living (Thomas et al., 2020).

Exhibit A.3

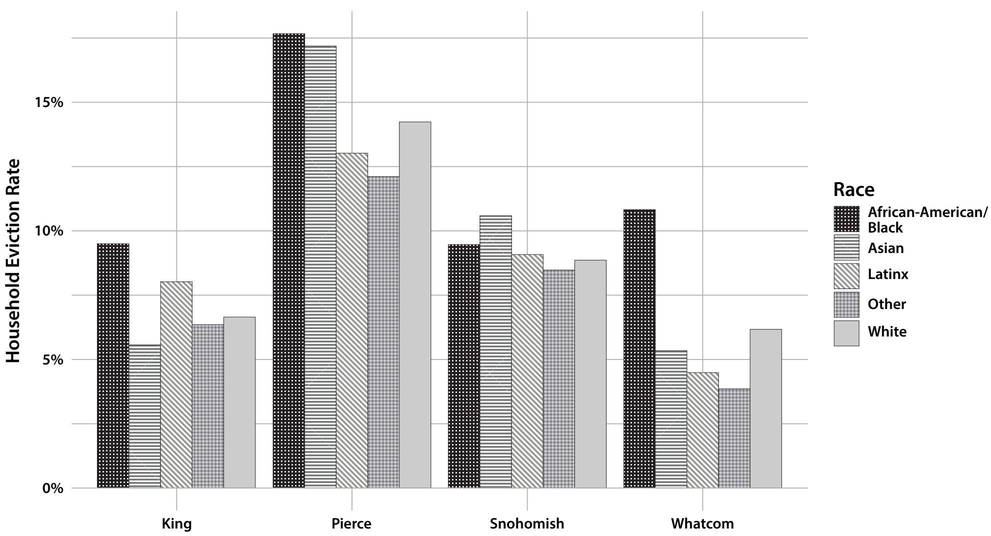
Unlawful Detainer Case Rates by Gender, 2013–17



Source: Estimated gender of heads of household with eviction filings over sex of renter heads of household (U.S. Census Bureau 2017 5-year American Community Survey)

Exhibit A.4

Unlawful Detainer Case Rates by Race, 2013–17



Source: Estimated race and ethnicity of heads of household with eviction filings over race and ethnicity of renter heads of household (U.S. Census Bureau 2017 5-year American Community Survey)

Acknowledgments

This work was funded in part by the Moore/Sloan foundation, Bill and Melinda Gates Foundation, and MacArthur Foundation. The authors would like to also thank the numerous colleagues and organizations that helped make this work possible, including the Snohomish, Pierce, and Whatcom Counties court clerks and tenant attorneys; Edmund Whitter and the King County Bar Association; John Stovall and Washington State Representative Nicole Macri; Jim Bamberger and the Washington Office of Civil Legal Aid; the University of Washington eScience Institute, iSchool, Department of Sociology, and Center for Studies in Demography and Ecology—particularly, Bill Howe, Jose Hernandez, Ian Kennedy, Sarah Stone, and Kyle Crowder; and The University of California Berkeley Institute of Governmental Studies; D-Lab; Berkeley Institute of Data Science; and Karen Chapple and the Urban Displacement Project.

Authors

Tim Thomas, Ph.D., is research director at the University of California, Berkeley (UCB) Urban Displacement Project and director of the Eviction Research Network. Alex Ramiller is a Ph.D. candidate in city and regional planning at UCB. Cheng Ren is a Ph.D. candidate at UCB Department of Social Welfare and lecturer at the University at Albany, State University of New York. Ott Toomet, Ph.D., is an assistant teaching professor at the University of Washington.

References

Balk, Gene. 2023. "When People Move Away from Seattle, Here's Where They Go." *The Seattle Times*, April 4. https://www.seattletimes.com/seattle-news/data/when-people-move-away-from-seattle-heres-where-they-go/.

———. 2017. "New Residents Pour in: Pierce, Snohomish Counties See Nation's Biggest Jump in Movers." *The Seattle Times*, March 27. https://www.seattletimes.com/seattle-news/data/new-residents-pour-in-pierce-snohomish-counties-top-the-nation/.

Bates, Lisa. 2023. "Evicted in Oregon." Portland State University. September 2023. https://www.evictedinoregon.com.

Benfer, Emily A., David Vlahov, Marissa Y. Long, Evan Walker-Wells, J. L. Pottenger, Gregg Gonsalves, and Danya E. Keene. 2021. "Eviction, Health Inequity, and the Spread of COVID-19: Housing Policy as a Primary Pandemic Mitigation Strategy," *Journal of Urban Health* 98 (1): 1–12. https://doi.org/10.1007/s11524-020-00502-1.

Davidson, Michael Scott. 2019. "Despite Changes, Nevada Eviction Law Still Favors Landlords," *Las Vegas Review Journal*, June 28. https://www.reviewjournal.com/local/local-nevada/despite-changes-nevada-eviction-law-still-favors-landlords-1697301/.

Desmond, Matthew. 2012. "Eviction and the Reproduction of Urban Poverty," *American Journal of Sociology* 118 (1): 88–133. https://doi.org/10.1086/666082.

Desmond, Matthew, and Carl Gershenson. 2017. "Who Gets Evicted? Assessing Individual, Neighborhood, and Network Factors," *Social Science Research* 62 (February): 362–77. https://doi.org/10.1016/j.ssresearch.2016.08.017.

Desmond, Matthew, Carl Gershenson, and Barbara Kiviat. 2015. "Forced Relocation and Residential Instability among Urban Renters," *Social Service Review* 89 (2): 227–62. https://doi.org/10.1086/681091.

Desmond, Matthew, and Tracey Shollenberger. 2015. "Forced Displacement From Rental Housing: Prevalence and Neighborhood Consequences," *Demography* 52 (5): 1751–72. https://doi.org/10.1007/s13524-015-0419-9.

Fowler, Katherine A., R. Matthew Gladden, Kevin J. Vagi, Jamar Barnes, and Leroy Frazier. 2015. "Increase in Suicides Associated With Home Eviction and Foreclosure During the US Housing Crisis: Findings From 16 National Violent Death Reporting System States, 2005–2010," *American Journal of Public Health* 105 (2): 311–16. https://doi.org/10.2105/ajph.2014.301945.

Fowler, Patrick J., David B. Henry, and Katherine E. Marcal. 2015. "Family and Housing Instability: Longitudinal Impact on Adolescent Emotional and Behavioral Well-Being," *Social Science Research* 53 (September): 364–74. https://doi.org/10.1016/j.ssresearch.2015.06.012.

Franzese, Paula A. 2018. "A Place to Call Home: Tenant Blacklisting and the Denial of Opportunity," *Fordham Urban Law Journal* 45 (3): 38.

Gromis, Ashley, Ian Fellows, James R. Hendrickson, Lavar Edmonds, Lillian Leung, Adam Porton, and Matthew Desmond. 2022. "Supplementary Information Estimating Eviction Prevalence across the United States," Eviction Lab. April 25. https://evictionlab.org/docs/Eviction_Lab_Methodology_Report_2022.pdf.

Hatch, Megan E., and Jinhee Yun. 2021. "Losing Your Home Is Bad for Your Health: Short- and Medium-Term Health Effects of Eviction on Young Adults," *Housing Policy Debate* 31 (3–5): 469–89. https://doi.org/10.1080/10511482.2020.1812690.

Hepburn, Peter, Renee Louis, and Matthew Desmond. 2020. "Racial and Gender Disparities among Evicted Americans," *Sociological Science* 7: 649–62. https://doi.org/10.15195/v7.a27.

Hepburn, Peter, Devin Q. Rutan, and Matthew Desmond. 2022. "Beyond Urban Displacement: Suburban Poverty and Eviction," *Urban Affairs Review*, March, 107808742210856. https://doi.org/10.1177/10780874221085676.

"ImageMagick." 2023. C. https://imagemagick.org/index.php.

Khanna, Kabir, Brandon Bertelsen, Santiago Olivella, Evan Rosenman, and Kosuke Imai. 2023. "Wru: Who Are You? Bayesian Prediction of Racial Category Using Surname and Geolocation." R. https://github.com/kosukeimai/wru.

Legal Services Corporation. 2023. “Civil Court Data Initiative.” Data. Eviction Tracker. September. https://civilcourtdata.lsc.gov.

Mullen, Lincoln, Cameron Blevins, and Ben Schmidt. 2023. “Gender: Predict Gender from Names Using Historical Data.” R. https://github.com/lmullen/gender.

National Academies of Sciences, Engineering, and Medicine. 2023. *Addressing the Long-Term Effects of the COVID-19 Pandemic on Children and Families*. Washington, DC: National Academies Press. https://doi.org/10.17226/26809.

Pan, Yuliya, Sabiha Zainulbhai, and Tim Robustelli. 2021. “Why Is Eviction Data so Bad?” Washington, DC: New America. https://d1y8sb8igg2f8e.cloudfront.net/documents/Why_is_Eviction_Data_so_Bad.pdf.

Parker, Brenda, and Janet Lynn Smith. 2021. “Policy Spotlight: Women’s Housing Precarity During and Beyond COVID-19.” SSRN Scholarly Paper 3896504. Rochester, NY: Social Science Research Network. https://doi.org/10.2139/ssrn.3896504.

Raymond, Elora, Sarah Stein, Victor P. Haley, Erik Woodworth, Gordon Zhang, R. Siva, and Subhro Guhathakurta. 2020. “Metro Atlanta Evictions Data Collective Database: Version 1.0.” School of City and Regional Planning: Georgia Institute of Technology. https://metroatlhousing.org/atlanta-region-eviction-tracker/.

Senate Committee on Housing Stability & Affordability. 2019. “Final Bill Report ESSB 5600.” Senate Bill Report C 356 L 19. Olympia, WA: Washington State Senate and House or Representatives. https://app.leg.wa.gov/billsummary?BillNumber=5600&Initiative=false&Year=2019.

Shinn, Marybeth, Andrew L. Greer, Jay Bainbridge, Jonathan Kwon, and Sara Zuiderveen. 2013. “Efficient Targeting of Homelessness Prevention Services for Families,” *American Journal of Public Health* 103 (S2): S324–30. https://doi.org/10.2105/AJPH.2013.301468.

“spaCy.” 2023. Python. https://spacy.io.

Tesseract-OCR. 2023. “Tesseract.” C++. Tesseract-OCR. https://github.com/tesseract-ocr/tesseract#license.

Thomas, Timothy. 2017. “Forced Out: Race, Market, and Neighborhood Dynamics of Evictions.” Thesis, Seattle, WA: University of Washington. https://digital.lib.washington.edu:443/researchworks/handle/1773/40705.

Thomas, Timothy, Ott Toomet, Ian Kennedy, and Alex Ramiller. 2020. “The State of Evictions: Results from the University of Washington Evictions Project.” Seattle, WA: University of Washington. https://evictionresearch.net/washington/.

U.S. Census Bureau. 2022. “More Than 19 Million Renters Burdened by Housing Costs.” December 8. https://www.census.gov/newsroom/press-releases/2022/renters-burdened-by-housing-costs.html.

Xie, Fangzhou. 2022. “Rethnicity: An R Package for Predicting Ethnicity from Names.” *SoftwareX* 17 (January): 100965. https://doi.org/10.1016/j.softx.2021.100965.

Eviction Practices in Subsidized Housing: Evidence From New York State

Ingrid Gould Ellen
Elizabeth Lochhead
Katherine O'Regan
NYU Wagner and Furman Center

Abstract

In response to growing evidence of the negative consequences of evictions, policymakers at all levels of government have directed a new effort toward eviction prevention in recent years. However, less attention has focused on evictions in subsidized housing, despite the low incomes of subsidized tenants and the arguably greater ability of government agencies to manage eviction practices in a stock they subsidize.

The authors aim to address this gap by leveraging multiple sources of administrative data to analyze patterns of nonpayment eviction filings and warrants in different types of place-based, subsidized housing in the state of New York between 2016 and 2022. Drawing on address-level eviction records from the New York State Office of Court Administration, unsubsidized property addresses from the New York City Department of Finance, subsidized property addresses from the National Housing Preservation Database, U.S. Department of Housing and Urban Development (HUD) administrative data, and the New York University Furman Center, the authors find that filing rates are consistently higher in public housing than in other types of subsidized housing. Filing rates in public housing are even higher than in unsubsidized multifamily properties, at least in New York City, where comparison data are available. However, the share of filings that result in a warrant is lowest in public housing, suggesting that public housing agencies (PHAs) may use eviction filing as a rent collection strategy. The analysis also uncovers substantial variation across cities and individual developments in New York City.

Introduction

The growing evidence on the cost of evictions has pushed policymakers at all levels of government in recent years to consider new efforts to prevent them (Collinson et al., 2023). At the city level, several cities, including New York, enacted new universal access to counsel laws for tenants facing eviction. At the state level, California and Oregon enacted statewide rent regulation (anti-rent gouging) laws in 2019. At the federal level, the U.S. Congress authorized the Emergency Rental Assistance program to assist individuals at risk of eviction across the country in response to the COVID-19 pandemic. However, surprisingly little scrutiny has been given to evictions in subsidized housing, even though government officials arguably have more policy levers to manage them and heightened interest in eviction practices within a stock they subsidize.

Several studies compare eviction patterns in subsidized and unsubsidized housing and find mixed results (Gromis, Hendrickson, and Desmond, 2022; Harrison et al., 2021; Preston and Reina, 2021). Previous research also identifies considerable variation in eviction patterns across markets and public housing agencies (PHAs). Among subsidized housing developments, eviction patterns may differ across properties depending on ownership, tenant incomes, and subsidy structures. The following section explains that higher eviction rates would be expected in privately owned developments, as well as those that serve lower-income tenants, and charge flat rents.

This article builds on this earlier work to compare eviction patterns in different types of place-based subsidized housing in New York City and in other cities and jurisdictions across New York State from 2016 through 2022. Observing whether patterns vary across different market conditions can be completed with this geographically diverse sample. Unlike earlier work, this article considers eviction filing rates and eviction warrants, permitting the examination of eviction filing outcomes. Across New York State jurisdictions, the authors can observe the share of eviction filings that lead to issued eviction warrants. In New York City, they can further determine the share of filings that resulted in executed warrants. Finally, the more recent data permit an examination of eviction patterns after the lifting of the COVID-19 moratoria.

In brief, eviction filing rates were consistently higher in public housing than in other types of subsidized housing in the prepandemic period. This finding is true in both New York City and the softer markets of upstate cities, although filing rate differences between public housing and low-income housing tax credit (LIHTC) developments were far lower within New York City. In New York City, eviction filings can also be observed in unsubsidized buildings and show that average filing rates in public housing were nearly three times as high as those for unsubsidized, multifamily buildings. In upstate cities, average filing rates in public housing were nearly four times as high as the overall filing rate for subsidized and unsubsidized buildings. That said, filing rates in individual public housing properties within New York City varied substantially, suggesting managerial discretion given that tenant incomes and demographics are very similar across the city's public housing developments. Importantly, the share of eviction filings resulting in issued warrants of eviction was consistently lower in public housing than in other stocks. In New York City, where data on executed warrants are available, approximately 1 percent of filings in public housing resulted in executed warrants. These facts, together with the lower dollar amounts sought, suggest that many PHAs use eviction filings as a strategy to collect back rent and underscore the value of administrative and policy reforms that might help to reduce these eviction filing rates and address nonpayment.

Theory and Background

Growing research illustrates the costs that evictions impose on individual households. Most notably, Collinson et al. (2023) use the random assignment of housing court cases to judges of varying leniency to estimate the impact of eviction on individual tenants. They show that an eviction order increases the risk of homelessness and hospital visits, reduces earnings, and increases indebtedness in the long term.

Even eviction filings themselves can harm tenants, regardless of whether a warrant is issued. Tenants threatened with eviction face both time and out-of-pocket costs associated with responding to court filings (Leung, Hepburn, and Desmond, 2021). Perhaps more significantly, filings can limit households' future housing options because many landlords use eviction records in screening prospective tenants (Desmond, 2016; Rosen, Garboden, and Cossyleon, 2021). Research documents the use of eviction filings as a rent collection strategy among private-sector landlords (Garboden and Rosen, 2019; Leung, Hepburn, and Desmond, 2021), a strategy that can lead to serial filings—repeated evictions filed against the same tenant. Such serial filings may compound the cost of evictions as each additional filing shows up on a tenant's record (Leung, Hepburn, and Desmond, 2021).

The federal government administers a range of subsidy programs that support the creation and operation of affordable developments for low-income renters. Eviction rates in place-based subsidized housing can be expected to differ from rates in unsubsidized rental properties for several reasons. On the one hand, receiving a housing subsidy may protect tenants from falling behind on rent and facing eviction, and owners of subsidized housing may be more reluctant to file evictions because of their perceived mission as service providers. On the other hand, tenants in subsidized housing generally have lower incomes and smaller saving buffers to manage income and expense shocks.

Depending on ownership, tenant composition, and rent structure, differences across properties may also be expected in the subsidized stock. In terms of ownership, although local government authorities own public housing, private entities own most subsidized housing in the United States. Private owners might be more apt than public agencies to file evictions and evict nonpaying tenants to maximize profits. If so, the expectation is that eviction rates would be lower in public housing compared with privately owned subsidized housing such as LIHTC and project-based Section 8 or 202 developments. However, not all private developers are for-profit owners—some are nonprofit or other mission-driven owners who may be less likely to pursue evictions.

It is also worth noting that PHAs may have additional administrative processes related to terminating tenancy. Leung et al. (2023) find some PHAs only file nonpayment cases in housing court after having exhausted their internal processes, which might lower filing rates. To the extent that required internal processes (for example, grievance processes) raise costs for PHAs in evicting tenants, it could also result in lower filings for those properties. That said, PHAs face additional programmatic incentives and requirements from the U.S. Department of Housing and Urban Development (HUD) in handling rent arrears. On the one hand, accounts receivable is one component of HUD's PHA assessment system, which puts pressure on PHAs to collect rent (HUD, 2015). On the other hand, evicting tenants and getting vacant units reoccupied is costly in terms of repairs and lost operating subsidy. Further, unit vacancy is also a component of a PHA's assessment, creating the dual pressures of maximizing rents collected and minimizing vacancies, which could lead to high filing rates but relatively lower executed evictions (Leung et al., 2023).

In addition, although HUD permits PHAs to enter into rental payment plans with public housing residents owing rent (and even encouraged such plans during the pandemic), HUD does not allow PHAs to forgive any portion of owed rent (HUD, n.d.; 2021). Hence, public housing tenants who fall behind may face increasing cumulative rental debt if they face another financial shock, perhaps contributing to additional eviction filings and higher rates of "serial" filings, consistent with Leung et al. (2023).

Tenant incomes may also shape eviction rates across subsidy programs. Within the subsidized stock, public housing and Section 8 developments house lower-income tenants who may be more likely to fall behind on rent due to lower savings and greater vulnerability to job loss or other financial shocks. However, income-based rents in public and Section 8 project-based housing provide greater stability, countering lower incomes. Tenants in these developments typically pay 30 percent of their income toward rent, so the tenant portion of rent should increase and decrease with income, assuming timely recertifications. In contrast, LIHTC tenants pay flat rents if they do not have other rent subsidies. Thus, if their income decreases, their rent burden increases.[1] The expectation is that LIHTC tenants would, therefore, be more likely to fall behind on rent than public and Section 8 project-based housing residents, holding all else constant. The same is expected for New York City's housing program-subsidized developments that typically charge flat rents as well.

Several recent studies have examined eviction practices in these types of place-based subsidized housing. The findings are somewhat mixed. Preston and Reina (2021) compared filing rates of federally subsidized and unsubsidized properties in Philadelphia. They found that between 2006 and 2017, annual eviction filing rates were one-third lower on average in subsidized multifamily properties than in unsubsidized multifamily properties.[2] Even after controlling for property and neighborhood characteristics, they found that tenants in public housing and project-based rental assistance properties were less likely to receive an eviction filing than unsubsidized renters in similar buildings and neighborhoods. Filing rates in LIHTC developments were higher than other subsidized properties and were closer to the citywide filing rate.

Gromis, Hendrickson, and Desmond (2022) found evidence of high filing rates, specifically in public housing. Using data on 1,243 PHAs in 26 states spanning from 2006 to 2016, they reported that PHAs file about 5.8 percent of eviction filings nationally, although only 3.5 percent of renters live in public housing. However, within individual markets, average PHA eviction filing rates are not significantly different from those of private landlords. In addition, they found considerable variation in eviction filing rates across individual PHAs, even within the same state or county,

[1] Otherwise, unsubsidized LIHTC tenants are also more likely to have income from employment and, hence, may be more likely to experience income shocks.

[2] As Leung et al. (2023) note, Preston and Reina (2021) do not appear to use HUD administrative data to capture all public housing addresses within a development, and they may undercount filings in public housing as a result.

suggesting that PHA-specific policies and practices are potentially quite important, although differences in local housing court practices could also contribute.

Harrison et al. (2021) examined filing rates in subsidized properties in Atlanta for 2016. They found that, on average, eviction filing rates were lower in subsidized buildings than in market-rate buildings. However, once they controlled for whether buildings targeted older populations, they saw no significant differences in filing rates between subsidized and unsubsidized properties. The most notable difference is between buildings targeting seniors and those serving families. Overall, senior multifamily developments had significantly lower annual average filing rates than non-senior developments.

Focusing specifically on eviction filings in public housing in 28 states, Leung et al. (2023) found that public housing residents face the same eviction filing risk as renters not residing in public housing but living in the same sets of census tracts that contain public housing. They also documented large variations across PHAs in their filing rates, even within the same state operating under the same tenancy laws. Their in-depth qualitative work on two PHAs in Ohio revealed very different eviction practices, with one PHA regularly filing eviction cases for each household that had not paid rent by the end of the month (as a rent collection strategy) and the other PHA only resorting to eviction filings after exhausting its internal administrative process.

Several of these studies also identified differences by race, suggesting racial discrimination and structural racism as additional factors in explaining filing rates across subsidized properties. Preston and Reina (2021) found that filing rates in subsidized properties are higher in census tracts where a greater share of subsidized households is African-American. Gromis, Hendrickson, and Desmond (2022) also showed that PHAs with greater shares of African-American residents have higher eviction filing rates on average, even after controlling for socioeconomic factors. Finally, Harrison et al. (2021) and Leung et al. (2023) showed that eviction filing rates are higher in properties in neighborhoods (either census tracts or block groups) with greater African-American population shares.

This article builds on the existing research to compare eviction patterns in subsidized housing in jurisdictions across New York State. The authors use individual eviction records with street addresses, allowing for the inclusion of precise controls for development attributes. Unlike earlier work, this study not only contains eviction filing data but also observes follow-on actions to cases, including the issuance of eviction warrants. It also draws on multiple years of data rather than a single year and observes more recent patterns. Finally, it compares eviction patterns across housing types in two very distinct markets: New York City and the softer markets of upstate cities.

This study examines four distinct types of place-based subsidized housing across New York State. Because it is possible for properties to have overlapping subsidy types, properties are grouped into four mutually exclusive categories for ease of analysis: (1) public housing; (2) project-based Section 8 but not Section 202 or public housing; (3) LIHTC but not public housing, Section 8, or Section 202; and (4) other federally subsidized housing.[3]

[3] The other federally subsidized category includes any other subsidy type listed in the National Housing Preservation Database, except for properties that have only project-based vouchers, including Section 236, HUD-insured properties, Section 202, Section 515, Section 538, HOME, and Moderate Rehabilitation properties.

In New York City, the authors have access to somewhat more detail on subsidy types and can also observe market-rate housing. Thus, they are able to compare evictions in seven different mutually exclusive categories for New York City: (1) public housing; (2) Section 202; (3) project-based Section 8, but not Section 202 or public housing; (4) LIHTC[4] but not Section 8, Section 202, or public housing; (5) other federally subsidized housing; (6) city-subsidized housing; and (7) unsubsidized multifamily (six or more units) housing. The authors separately analyze evictions in subsidized housing types for New York City to account for differences in market conditions in the city compared with the other cities in the state, which are far smaller and have much softer housing markets. Consider that in 2022, the U.S. Census Bureau-reported rental vacancy rate was 3.5 percent for the New York City metropolitan area compared with 9.2, 7.1, and 5 percent for the Albany, Buffalo, and Rochester metropolitan areas, respectively.

Eviction Process in New York State

Eviction cases in New York State are divided into two categories: nonpayment cases filed due to nonpayment of rent and holdover cases filed due to any violations other than nonpayment. This article focuses on nonpayment cases, which represent most eviction cases. During the study period, nonpayment cases represented 85 percent in New York City, 89 percent in Buffalo, 91 percent in Rochester, and 86 percent in Albany.

Before a landlord can begin a formal nonpayment eviction case, they must first send the tenant a notice of eviction indicating that rent is due. If the tenant does not pay the rent amount within 14 days, the landlord may then file a petition with the court to evict the tenant. Once the eviction case is filed, the tenant receives notice of the hearing. The tenant is then required to answer the petition orally or in writing, either before or at the hearing, depending on the jurisdiction.

Eviction cases in New York City are heard in the housing part of the New York City Civil Courts. Outside New York City, a lower-level civil court in the jurisdiction where the unit is located hears eviction cases. At the conclusion of a hearing, the court may enter a judgment in favor of the tenant, dismiss the petition, or enter a judgment of possession for the landlord. The court may also award a money judgment for the tenant's rent due. A judge may also enter a default judgment against a tenant if they fail to answer, fail to appear for their hearing, or fail to appear or respond at subsequent stages of the case. The eviction is stayed if the tenant pays their outstanding rent before warrant issuance, even if they pay after a judgment has been entered.[5]

Although eviction filings appear on tenants' records and may dissuade landlords from approving them in the future, filings do not constitute evictions, and most filings (60 percent) in New York State do not result in the issuance of a judgment. An eviction can only happen if the court enters a judgment of possession for the landlord, and then the landlord requests that the eviction be executed, and the court issues a warrant of eviction. A tenant may still ask the court to stay the eviction after issuing a warrant, and the court can approve such a stay. If no stay of the eviction is

[4] A version of the New York City analysis separated LIHTC 9 percent from LIHTC 4 percent properties. Filing rates and amounts sought were similar across these groups.

[5] Alternatively, the landlord and tenant may reach an agreement called a "Stipulation of Settlement" that outlines each party's rights and responsibilities for resolving the case. For example, the settlement may require the tenant to pay rent or cure lease violations and require the landlord to make repairs and list the dates and times when they will need to access the unit.

granted, a county sheriff, constable, or city marshal removes the people named in the judgment from the premises.

New York State does not permit nonjudicial termination of tenancy. Given this, any PHA administrative termination of tenancy process must operate *in addition* to housing court proceedings, rather than as a replacement. In New York State, public housing tenancies cannot be ended without an affirmative finding to do so through the housing court.

Various rental assistance and eviction prevention programs are available to tenants facing eviction in New York State. In New York City, short-term financial assistance is available through the Homebase homeless prevention program. In addition, tenants in New York City can access emergency loans through the Human Resources Administration (HRA). Although having a pending eviction case is not a requirement to qualify, tenants who receive an eviction notice can go to an HRA office within the court to potentially expedite the process.

Data

The analysis relies on a range of administrative data sources on both evictions and subsidized housing developments, as well as census data to capture neighborhood composition.

Office of Court Administration Eviction Data

Data on eviction cases come from the New York State Office of Court Administration (OCA). The data include address-level information on the case filing date, dollar amount sought, type of case (nonpayment versus holdover), and follow-on actions, including issuing an eviction warrant. In New York City, the data also include the date any warrant was executed. The data cover all eviction cases filed from 2016 to present in all city courts in New York State, Nassau, and Suffolk Counties. The data do not capture cases filed in other county courts against tenants living in rural areas, towns, or villages. The geographies included in the OCA data cover approximately 83 percent of renters in New York State overall and 53 percent of renters living outside New York City. This analysis focuses on New York City and upstate cities in the OCA data (see exhibit A-1). It excludes Nassau, Suffolk, and Westchester Counties because they have far fewer subsidized properties.

Subsidized Properties—Upstate Cities

The authors use the National Housing Preservation Database (NHPD) to identify subsidized properties in Upstate New York cities that OCA data cover. NHPD provides information on all subsidized properties in the state and identifies any properties with overlapping subsidies. One weakness of NHPD is that it provides only one street address per property, even for subsidized properties containing multiple buildings. These addresses are supplemented with additional addresses for public housing and LIHTC properties,[6] using HUD administrative data to address this weakness.[7] Because additional building addresses are only available for LIHTC properties that

[6] Note that other published papers do not appear to have utilized additional LIHTC building addresses, potentially undercounting filings in LIHTC properties.

[7] These additional addresses are merged using the HUD property identification field included in NHPD.

appear in HUD LIHTC tenant data, this analysis limits LIHTC properties in Upstate New York to the 82 developments (out of 403) included in HUD LIHTC tenant data.[8] Although public housing and LIHTC are the subsidy types most likely to be associated with multibuilding properties, other subsidized properties may also have additional addresses not in NHPD, which could downwardly bias estimates of their eviction filing rates if filings associated with all property addresses are not captured. The subsidy start- and end-date fields are used to determine the years in which each subsidy type was attached to the property.

Subsidized Properties—New York City

To identify subsidized properties in New York City, the authors use the Subsidized Housing Database from CoreData.nyc that the Furman Center maintains. This dataset lists subsidized properties at the borough-block-lot (BBL) level, including any overlapping subsidies and subsidy start and end dates. In New York City, unsubsidized multifamily rental properties can also be identified using the New York City Department of Finance Property Tax System. The authors estimate the set of BBLs likely to be rental properties by filtering out nonresidential properties, vacant land, condominiums, and co-ops, then limit the results to multifamily buildings defined as having six or more units. Properties from this filtered list that do not appear in the Subsidized Housing Database are considered unsubsidized.

Neighborhood Characteristics—American Community Survey

Eviction rates might be higher in developments with more economically vulnerable residents or more economically disadvantaged neighborhoods. Unfortunately, consistent information about tenant characteristics for all buildings is not available. To help control for population demographics and neighborhood factors identified in other research, the authors draw on 2016-2020 American Community Survey 5-year estimates to obtain characteristics of the surrounding census tract, including tract poverty rate and the shares of residents that are African-American and Hispanic or Latinx.

Methods

Analyzing eviction patterns across subsidized properties requires careful matching of eviction filings to buildings as well as calculating a variety of relevant eviction measures. The following section outlines the methods used to match evictions to properties and to estimate regressions that compare filing and other eviction outcomes across developments.

Geocoding and Matching Eviction Cases to Subsidized Properties

To match eviction filings to subsidized properties, the authors first parse and standardize all street addresses in OCA eviction data, New York State subsidized property data, and New York City subsidized property data to ensure addresses are consistently formatted across datasets. The addresses in each dataset are then geocoded. For upstate New York cities, eviction filings

[8] Developments in the study sample are in census tracts with very similar poverty rates and racial compositions to census tracts where other LIHTC developments in upstate cities are located. The included developments are somewhat more likely to be 4 percent LIHTC developments, however.

and subsidized properties are geocoded using the geocoder available from the Census Bureau. Eviction filings are matched to subsidized addresses using the standard address string assigned in the geocoding process.[9] In New York City, eviction filings are geocoded using New York City's Geoclient REST API. Then, eviction filings are matched to subsidized properties by BBL.[10] The eviction data are aggregated to the property level to compile a panel dataset at the property-year level, covering 2016 through 2022.

Calculating Eviction Measures

The authors calculate several different annual eviction measures at the property-year level. First, the *eviction filing rate*—the number of nonpayment filings in a year divided by total units at the property, including subsidized and unsubsidized units—is calculated. Second, the *average amount sought*, the total dollar amount sought divided by the total number of filings, is also calculated.[11] Third, the *issued warrant rate*, the number of issued eviction warrants divided by the total units at the property, and the *issued warrant share*, the share of eviction filings that result in an issued warrant, are also calculated.[12] Because a case can still be resolved after a warrant has been issued, it is not a precise measure of the number of evictions carried out but rather an indicator of whether the case progresses to the stage of a warrant. In New York City, an additional measure, the *executed warrant share*, which is the share of filings that result in warrants ultimately executed, can be calculated. This measure also allows for a comparison of the share of filings that result in executions across subsidy types within New York City.

Comparisons Across Subsidy Types

To compare these measures across subsidized housing types, the authors begin with simple descriptive statistics for eviction measures, shown separately for the two geographies: Upstate cities and New York City. This study focuses on the period between 2016 and 2019 to determine averages before pandemic-era eviction protections and court backlogs, but also considers later years to assess patterns during and after the pandemic.

Ordinary least squares regressions are then estimated to better control for local factors, including the economic and demographic characteristics of the property's neighborhood. Specifically, the authors regress each of the eviction measures on a categorical variable indicating the property subsidy type, controlling for the year and tract characteristics (poverty rate and shares of residents that are African-American and Hispanic).

These regressions are run separately for the two geographies. The upstate model includes a control for jurisdiction. In New York City, the authors also include a control for the community district, the 59 sub-city areas represented by local community boards, to better control for neighborhood conditions. The marginal means of each eviction measure are plotted by subsidy type in each year.

[9] Ninety-four percent of eviction filings and 93 percent of subsidized properties are geocoded successfully.

[10] Ninety-six percent of eviction filings are geocoded successfully within New York City.

[11] Small average amounts suggest potentially avoidable evictions (Badger, 2019).

[12] Eviction warrants may be issued weeks or months after the initial filing. For this analysis, an eviction filing that has a reported warrant issue date at any point after the filing date and before the summer of 2023 is considered to have an issued warrant.

The marginal mean represents the mean of the predicted value for each subsidy type in each year, holding the other factors constant. This value allows for the comparison of levels and trends in eviction filings, warrant shares, and the amount sought for properties with different subsidy types. The regression analysis includes all years from 2016 to 2022 to analyze both prepandemic and more recent trends.

Results

The following sections summarize key findings, starting with upstate cities. The tables present raw, uncontrolled means for the 4-year period extending from 2016 to 2019, and the event study exhibits show annual regression-controlled marginal means extending to 2022.

Upstate Cities

Exhibit 1 presents mean annual filings and mean share of filings that resulted in warrants for each subsidy type for upstate cities for filings occurring from 2016 through 2019. On average, during this 4-year period, public housing properties had by far the highest eviction filing rates of the subsidy types, with 40.5 filings per 100 units. Filing rates per unit were more than *10* times higher in public housing than in any other affordable portfolio. Public housing also has the highest issued warrant rate per 100 units, although the share of public housing filings resulting in warrants is the lowest among subsidized housing types.

Although eviction rates for individual unsubsidized properties in upstate cities are not available, these filing rates in subsidized housing can be compared with the overall filing rates in the same cities. On average, upstate landlords filed 10.5 evictions per 100 renter-occupied units each year, far fewer than PHAs but more than private owners of subsidized housing. In upstate cities, 44 percent of eviction filings resulted in warrants each year on average, identical to the share for LIHTC developments and significantly higher than the share for public housing.

Exhibit 1

Eviction Filings and Issued Warrants by Subsidy Type—Upstate Cities, 2016–19

Program	Number of Properties (2019)	Mean Annual Number of Filings (2016–19)	Mean Annual Filings per 100 Units (2016–19 Average)	Mean Annual Number of Issued Warrants (2016–19)*	Mean Annual Share of Filings Resulting in Issued Warrants (2016–19 Average)*	Mean Annual Issued Warrants per 100 Units (2016–19 Average)
Public Housing (PH)	106	7,652	40.5	2,249	29.4%	11.9
Section 8 but not Section 202 or PH	161	355	2.1	148	41.5%	0.8
LIHTC** but not Section 202, Section 8, or PH	73	192	3.5	85	44.1%	1.5
Other Subsidized	403	170	1.3	61	35.9%	0.5

**Represents the total number of eviction cases filed from 2016 to 2019 resulting in warrants by mid-2023 (the warrant date may be after 2019).*

***Only LIHTC properties included in U.S. Department of Housing and Urban Development tenant dataset.*

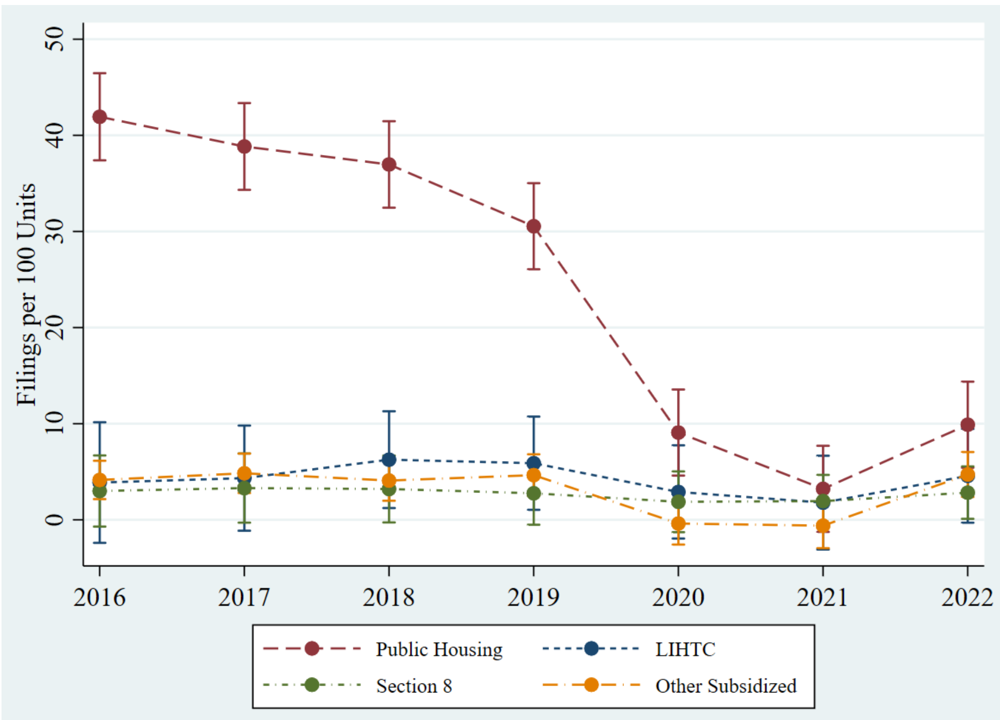
LIHTC = low-income housing tax credit.

Sources: National Housing Preservation Database; New York State Office of Court Administration

These same patterns hold once neighborhood characteristics, year, and city are controlled for in the regression analysis. Exhibit 2 plots the regression-controlled marginal means of filing rates by subsidy type for the years from 2016 through 2022, with 95-percent confidence intervals. Exhibit 2 plots the marginal means of the issued warrant rate per 100 units. Importantly, regression analyses also confirm that the share of filings that resulted in warrants is lowest for public housing. Exhibits 2 and 3 also show the sharp drop off of filings and warrants during the pandemic in 2020 and 2021. In 2022, filings and warrants in public housing were once again the highest among the subsidized housing types, but they have not returned to prepandemic levels, although the 2022 averages conceal considerable variation across cities.

Exhibit 2

Eviction Filings per 100 Units by Subsidy Type—Upstate Cities

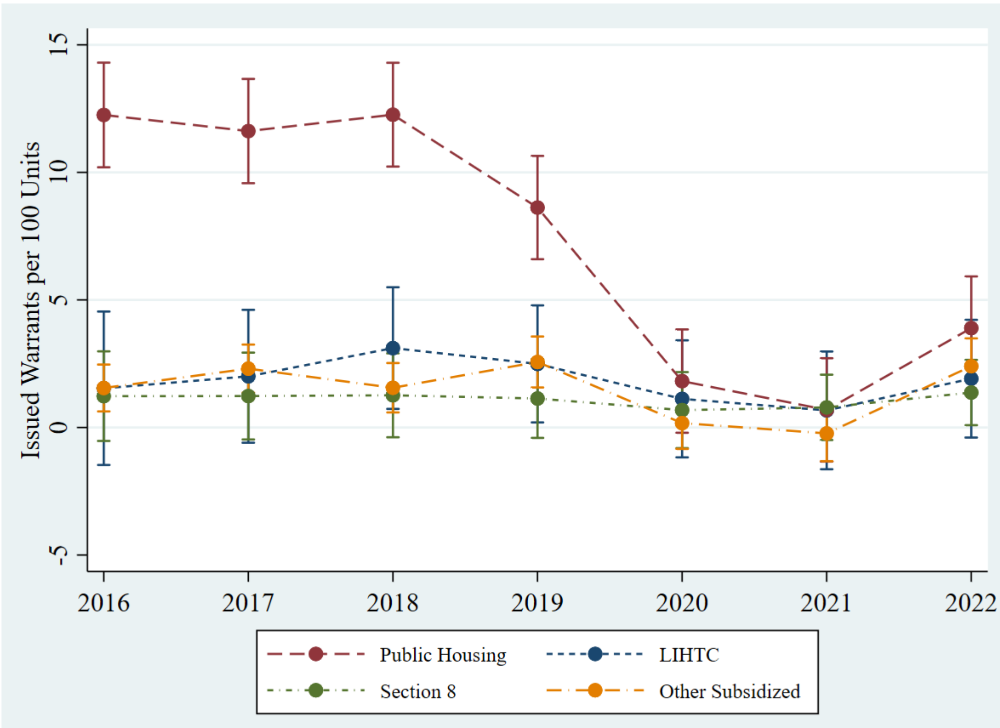


LIHTC = low-income housing tax credit.

Sources: National Housing Preservation Database; New York State Office of Court Administration; American Community Survey

Exhibit 3

Issued Warrants per 100 Units by Subsidy Type—Upstate Cities

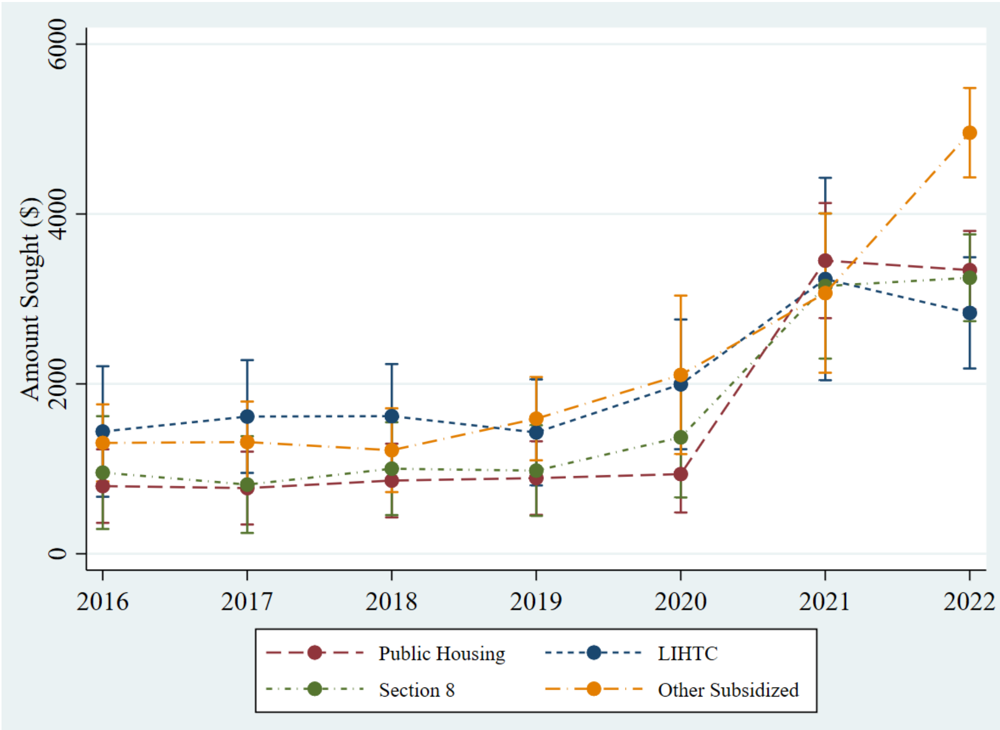


LIHTC = low-income housing tax credit.
Sources: National Housing Preservation Database; New York State Office of Court Administration; American Community Survey

Exhibit 4 displays the regression-controlled average dollar amount sought in nonpayment cases across subsidy types. Evictions filed at public housing properties had the lowest amount sought on average, perhaps due to lower rents. LIHTC and other subsidized properties tended to have relatively greater amounts sought, controlling for tract characteristics, year, and city. The amount sought in eviction cases increased in 2021, especially in public housing properties, which rose from approximately $1,000 in 2020 to nearly $3,500 in 2021. By contrast, the average rent public housing households paid in New York State barely increased during the same period, from $518 in 2020 to $523. The ratio of the average arrears sought in public housing cases relative to the average rent paid increased from slightly under 2.0 in 2020 to more than 6.5 in 2021.

Exhibit 4

Average Amount Sought in Nonpayment Cases by Subsidy Type—Upstate Cities



LIHTC = low-income housing tax credit.

Sources: National Housing Preservation Database; New York State Office of Court Administration; American Community Survey

New York City

Exhibit 5 presents mean annual filings and mean share of filings that resulted in issued and executed warrants for each subsidy type for properties in New York City from 2016 through 2019. Eviction activity in unsubsidized multifamily rental properties and other types of subsidized properties, including city-subsidized buildings and Section 202 properties, can be observed using additional city-specific data. Section 202 properties are grouped in the other subsidized category in the upstate cities analyses due to the small sample size.

Exhibit 5

Eviction Filings, Issued Warrants, and Executed Warrants by Subsidy Type—New York City, 2016–19

Program	Number of Properties (2019)	Mean Annual Number of Filings (2016–19)	Mean Annual Filings per 100 Units (2016–19 Average)	Mean Annual Number of Issued Warrants (2016–19)*	Mean Annual Share of Filings Resulting in Issued Warrants (2016–19)	Mean Annual Issued Warrants per 100 Units (2016–19 Average)	Mean Annual Number of Executed Warrants (2016–19)**	Mean Annual Share of Filings Resulting in Executed Warrants (2016–19 Average)	Mean Annual Executed Warrants per 100 Units (2016–19 Average)
Public Housing (PH)	505	31,424	19.2	11,611	36.9%	7.1	290	0.92%	0.2
Section 202 but not PH	100	156	2.3	66	42.5%	1.0	11	7.2%	0.2
Section 8 but not Section 202 or PH	386	2,530	12.5	1,194	47.2%	5.9	65	2.6%	0.3
LIHTC but not Section 8, Section 202, or PH	2,563	23,851	14.8	11,300	47.4%	7.0	960	4.0%	0.6
Other Federal Subsidized	317	1,537	5.3	736	47.9%	2.5	80	5.2%	0.3
City Subsidized	9,134	23,542	6.4	11,094	47.1%	3.0	1,211	5.1%	0.3
Unsubsidized (6 or more unit residential buildings)	44,333	74,713	6.5	35,079	46.9%	3.0	4,546	6.1%	0.4

**Represents the total number of eviction cases filed from 2016 to 2019 resulting in issued warrants by mid-2023 (the warrant date may be after 2019).*

***Represents the total number of eviction cases filed from 2016 to 2019 resulting in executed warrants by mid-2023 (the execution date may be after 2019).*

LIHTC = low-income housing tax credit.

Sources: CoreData.nyc; New York City Department of Finance Property Tax System; New York State Office of Court Administration

As in the case of other cities across the state, public housing developments in New York City had the highest eviction filing rates, although the difference between filing rates in public housing and the other subsidized types (particularly Section 8 and LIHTC) was smaller. Consistent with patterns elsewhere in the state, very few eviction filings in public housing resulted in actual evictions. Public housing developments had the lowest share of filings that resulted in issued warrants (36.9 percent). Because of their high filing rates, public housing developments still saw the highest issued warrant rate per unit (7.1 percent). But, the *executed* warrant rate per 100 units was lower in public housing than all but section 202 housing because only a very small share (less than 1 percent) of public housing filings resulted in an executed eviction.[13] Public housing developments had the lowest annual executed warrant rate per unit among housing types (0.2 executed warrants per 100 units, or one executed warrant each year per 500 units) and LIHTC developments had the highest (0.6 executed warrants each year per 100 units). Therefore, although New York City public housing developments generally saw more filings per unit than other housing types, they also saw the lowest rates of actual executed evictions per unit.

One potential contributor to the low warrant execution rates in New York City public housing is that the New York City Housing Authority (NYCHA) may file cases of chronic nonpayment as holdover evictions rather than nonpayment cases, which are this study's focus.[14] The exclusion of these chronic nonpayment cases from the data could bias the executed warrant rate downward. Unfortunately, this subcategory of holdover cases cannot be identified in the court filing data. However, only 3 percent of public housing eviction cases in New York City are holdover cases, and even if *all* holdover cases are included in the analysis, the executed warrant share rises only slightly, from 0.92 percent to 1.4 percent of all New York City public housing court filings.

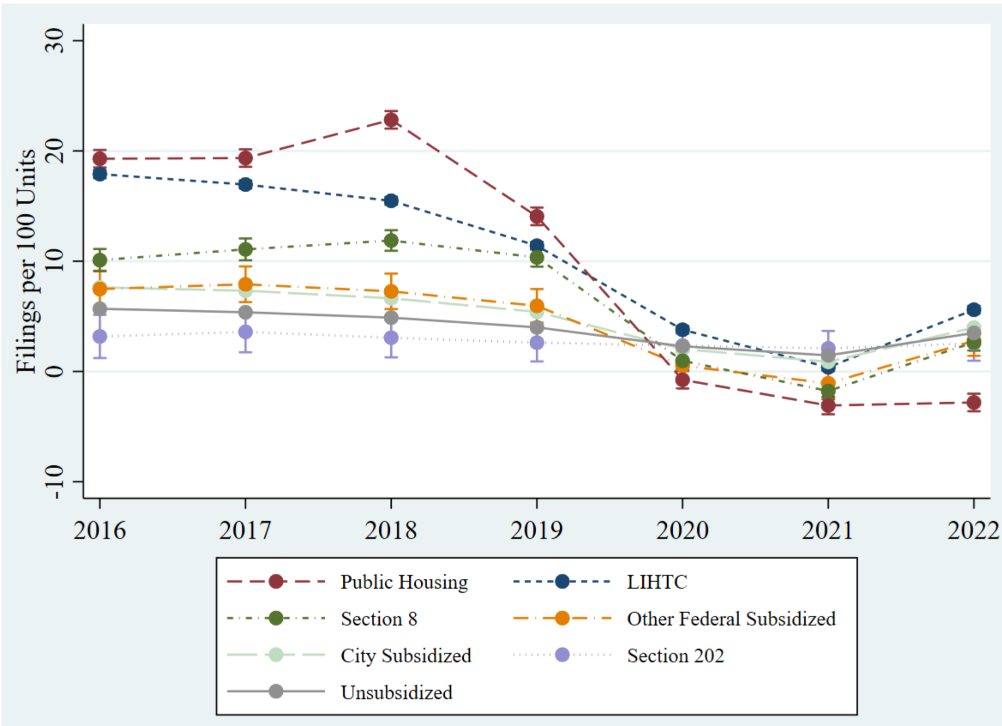
Notably, all subsidized housing portfolios except Section 202, which serves older adults who are typically on fixed incomes and do not see the same volatility in earnings, had higher average filing rates than unsubsidized multifamily housing. Although it may be due to the lower incomes of subsidized tenants, controls for neighborhood poverty rates and demographic composition should help to control for some of those tenant differences. As discussed further in the following section, it is possible that public housing agencies and owners of subsidized housing may be more likely than private-market landlords to file evictions as a rent collection strategy because they think it is the quickest way to get tenants access to rental assistance.

[13] On average, only about 5 percent of filings across types and less than 1 percent of public housing filings ultimately resulted in an executed warrant (not shown in table).

[14] A consent decree requires NYCHA to file chronic nonpayment cases as holdover cases if it chooses to use its administrative process to initiate terminating a tenancy based on chronic nonpayment, defined as failure to pay rent four times in a 12-month period. *Escalera v. NYCHA*, 425 F.2d 853 (2d Cir. 1970); *Tyson v. NYCHA*, 73 Civ. 859 (SDNY. 1976); *Randolph v. NYCHA*, 74 Civ. 1856 (SDNY 1976).

Exhibit 6

Eviction Filings per 100 Units by Subsidy Type—New York City



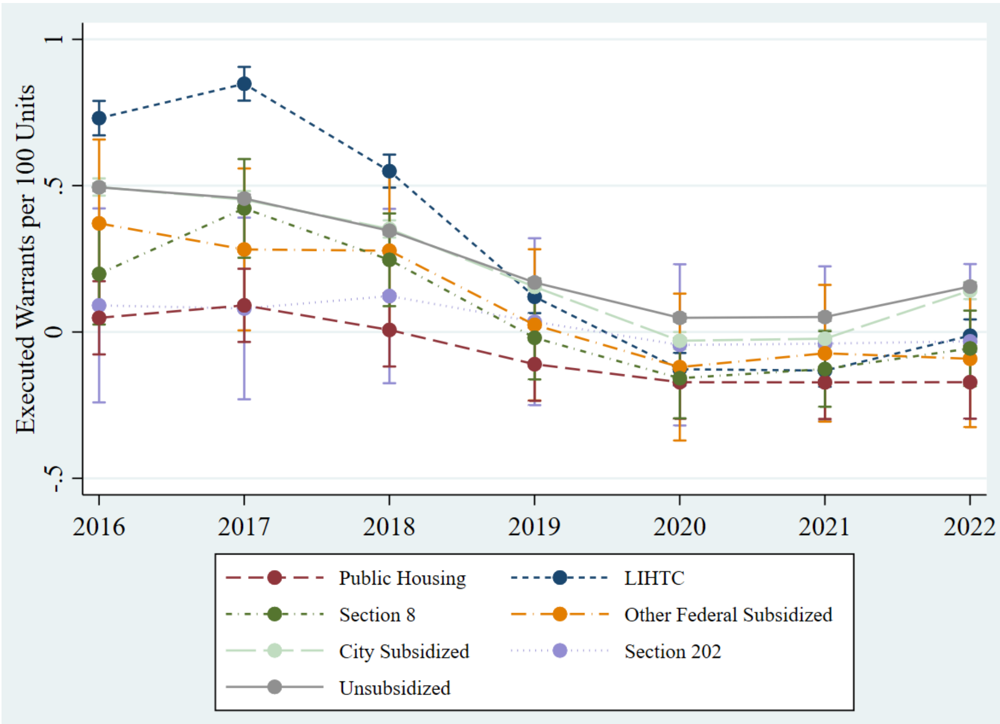
LIHTC = low-income housing tax credit.
Sources: CoreData.nyc; New York City Department of Finance Property Tax System; New York State Office of Court Administration; American Community Survey

As for trends over time, filing rates decreased across nearly all property types in 2019, even before the pandemic, with public housing properties seeing particularly sharp declines. All but Section 202 developments saw a sharper drop at the start of the pandemic in 2020. Filing rates in public housing have remained low since the pandemic, even as the other subsidized and unsubsidized types have started increasing again. This difference is likely due to a NYCHA decision to prioritize cases that involve more than 2 years of arrears and to centralize eviction decisions in the post-pandemic era to ensure more consistency across developments and encourage nonjudicial resolutions to nonpayment issues. In February 2022, NYCHA discontinued 90 percent of the 34,000 nonpayment cases pending in housing court (NYCHA, 2022).

Exhibit 7 plots the marginal means for the executed warrant rate (executed warrants per 100 units). Between 2016 and 2019, the executed rate was highest in LIHTC properties and lowest in public housing. All subsidized types other than LIHTC had executed rates below prepandemic unsubsidized multifamily properties. Since 2020, executed rates for all subsidy types have remained below the unsubsidized average, although not by much.

Exhibit 7

Executed Warrants per 100 Units by Subsidy Type—New York City



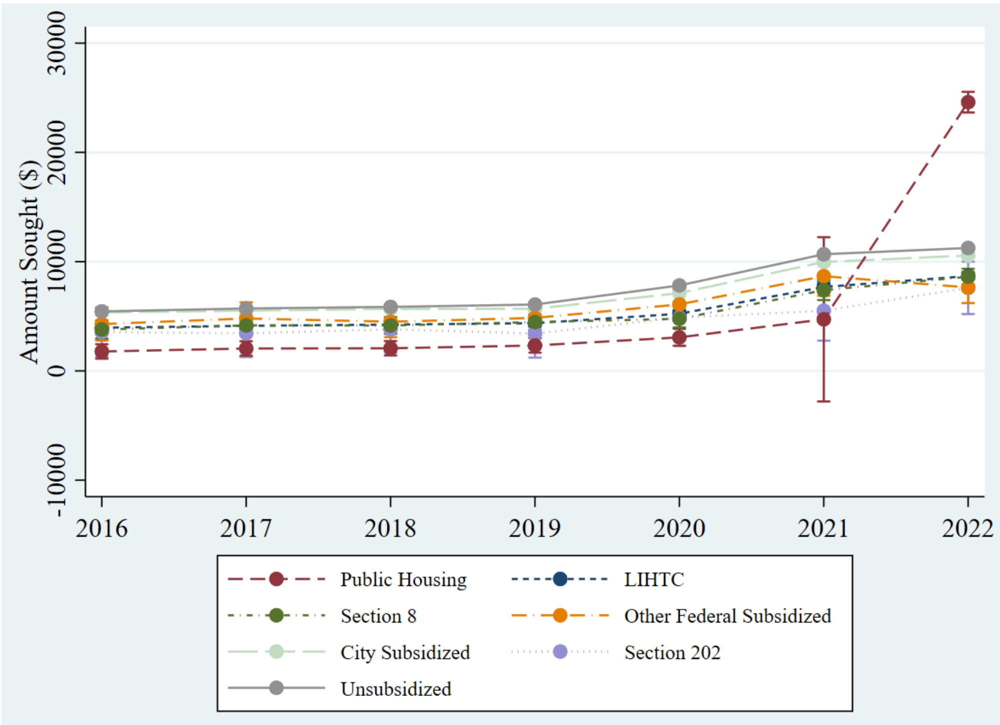
LIHTC = low-income housing tax credit.

Sources: CoreData.nyc; New York City Department of Finance Property Tax System; New York State Office of Court Administration; American Community Survey

Exhibit 8 shows the average dollar amount sought in nonpayment cases across subsidy types. In the prepandemic period, evictions filed at public housing properties were for the lowest amounts on average, similar to the results from the upstate cities. Eviction filings at unsubsidized (likely higher rent) properties tend to be for the highest dollar amount, as expected. NYCHA's shift to focus on higher arrears cases starting in 2022 also explains the sharp jump in the average amount sought for public housing.

Exhibit 8

Average Amount Sought in Nonpayment Cases by Subsidy Type—New York City



LIHTC = low-income housing tax credit.

Sources: CoreData.nyc; New York City Department of Finance Property Tax System; New York State Office of Court Administration; American Community Survey

Variation in Public Housing Eviction Patterns Across Upstate Cities and in New York City

The averages presented previously conceal quite a bit of variation in eviction actions in public housing, both across upstate cities and across developments in New York City. Exhibit 9 shows percentile cutoffs for these actions for public housing developments from 2016 to 2019.[15]

[15] Includes only cities with more than 20 total filings between 2016 and 2019.

Exhibit 9

Distributions of Eviction Measures in Public Housing by City and BBL in New York City (2016–19)

Geography	Measure	Minimum	10th Percentile	25th Percentile	50th Percentile	75th Percentile	90th Percentile	Maximum
Upstate Cities (jurisdictions with > 20 filings 2016–19, or 28 of 52 cities)	City-Level Mean Annual Filings per 100 Units—Public Housing (2016–19)	1.0	2.1	4.6	9.8	69.1	87.6	259.4
	City-Level Mean Annual Share of Filings Resulting in Warrants—Public Housing (2016–19)	0%	2.0%	22.0%	35.6%	50.0%	61.0%	75.7%
New York City	BBL-Level Mean Annual Filings per 100 Units—Public Housing (2016–19)	0	6.0	14.2	21.0	27.3	35.0	175.0
	BBL-Level Mean Annual Share of Filings Resulting in Executed Warrants—Public Housing (2016–19)	0%	0%	0%	0.3%	1.2%	2.3%	14.3%

BBL = borough-block-lot.
Sources: National Housing Preservation Database; CoreData.nyc; New York City Department of Finance Property Tax System; New York State Office of Court Administration

For upstate cities, one-half saw fewer than 10 annual filings per 100 units in public housing on average. At the other end of the spectrum, 10 percent of cities (two cities total) saw average filing rates of more than 100 filings per 100 units annually, indicating multiple (serial) filings issued to the same units in 1 year. Differences in rental arrears do not appear to drive these differences—filing rates are actually higher in PHAs where the amounts sought are lower. The differences in filing rates more likely reflect differences in approaches to managing arrears, consistent with Leung et al. (2023).

The share of filings receiving warrants was somewhat more evenly distributed across upstate cities, but it still ranged widely from 0 to 75.7 percent. Recall that an issued warrant does not necessarily mean that the eviction was ultimately executed. In New York City, only 10 percent of warrants issued in nonpayment cases were executed during the period of this study.[16]

Within New York City, variation is observed in eviction patterns across individual public housing properties, at least prepandemic, suggesting discretion on the part of individual property managers in deciding how quickly to file eviction notices when tenants owe back rent. The median number of filings per 100 units for New York City public housing properties between 2016 and 2019 was

[16] Whether warrants are executed in upstate cities is unobservable.

21.0, and most were clustered around this level; however, 10 percent of properties had filing rates of 35.0 or higher, and 10 percent of properties had rates of under 6.0 filings per 100 units.

The prepandemic median share of filings resulting in executed warrants in New York City public housing properties was only 0.3 percent. Ninety percent of public housing properties had executed warrant shares below 3 percent, although the executed warrant share reached 14 percent in one property, again suggesting managerial discretion. In 2022, execution rates across all properties are very low.

Serial Filings

The combination of high filing rates, low warrant rates, and even lower warrant execution rates in public housing suggest that PHAs use eviction filings more often as a rent collection strategy rather than a way to actually remove tenants. Indeed, some PHAs issue more eviction notices in a given year than total public housing units, with multiple filings against the same unit in a single year.

Exhibit 10 reports "serial" filings in New York State, defined as the number of cases filed against a unit that had already received a filing in that year. The table shows the share of eviction filings that are serial filings by subsidy type from 2016 to 2019.[17] Across both geographies, the highest rates of serial filing are in public housing. Serial filing rates in public housing are particularly high in upstate cities (exhibit 10).

Exhibit 10

Serial Filings (2016–19) (1 of 2)

Geography	Program	Mean Annual Number of Filings (2016–19)	Mean Annual Number of Filings With Nonmissing Apartment Number* (2016–19)	Mean Annual Number of Serial Filings (2016–19)	Mean Annual Share of Filings With Nonmissing Apartment Number That Are Serial Filings** (2016–19 Average)
Upstate Cities	Public Housing	7,652	4,969	3,058	61.6%
	Section 8 but not Section 202 or public housing	355	351	74	21.3%
	LIHTC** but not Section 202, Section 8, or public housing	192	179	29	16.6%
	Other Subsidized	170	155	27	17.6%

[17] Note that these serial filing rates can only be calculated for filings with nonmissing apartment numbers.

Exhibit 10

Serial Filings (2016–19) (2 of 2)

Geography	Program	Mean Annual Number of Filings (2016–19)	Mean Annual Number of Filings With Nonmissing Apartment Number* (2016–19)	Mean Annual Number of Serial Filings (2016–19)	Mean Annual Share of Filings With Nonmissing Apartment Number That Are Serial Filings** (2016–19 Average)
New York City	Public Housing	31,424	31,396	7,483	23.8%
	Section 202 but not public housing	156	156	10	6.6%
	Section 8 but not Section 202 or public housing	2,530	2,524	385	15.2%
	LIHTC but not Section 8, Section 202, or public housing	23,851	23,788	4,547	19.1%
	Other Federal Subsidized	1,537	1,533	229	15.0%
	City Subsidized	23,542	23,403	4,295	18.4%
	Unsubsidized (6 or more unit residential buildings)	74,713	74,567	13,909	18.7%

*Apartment number is nonmissing if (1) apartment number is listed; or (2) apartment number is listed but subsidized development is only one unit; or (3) apartment number is not listed, but the total number of unique addresses at a development is equal to the total units at the development (the development includes multiple single-family buildings, each with its own address).

**A filing is considered serial if it is a subsequent filing in the same year against a unit that already had a filing in that year. The initial filing at that unit in that year is not counted as a serial filing.

LIHTC = low-income housing tax credit.

Sources: National Housing Preservation Database; CoreData.nyc; New York City Department of Finance Property Tax System; New York State Office of Court Administration

Discussion and Policy Implications

This article hopefully shows the value of using and linking different sources of administrative data to analyze eviction patterns, including the outcomes of eviction filings. The analysis shows that eviction filing rates in subsidized housing in New York State are higher than many might expect. In New York City, eviction filing rates are higher in subsidized housing than in unsubsidized multifamily buildings. Eviction filing rates vary across subsidy types, with public housing having the highest filing rates across markets, even after controlling for neighborhood characteristics. However, public housing also has the lowest share of filings that result in eviction warrants. In New York City, less than 1 percent of filings in public housing result in executed warrants. Private owners of subsidized housing are less likely to file eviction notices than PHAs and are also somewhat more likely to follow them with court-issued eviction warrants.

High filing rates, low warrant rates, and even lower warrant execution rates in public housing suggest that PHAs use eviction filings more often as a rent collection strategy than private owners of other subsidized housing. Conversations with local housing agency officials and exhibit 10 provide some support for this hypothesis. However, this strategy is costly for both PHAs and tenants. In New York City, filing rates in project-based Section 8 and LIHTC properties are high, as well as in public housing. These high rates may result from landlords believing that filing an eviction is the best way for tenants to access emergency rental assistance from the city. Although an active eviction filing is not a formal requirement to access the program, many landlords believe that filing can facilitate faster approval (Chen, 2023).

It is worth noting that cities vary considerably in public housing filing and warrant rates, suggesting that local housing court regimes may shape outcomes or that individual PHAs within the same state may adopt different policies to address back rent. Filing rates even varied across individual public housing properties in New York City during the prepandemic period, reflecting the discretion traditionally afforded to individual property managers. Notably, since the 2022 NYCHA decision to begin prioritizing only cases involving several years of arrears, filing rates in New York City public housing properties have remained much lower than prepandemic averages and below rates of other subsidized and unsubsidized properties, with less variation across properties. In upstate cities, filing rates in public housing have begun to rise again, although the cities vary, perhaps due to differences in local policies.

This analysis shows the value of using administrative data to shed light on local eviction patterns. It also offers several policy implications. To start, HUD could take numerous steps to align better public housing management with affordable housing goals. The first is reconsidering its statutory interpretation that PHAs cannot forgive any portion of rent owed in public housing, removing a critical tool from the hands of the landlords housing our nation's most vulnerable populations.[18] The second is providing clear guidance strongly encouraging PHAs to exhaust other administrative routes to rent collection before filing evictions. It would likely lower unnecessary filings and lead to more consistency across developments and PHAs. Indeed, HUD could consider incorporating metrics into its PHA assessment system that incentivize using nonjudicial means of addressing rental arrears. Third, HUD should also require that PHAs provide data on their filings to permit HUD to monitor and study PHA eviction practices. Fourth, HUD could update its guidance on how landlords of any HUD-subsidized properties (and perhaps LIHTC properties) can—and cannot—use eviction records in screening prospective tenants. Akin to HUD's guidance distinguishing the use of arrest records from convictions and the importance of the specific context of a criminal conviction, HUD could provide guidance on a more nuanced use of eviction actions (HUD, 2016).

States could act, too. State housing finance agencies (HFAs) could add restrictions on using eviction filings in tenant selection, which could be incorporated into state Qualified Allocation Plans, as the state of Ohio has done for pandemic-era evictions. More generally, states could pass laws to expunge evictions automatically from public records after a certain number of years or when courts decide in favor of tenants. States could also reduce the fees and administrative barriers tenants face in requesting expungements.

[18] Private-sector landlords can forgive voucher households' past rent.

Finally, the findings suggest that PHAs across New York State—and across the country—should examine eviction patterns across their properties and compare them with those in the broader market, using data like those in this article. They should scrutinize their eviction practices and explore other administrative mechanisms to collect back rent. Housing agencies depend on rent revenues, but collecting rent through eviction filings is administratively costly and imposes significant burdens on tenants, even if they are not ultimately forced to leave their homes.

Appendix

Exhibit A-1

List of Upstate Cities in the Analysis

Albany	Middletown
Amsterdam	Newburgh
Auburn	Niagara Falls
Batavia	North Tonawanda
Beacon	Norwich
Binghamton	Ogdensburg
Buffalo	Olean
Canandaigua	Oneida
Cohoes	Oneonta
Corning	Oswego
Cortland	Plattsburgh
Dunkirk	Port Jervis
Elmira	Poughkeepsie
Geneva	Rensselaer
Glens Falls	Rochester
Gloversville	Rome
Hornell	Salamanca
Hudson	Saratoga Springs
Ithaca	Schenectady
Jamestown	Sherrill
Johnstown	Syracuse
Kingston	Tonawanda
Lackawanna	Troy
Little Falls	Utica
Lockport	Watertown
Mechanicville	Watervliet

Source: Authors

Acknowledgments

The authors thank Ryan Brenner, Jiaqi Dong, and Avery Lamb for their exceptional work assisting us with data and analysis of New York State laws and housing agency practices. We also thank two anonymous reviewers for their very helpful comments.

Authors

Ingrid Gould Ellen is the Paulette Goddard Professor of Urban Policy and Planning at NYU's Wagner School of Public Service and a faculty director at the NYU Furman Center. Elizabeth Lochhead is a doctoral student at NYU's Wagner School of Public Service and a doctoral research fellow at the NYU Furman Center. Katherine O'Regan is a Professor of Public Policy and Planning at NYU's Wagner School of Public Service and a faculty director at the NYU Furman Center.

References

Badger, Emily. 2019. "Many Renters Who Face Eviction Owe Less Than $600," *The New York Times*. https://www.nytimes.com/2019/12/12/upshot/eviction-prevention-solutions-government.html.

Chen, Stefanos. 2023. "Why a 'Startling Number' of Low-Income Tenants Face Eviction Cases," *The New York Times*. https://www.nytimes.com/2023/05/31/nyregion/nyc-evictions-supportive-housing.html.

Collinson, Robert, John E. Humphries, Nicholas Mader, Davin Reed, Daniel I. Tannenbaum, and Winnie van Dijk. 2023. Eviction and Poverty in American Cities. NBER Working Paper No. w30382. Cambridge, MA: National Bureau of Economic Research. https://doi.org/10.2139/ssrn.4196326.

Desmond, Matthew. 2016. *Evicted: Poverty and Profit in the American City*. New York: Crown.

Garboden, Philip M., and Eva Rosen. 2019. "Serial Filing: How Landlords Use the Threat of Eviction," *City & Community* 18 (2): 638–661. https://doi.org/10.1111/cico.12387.

Gromis, Ashley, James R. Hendrickson, and Matthew Desmond. 2022. "Eviction From Public Housing in the United States," *Cities* 127 (103749). https://www.sciencedirect.com/science/article/pii/S0264275122001883?via%3Dihub.

Harrison, Austin, Dan Immergluck, Jeff Ernsthausen, and Stephanie Earl. 2021. "Housing Stability, Evictions, and Subsidized Rental Properties: Evidence From Metro Atlanta, Georgia," *Housing Policy Debate* 31 (3–5): 411–424. https://doi.org/10.1080/10511482.2020.1798487.

Leung, Lillian, Peter Hepburn, and Matthew Desmond. 2021. "Serial Eviction Filing: Civil Courts, Property Management, and the Threat of Displacement," *Social Force* 100 (1): 316–344. https://doi.org/10.1093/sf/soaa089.

Leung, Lillian, Peter Hepburn, James Hendrickson, and Matthew Desmond. 2023. "No Safe Harbor: Eviction Filing in Public Housing," *Social Service Review* 97 (3): 456–497. https://doi.org/10.1086/725777.

New York City Housing Authority (NYCHA). 2022. "NYCHA Discontinues More Than 31,000 Non-Payment Cases, Reducing Caseload by 90 Percent." https://www.nyc.gov/site/nycha/about/press/pr-2022/pr-20220203.page.

Preston, Gregory, and Vincent J. Reina. 2021. "Sheltered From Eviction? A Framework for Understanding the Relationship Between Subsidized Housing Programs and Eviction," *Housing Policy Debate* 31 (3–5): 785–817. https://doi.org/10.1080/10511482.2021.1879202.

Rosen, Eva, Philip M.E. Garboden, and Jennifer E. Cossyleon. 2021. "Racial Discrimination in Housing: How Landlords Use Algorithms and Home Visits to Screen Tenants," *American Sociological Review* 86 (5): 787–822. https://doi.org/10.1177/00031224211029618.

U.S. Department of Housing and Urban Development (HUD). n.d. "PHA Guide to Preventing Homelessness After the Eviction Moratorium Expires." https://www.hud.gov/sites/dfiles/PIH/documents/Attachment1-PHABrochure.pdf.

———. 2021. "Questions on the U.S. Treasury's Emergency Rental Assistance (ERA) and Other Rental Assistance Programs." https://www.hud.gov/sites/dfiles/PIH/documents/ERAP_PIH_ERAP_FAQs.pdf.

———. 2016. "Office of General Counsel Guidance on Application of Fair Housing Act Standards to the Use of Criminal Records by Providers of Housing and Real Estate-Related Transactions." https://www.hud.gov/sites/documents/HUD_OGCGUIDAPPFHASTANDCR.PDF.

———. 2015. "Lead the Way: PHA Governance and Financial Management - A Training for Board Members." https://www.hudexchange.info/trainings/courses/lead-the-way-pha-governance-and-financial-management.

Commentary: Using Eviction Court Records to Inform Local Policy

Peter Hepburn
Rutgers University-Newark

Introduction

Landlord-tenant court records provide an enormously useful source of administrative data, allowing researchers, activists, and policymakers insight into the prevalence, causes, and consequences of housing instability among renters. The three articles collected here describe ways in which these data can be accessed and analyzed to inform local decisionmaking. This brief commentary reflects on the contributions that each makes and describes other ways such data can, and should, be harnessed to shape policymaking—housing-related and otherwise.

To begin, however, it is worth acknowledging and celebrating enormous strides that have been made both in recognizing eviction as a meaningful social and policy problem and accessing data about its prevalence—two processes imbricated deeply with one another. The Anti-Eviction Mapping Project began its work in San Francisco in 2013. The Eviction Lab launched its national map of eviction rates in 2018. During the COVID-19 pandemic, research and advocacy organizations nationwide began collecting eviction records from local courts, most notably the Legal Services Corporation, through their Civil Court Data Initiative.

Largely thanks to these efforts, we now know far more about eviction in America. In a typical year in the 2010s, 2.7 million renter households including 7.6 million people—2.9 million of them children, a disproportionate number of them African-American—faced an eviction case in this country (Graetz et al., 2023; Gromis et al., 2022). Most of these cases are brought to court due to nonpayment of rent, and few tenants have lawyers (Deluca and Rosen, 2022; Ellen et al., 2021). Two decades ago, Hartman and Robinson (2003), noting how little we knew about the prevalence of eviction, wrote that “the existence of numerical information in itself permits and encourages the media, policymakers, public officials, researchers, and the general public to pay attention to a problem that is now well beneath the surface” (489). Over the last decade, we have moved closer to seeing the problem, rendering it much harder to ignore.

Fighting Crime or Punishing the Poor?

In their article, "Analyzing the Effect of Crime-Free Housing Policies on Completed Evictions using Spatial-First Differences," Griswold and colleagues offer compelling evidence that implementing crime-free housing policies is associated with increased evictions. Using data from four jurisdictions in California, they show that census block groups with crime-free certified rental units have higher rates of executed evictions than otherwise equivalent spaces. Differences are both statistically and practically significant; they estimate an average 24.9-percent increase in evictions in such block groups and show similar effects across jurisdictions.

The article contributes to a growing body of evidence detailing the harms caused by third-party policing generally and crime-free housing policies in particular (Archer, 2019; Prochaska, 2023; Werth, 2013), including with respect to eviction (Desmond and Valdez, 2012; Kroeger and La Mattina, 2020). Such policies extend the reach of the carceral state, recruiting landlords to surveil and punish tenants construed as deviant (Cullen, 2022; Wacquant, 2009). Unlike nuisance ordinances applied city- or neighborhood-wide, crime-free housing policies depend on landlords opting in. The authors show that selection is nonrandom: certain landlords in certain neighborhoods—particularly low-income neighborhoods with more non-White residents and renters—are more likely to participate. The net effect is further destabilization of already marginalized communities.

The authors argue that their results undercut claims about the cost efficiencies of crime-free housing policies because evictions cost money to execute, and these cases entail meaningful downstream costs (e.g., more funding for homeless shelters). An even more direct case to be made against such policies is that landlords who file more eviction cases see significantly higher rates of assault, burglaries, robberies, and theft at their properties (Gomory and Desmond, 2023). By destabilizing tenancies, increasing mobility, and undermining collective efficacy, crime-free housing policies may not only increase evictions but also intensify violent crime within neighborhoods (Semenza et al., 2022). Although the moral, financial, and legal arguments against such policies are increasingly clear, highlighting their direct criminogenic potential is important.

Low-Hanging Fruit: Subsidized Housing Policy Reform

Are residents of subsidized housing protected from the threat of eviction? Judging by Ellen and colleagues' article, "Eviction Practices in Subsidized Housing: Evidence from New York State," the answer to that question depends on location, subsidy type, and how one measures eviction risk. Drawing on data from across New York State, the authors demonstrate strikingly high eviction filing rates from public housing, although they also show that a very small number of such filings result in issued eviction warrants. Filing rates from other subsidized developments—whether funded through Section 8, Low-Income Housing Tax Credit, or other programs—are lower than those observed in public housing but more likely to lead to an eviction warrant.

For at least two reasons, the question of subsidized housing's effect on eviction has attracted considerable recent attention (Gromis, Hendrickson, and Desmond, 2022; Harrison et al., 2021; Leung et al., 2023; Lundberg et al., 2021; Preston and Reina, 2021). First, it speaks to a basic premise: many people get evicted because they cannot pay rent. These programs may allow residents to manage expenses and stay housed stably by reducing housing cost burden. Ellen and colleagues' findings clearly indicate this not to be the case. The extraordinarily high eviction filing rates documented in public housing—where rents can rise and fall with resident income—demonstrate the limits of reducing housing cost burden alone. According to the U.S. Department of Housing and Urban Development (HUD) Picture of Subsidized Households, the average household living in public housing in New York State in 2016 (the start of the study period) had an annual income of $22,816, less than a quarter of local median household income (HUD, 2022). Even when paying only a few hundred dollars a month in rent, a family will struggle to make ends meet on that sum, especially in a place as expensive as New York City. We should consider how supplemental income programs can facilitate housing stability and might pair effectively with housing subsidies.

The second reason subsidy programs have attracted scholarly attention is because they represent a space where it may be possible to make broad-stroke changes affecting a relatively large population. Although state and local governments set most eviction regulations, housing subsidy programs leave open an avenue for meaningful federal reform. Ellen and colleagues' findings suggest several areas for improvement. First, HUD should move to reduce the number of eviction cases filed by public housing authorities. As is clear from their article and other recent research, public housing authorities routinely turn to serial eviction filing to facilitate rent collection, rarely seeing cases through to removal. Even if such a strategy were effective—Leung and colleagues (2023) find no evidence that it is—that does not mean it is the only or best option available. HUD should actively describe and promote alternatives and, as Ellen and colleagues argue, reconsider the possibility of rent forgiveness. Second, HUD should collect additional data on evictions from federally assisted housing, a possibility that the Department has already begun to explore (Office of Policy Development and Research, 2021a). Especially with improved transparency, such data could allow local policymakers to channel resources to programs and developers with a proven track record of improving residential stability.

As with other recent research on eviction and subsidized housing, Ellen and colleagues are unable to examine the effects of housing choice vouchers, by far the largest of the deep subsidy programs (Schwartz, 2021: 8). Voucher use is concentrated heavily in a set of neighborhoods that also see high eviction rates (Rosen, 2020). Future research should work to establish whether recipient landlords are more or less likely to bring eviction cases than nonrecipients.

New Means of Accessing Data

The two preceding articles relied on structured eviction data—data already in a machine-readable, spreadsheet-type format. In their contribution, "Toward a National Eviction Data Collection Strategy using Natural Language Processing," Thomas and colleagues describe a set of tools that allow for creating such structured data. Specifically, they detail a natural language processing approach where scanned images of court documents can be converted to text and then mined for relevant data. Using court records from four counties in Washington State, the authors demonstrate how such tools can be employed and validated.

A great deal of data exists about eviction cases that we are not systematically collecting. Thomas and colleagues frame their contribution as allowing us to study places where scanned documents may be available but structured data are not. However, the potential of natural language processing technologies extends beyond that: These tools offer the hope of extracting considerably more information that courts collect in various documents but do not systematically record in databases.

Still, natural language processing technologies have restrictions. As the authors note, such tools do not work with handwritten records, thus limiting their value. Even when records are readable, data processing is time-intensive and error-prone and requires significant validation. Those problems are by no means unique to natural language processing technologies (Porton, Gromis, and Desmond, 2021), and the possibility of rapid improvements in the development of local large language models may facilitate this approach in the not-so-distant future. However, it is hard to imagine local policymakers, few of whom possess significant technical capacity, taking advantage of such tools. Therefore, it remains imperative that we push for the creation and distribution of more and higher-quality structured datasets (New America, 2023; Office of Policy Development and Research, 2021b).

Further Applications: Proven and Potential Uses of Eviction Records

How else can eviction records be used to drive local insights? Existing research points to a number of avenues with immediate policy implications.

One of the simplest questions that a local policymaker should be able to answer using court data is, "Where are evictions happening?" Previous research highlights the extreme geographic concentration of eviction cases within a relatively small number of neighborhoods and even buildings (Teresa and Howell, 2021). In Tucson, Arizona, the same 295 buildings were consistently responsible for two-thirds of all evictions every year between 2004 and 2013 (Rutan and Desmond, 2021). Over the last year, two in every five eviction cases filed in Memphis, Tennessee, originated from just 100 buildings (Hepburn et al., 2023). When working to target rental assistance, legal aid, or re-housing services—or directing tax enforcement or building inspections—it pays to know exactly where people are being evicted.

Court data can also allow policymakers to understand better how the eviction process is being used by local landlords and, as such, to tailor responses. For example, in some areas—particularly in places where the court process is cheap, fast, and easy (Leung, Hepburn, and Desmond, 2021)—landlords file repeated, serial eviction cases against the same tenants at the same units (Garboden and Rosen, 2019; Immergluck et al., 2019). An eviction diversion or right-to-counsel program may be better suited to a jurisdiction where a large portion of cases are seen through to removal, while regulations to slow the eviction process or raise eviction filing fees might be more appropriate in areas with high serial eviction filing rates.

As Griswold and colleagues' article makes clear, these data do not pertain only to housing policy. Eviction has far-reaching consequences for renters' employment and financial well-being

(Collinson et al., 2023), their health and access to insurance (Hatch and Yun, 2021; Schwartz et al., 2022a), and their children (Graetz et al., 2023; Schwartz et al., 2022b). These data can inform decisions around health care, policing, social services, and schools.

However, it is also important to recognize the limits of these data. Court data will always leave hidden events that occur outside of the courts. An unknown number of renters face informal and illegal evictions each year, and attempts to measure these events using surveys have seen mixed success (Gromis and Desmond, 2021). Even within the courts, the exact outcome of cases is often difficult to ascertain (Nelson et al., 2021; Summers, 2023). Still, for local leaders interested in addressing housing instability and its repercussions, access to and careful analysis of these sorts of administrative court data can be invaluable.

Acknowledgments

Thank you to the editor for the invitation to provide this commentary and to Carl Gershenson for reviewing a draft. All opinions and errors are my own.

Author

Peter Hepburn is an assistant professor of sociology at Rutgers University-Newark and associate director of the Eviction Lab at Princeton University.

References

Archer, Deborah N. 2019. "The New Housing Segregation: The Jim Crow Effects of Crime-Free Housing Ordinance," *Michigan Law Review* 118 (2): 173–232.

Collinson, Robert, John Eric Humphries, Nicholas Mader, Davin Reed, Daniel I. Tannenbaum, and Winnie van Dijk. 2023. *Eviction and Poverty in American Cities*. 30382.

Cullen, Kimberly J. 2022. "State-Sponsored Surveillance and Punishment: How Municipal Crime-Free Ordinances Exacerbate the Carceral Continuum," *Boston University Public Interest Law Journal* 31 (1): 47–80.

Deluca, Stefanie, and Eva Rosen. 2022. "Housing Insecurity Among the Poor Today," *Annual Review of Sociology* 48 (20): 1–29.

Desmond, Matthew, and Nicol Valdez. 2012. "Unpolicing the Urban Poor: Consequences of Third-Party Policing for Inner-City Women," *American Sociological Review* 78 (1): 117–141.

Ellen, Ingrid Gould, Katherine O'Regan, Sophia House, and Ryan Brenner. 2021. "Do Lawyers Matter? Early Evidence on Eviction Patterns After the Rollout of Universal Access to Counsel in New York City," *Housing Policy Debate* 31 (3–5): 540–561.

Garboden, Philip M.E., and Eva Rosen. 2019. "Serial Filing: How Landlords Use the Threat of Eviction," *City & Community* 18 (2): 638–661.

Gomory, Henry, and Matthew Desmond. 2023. "Neighborhoods of Last Resort: How Landlord Strategies Concentrate Violent Crime," *Criminology* 61 (2): 270–294.

Graetz, Nick, Carl Gershenson, Peter Hepburn, Sonya R. Porter, Danielle H. Sandler, and Matthew Desmond. 2023. *A Comprehensive Demographic Profile of the US Evicted Population*.

Gromis, Ashley, and Matthew Desmond. 2021. "Estimating the Prevalence of Eviction in the United States: New Data from the 2017 American Housing Survey," *Cityscape* 23 (2): 279–289.

Gromis, Ashley, Ian Fellows, James R. Hendrickson, Lavar Edmonds, Lillian Leung, Adam Porton, and Matthew Desmond. 2022. "Estimating Eviction Prevalence Across the United States," *Proceedings of the National Academy of Sciences* 119 (21): 1–8.

Gromis, Ashley, James R. Hendrickson, and Matthew Desmond. 2022. "Eviction from Public Housing in the United States," *Cities* 127: 1–13.

Harrison, Austin, Dan Immergluck, Jeff Ernsthausen, and Stephanie Earl. 2021. "Housing Stability, Evictions, and Subsidized Rental Properties: Evidence From Metro Atlanta, Georgia," *Housing Policy Debate* 31 (3–5): 411–424.

Hartman, Chester, and David Robinson. 2003. "Evictions: The Hidden Housing Problem," *Housing Policy Debate* 14 (4): 461–501.

Hatch, Megan E., and Jinhee Yun. 2021. "Losing Your Home Is Bad for Your Health: Short- and Medium-Term Health Effects of Eviction on Young Adults," *Housing Policy Debate* 31 (3–5): 469–489.

Hepburn, Peter, Jacob Haas, Renee Louis, Adam Chapnik, Danny Grubbs-Donovan, Olivia Jin, Jasmine Rangel, and Matthew Desmond. 2023. "Eviction Tracking System: Version 2.0." https://evictionlab.org/eviction-tracking/.

Immergluck, Dan, Stephanie Earl, Jeff Ernsthausen, and Allison Powell. 2019. *Multifamily Evictions, Large Owners, and Serial Filings: Findings from Metropolitan Atlanta*.

Kroeger, Sarah, and Giulia La Mattina. 2020. "Do Nuisance Ordinances Increase Eviction Risk?" *AEA Papers and Proceedings* 110: 452–456.

Leung, Lillian, Peter Hepburn, and Matthew Desmond. 2021. "Serial Eviction Filing: Civil Courts, Property Management, and the Threat of Displacement," *Social Forces* 100 (1): 316–344.

Leung, Lillian, Peter Hepburn, James Hendrickson, and Matthew Desmond. 2023. "No Safe Harbor: Eviction Filing in Public Housing," *Social Service Review* 97 (3): 456–497.

Lundberg, Ian, Sarah L. Gold, Louis Donnelly, Jeanne Brooks-Gunn, and Sara S. McLanahan. 2021. "Government Assistance Protects Low-Income Families from Eviction," *Journal of Policy Analysis and Management* 40 (1): 107–127.

Nelson, Kyle, Philip Garboden, Brian J. McCabe, and Eva Rosen. 2021. “Evictions: The Comparative Analysis Problem,” *Housing Policy Debate* 31 (3–5): 696–716.

New America. 2023. “Court Eviction Data Standards: Recommended Eviction Data Elements and Definitions.” https://www.newamerica.org/future-land-housing/briefs/court-eviction-data-standards/.

Office of Policy Development and Research. 2021a. *Congressional Report on Tracking Data on Evictions from Federally Assisted Housing for House Report 116-106*.

———. 2021b. *Report to Congress on the Feasibility of Creating a National Evictions Database*.

Porton, Adam, Ashley Gromis, and Matthew Desmond. 2021. “Inaccuracies in Eviction Records: Implications for Renters and Researchers,” *Housing Policy Debate* 31 (3–5): 377–394.

Preston, Gregory, and Vincent J. Reina. 2021. “Sheltered From Eviction? A Framework for Understanding the Relationship Between Subsidized Housing Programs and Eviction,” *Housing Policy Debate* 31 (3–5): 785–817.

Prochaska, Jenna. 2023. “Breaking Free from ‘Crime-Free’: State-Level Responses to Harmful Housing Ordinances,” *Lewis & Clark Law Review* 27 (1): 259–326.

Rosen, Eva. 2020. *The Voucher Promise: “Section 8” and the Fate of an American Neighborhood*. Princeton, NJ: Princeton University Press.

Rutan, Devin, and Matthew Desmond. 2021. “The Concentrated Geography of Eviction,” *Annals of the American Academy of Political and Social Science*.

Schwartz, Alex F. 2021. *Housing Policy in the United States*. 4th ed. New York: Routledge.

Schwartz, Gabriel L., Justin M. Feldman, Scarlett S. Wang, and Sherry A. Glied. 2022a. “Eviction, Healthcare Utilization, and Disenrollment Among New York City Medicaid Patients,” *American Journal of Preventive Medicine* 62 (2): 157–164.

Schwartz, Gabriel L., Kathryn M. Leifheit, Jarvis T. Chen, Mariana C. Arcaya, and Lisa F. Berkman. 2022b. “Childhood Eviction and Cognitive Development: Developmental Timing-Specific Associations in an Urban Birth Cohort,” *Social Science and Medicine* 292: 114544.

Semenza, Daniel C., Richard Stansfield, Jessica M. Grosholz, and Nathan W. Link. 2022. “Eviction and Crime: A Neighborhood Analysis in Philadelphia,” *Crime & Delinquency* 68 (4): 707–732.

Summers, Nicole. 2023. “Civil Probation,” *Stanford Law Review* 75: 847–915.

Teresa, Benjamin F., and Kathryn L. Howell. 2021. “Eviction and Segmented Housing Markets in Richmond, Virginia,” *Housing Policy Debate* 31 (3–5): 627–646.

U.S. Department of Housing and Urban Development (HUD). 2022. “Picture of Subsidized Households.”

Wacquant, Loïc. 2009. *Punishing the Poor: The Neoliberal Government of Social Insecurity*. Durham, NC: Duke University Press.

Werth, Emily. 2013. *The Cost of Being "Crime Free:" Legal and Practical Consequences of Crime Free Rental Housing and Nuisance Property Ordinances*. Chicago, IL: Sargent Shriver National Center on Poverty Law.

Developing and Improving Datasets

Racial Disparities in Automated Valuation Models: New Evidence Using Property Condition and Machine Learning

Local Landscapes of Assisted Housing: Reconciling Layered and Imprecise Administrative Data for Research Purposes

Who Owns Our Homes? Methods to Group and Unmask Anonymous Corporate Owners

How Data Architects Are Crafting Equitable Housing Policy Research

Racial Disparities in Automated Valuation Models: New Evidence Using Property Condition and Machine Learning

Linna Zhu
Michael Neal
Urban Institute

Caitlin Young
Yale Law School

Abstract

Automated valuation models (AVMs), which exclude an appraiser's input in estimating a home's price, hold great promise for reducing costs and increasing the accuracy of home valuations. However, AVMs can manifest racial disparities, even when the algorithm remains agnostic to the neighborhood's majority race or the homebuyer's race. This study provides a quantifiable measure for auditing the performance of AVMs in majority-Black neighborhoods compared with their majority-White counterparts. The authors find that including data on property condition and employing more sophisticated machine learning techniques can help more accurately assess the percentage of the magnitude of AVM error and its underlying contributors. In addition, even with data improvement and machine learning, the authors still find evidence that AVMs yield larger valuation errors in majority-Black neighborhoods.

Introduction

The racial gap in homeownership rates is wide and persistent. Even when Black households do achieve homeownership, the value of their homes is typically less than that of White households. Although household, property, and community differences play key roles in explaining the home value gap, evidence also suggests that differences in appraised value can contribute to the racial disparity in home values.

One potential solution to appraisal bias is using automated valuation models (AVMs). An AVM is a computer-driven mathematical formula that uses property characteristics, local market information, and price trends to arrive at an estimated value for a property. With a surplus of data, an AVM can provide a standard approach and faster property value estimation.

Hypothetically, AVMs should help address instances of racial bias. By eliminating the appraiser's input, a potential source of racial inequity, the AVM should produce property value estimates agnostic to the community's racial makeup. In addition, because an AVM uses a standardized approach to estimate a property's value, the supposed elimination of racial differentiation should scale to communities more broadly.

The prospect for greater accuracy and standardization also indicates that AVMs can bring greater efficiency to the market. In addition, AVMs may quicken property valuation, because a task that may take a significant amount of time with an appraiser can be completed more quickly with an AVM. The AVM can also be centralized, eliminating the need for much of the appraisal profession.

The efficiency proposition AVMs pose is enhanced with the use of artificial intelligence (AI). AI, specifically machine learning (ML), can further speed up calculating a property's value. In addition, it can extend the data inputs to nontraditional data sources that could more accurately capture nuances in communities of color. By extension, the flexibility provided by AI and ML can make an AVM valuable at high levels of geography and across different regions of the country. On balance, AVMs may improve the functioning of the housing market because of their varied uses. Consumers—potential homebuyers or sellers—may assess a home's value on popular multiple listing service websites. These sites often use an AVM to estimate a property's value. In addition, an AVM may be used in the process of underwriting a mortgage loan.

Greater and varied use of AVMs throughout the housing industry has led to a need for auditing tools to assess the potential for racial inequity. This set of tools could improve the development of future AVMs, inform federal authorities governing these tools, and ensure that all share the benefits.

Overview of the Literature

Emerging literature suggests the potential for variation in property valuation due to race. Earlier research motivating the scholarly and policy analysis in the area of appraisals suggested that homes in Black neighborhoods were devalued by as much as $48,000, amounting to $156 billion in cumulative losses (Perry, Rothwell, and Harshbarger, 2018). Additional analysis affirmed the potential for racial bias, indicating that majority-Black and Hispanic neighborhoods were more likely to experience undervaluation compared with White neighborhoods (Freddie Mac, 2021). However, the appraiser's race may not inform the potential for bias, because research shows little difference in appraisal valuation discounts by the appraiser's race (Ambrose et.al., 2021). Although most appraisers are White, these results may also suggest that the potential for a racial issue is embedded in the process. Moreover, a lower valuation may provide additional benefits for the homebuyer if the property is negotiated to a lower contract price (Fout, Mota, and Rosenblatt, 2022).

The prospect of racial bias in appraisals coincides with policy analysis assessing AVMs. Research suggests that AVMs can help improve variation in future delinquencies (Bogin and Shui, 2020).

Although an AVM audit suggested the models may produce greater percent error in majority-Black communities than in majority-White ones (Neal et al., 2020), additional research suggests that despite inaccuracies—at least for refinances—AVMs may be more accurate than appraisers (Williamson and Palim, 2022).

The potential for appraisal bias has emerged as a key policy area for the Biden Administration, which created the Interagency Task Force on Property Appraisal and Valuation Equity (PAVE) to address the possibility of discrimination in home valuation. The first PAVE report discussed inequities in home valuation and highlighted the background and administrative steps the task force planned to take regarding appraisals and AVMs. Comments by Vice President Kamala Harris further amplified PAVE activities, highlighting the federal government's desire to prevent algorithmic bias in home valuation.[1] In response, six federal agencies—the National Credit Union Administration, Office of the Comptroller of the Currency, Federal Housing Finance Agency, Federal Reserve, Federal Deposit Insurance Corporation, and Consumer Financial Protection Bureau—issued a comment request on a proposed rule designed to ensure quality control standards for AVMs.[2]

The federal government also indicated the potential for discrimination in AI. Through its Blueprint for an AI Bill of Rights,[3] the Administration hopes to make automated systems work for the American people. The AI Bill of Rights advocates for algorithmic discrimination protections and seeks to ensure that automated systems—including AVMs—are used and designed in an equitable way (Engler, 2022). These steps support the activities of the National Institute of Standards and Technology, an agency that developed a voluntary AI risk management framework,[4] with a focus on fairness, equality, and equity that addresses issues such as harmful bias and discrimination.

This research proposes an AI-based method for assessing the potential for error in AVMs, building on earlier research auditing AVM accuracy (Neal et al., 2020). However, the team shows how the auditing technique may not be the best approach given the structure of the housing data. This research incorporates greater flexibility to better fit housing market data through the use of an AI-ML technique, which presents richer auditing results that can be useful to the broader housing ecosystem. This research sits at the intersection of appraisal bias, AI, and racial equity and is a part of the growing discussion on appraisal bias and finding solutions where this problem exists. It is also a part of the research on AI in the housing industry and has important policy implications. Addressing steps toward improving the appraisal system contributes to the policy discussion on appraisal bias. At the same time, by using AI in its analysis, the research also informs the AI policy area.

This inquiry is of great importance in that AVMs, like human appraisals, seek to answer the question, "What will this property sell for under current market conditions?" This question is

[1] https://www.whitehouse.gov/briefing-room/statements-releases/2023/06/01/fact-sheet-biden-harris-administration-takes-sweeping-action-to-address-racial-bias-in-home-valuations/.

[2] https://ncua.gov/newsroom/press-release/2023/agencies-request-comment-quality-control-standards-automated-valuation-models-proposed-rule.

[3] https://www.whitehouse.gov/ostp/ai-bill-of-rights/.

[4] https://nvlpubs.nist.gov/nistpubs/ai/NIST.AI.100-1.pdf.

important, because it addresses the system in which homeownership transactions occur, the ways in which property values are currently determined, and the racial equity implications therein.

However, an overlapping but separate focus is on the value disparity of properties in Black communities compared with White communities. Although this article seeks to answer the preceding question, a robust assessment of the Black-White housing wealth gap answers a more fundamental question: "What is the true value of this property?" Although a property's selling price and its true value can be similar, certain conditions, including the history of structural racism, serve as a reminder of why they may be radically different as well. This article uses sales price as the benchmark to assess the accuracy of AVM estimates and recommends that, in the future, literature in this area move from the first question anchored to a property's sale price to the second question exploring fundamental value.

The next section of this article describes the AVM data, followed by a discussion of the research team's methodology, findings, conclusions, and implications.

Data Description

This study uses property records data to capture home sales information and produces property-level pairings with AVM estimates and property condition from private data vendors. In addition, this study relies on the American Community Survey to capture neighborhood-level characteristics.

Property Records Data

To examine whether AVM accuracy differs by race, this research compares AVM values with sales prices associated with arm's-length transactions at the property level between majority-Black and majority-White neighborhoods. The team analyzed Atlanta, Georgia, and Memphis, Tennessee. Each city had a significant Black population share and produced solid property-level pairings between AVM estimates and sales prices to analyze. In each city, instead of using the entire core-based statistical area (CBSA), the team used the counties with strong historical deeds data that could be matched with the AVM data. These counties are a small proportion of the total number of counties in each CBSA but account for most of the CBSA population. The Atlanta and Memphis counties account for 17 and 22 percent of the total counties in their CBSAs, respectively, and 63 and 74 percent of their CBSAs' respective populations.

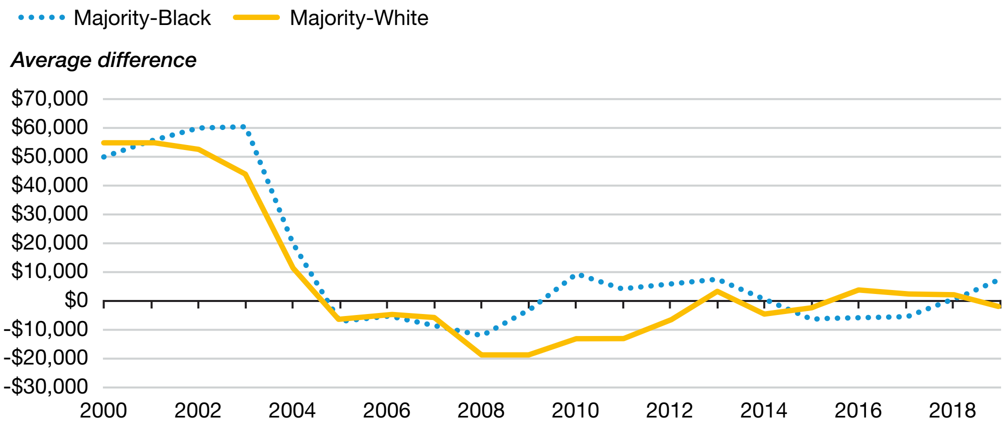
The research team employed property records data from a major data provider to combine information on AVM values, sales prices, and transaction dates for each traded property in 2018 for those selected counties within each city. The team then used 5-year (2014–18) American Community Survey data to extract the share of Black and White homeowners at the census tract level and merge the racial composition information with the property records data.

To characterize the differences between AVM values and sales prices, the research team first calculated directional inaccuracy, the difference between the AVM estimate and the corresponding sale price. For example, if one home is undervalued by $20,000 and a second is overvalued by $20,000, the average directional inaccuracy across these two properties is zero. As exhibit 1 illustrates, direction inaccuracy does not systematically differ according to neighborhood racial composition in Atlanta or Memphis and has not been significant between 2005 and 2018. In addition, the average inaccuracy across majority-Black neighborhoods fluctuated around zero during this period. In the Memphis CBSA, the average error across majority-Black neighborhoods has been systematically below zero over time but only to a modest degree. Exhibit 1 also shows that the average difference in both Atlanta and Memphis across majority-Black neighborhoods is neither consistently above nor consistently below that of majority-White neighborhoods.

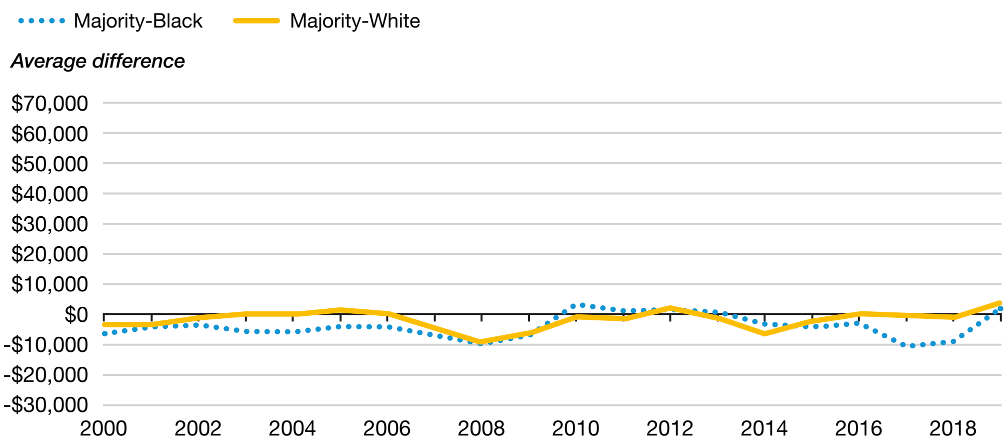
Exhibit 1

Directional Automated Valuation Model Inaccuracy, by Majority Race in Neighborhood

Atlanta-Sandy Springs-Roswell, GA



Memphis, TN-MS-AR

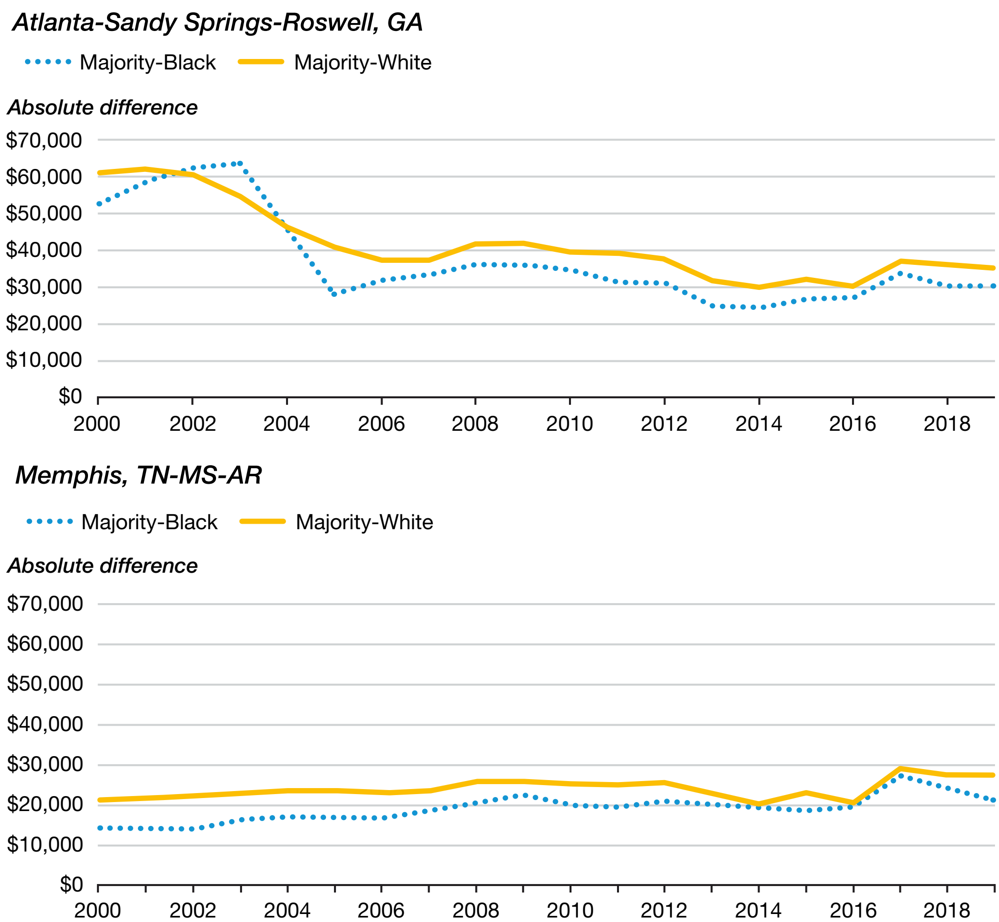


Sources: Authors' calculations of property records data; 2014–18 American Community Survey data

In contrast, the magnitude of inaccuracy, measured as the absolute difference between the AVM estimate and corresponding sale price, in majority-Black neighborhoods is consistently below the absolute inaccuracy in majority-White neighborhoods in the two CBSAs analyzed. For example, when one home is undervalued by $20,000 and a second home is overvalued by $20,000, the absolute value difference is $20,000, irrespective of whether the property is undervalued or overvalued by $20,000. Exhibit 2 shows that, except from 2002 to 2004, AVM inaccuracy in majority-Black neighborhoods in the Atlanta CBSA was smaller than in majority-White neighborhoods. Data from Memphis also reveal that the magnitude of inaccuracy in majority-Black neighborhoods was smaller than in majority-White neighborhoods.

Exhibit 2

Magnitude of Automated Valuation Model Inaccuracy, by Majority Race in Neighborhoods



Sources: Authors' calculations of property records data; 2014–18 American Community Survey data

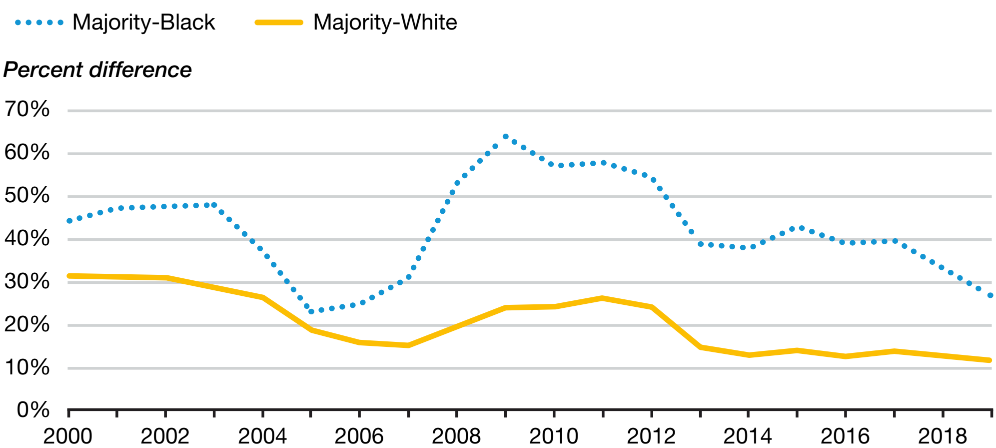
Finally, the research team calculated the percentage magnitude of inaccuracy, or the absolute difference between the AVM estimate and the corresponding sale price divided by the sale price. For example, if one home is undervalued by $20,000 and worth $200,000, the extent of error is greater than that of another home overvalued by $20,000 and worth $300,000. Exhibit 3 illustrates the percentage magnitude of inaccuracy in majority-White and majority-Black neighborhoods in

Atlanta and Memphis between 2000 and 2018. It demonstrates that the magnitude of inaccuracy is much higher in majority-Black neighborhoods in Atlanta and Memphis. The higher percentage magnitude of inaccuracy in majority-Black neighborhoods is attributable to lower average home values than in majority-White neighborhoods. The percentage magnitude of inaccuracy is roughly twice as large in majority-Black neighborhoods as in majority-White neighborhoods and is notably more volatile. For example, in 2009, the percentage magnitude of inaccuracy in majority-Black areas in Atlanta was 64 percent compared with 24 percent in majority-White neighborhoods. Although it has steadily improved since then, as of 2019, it was still more than twice the size in majority-Black neighborhoods than in majority-White neighborhoods. These results are consistent over time in both the Atlanta and Memphis CBSAs.

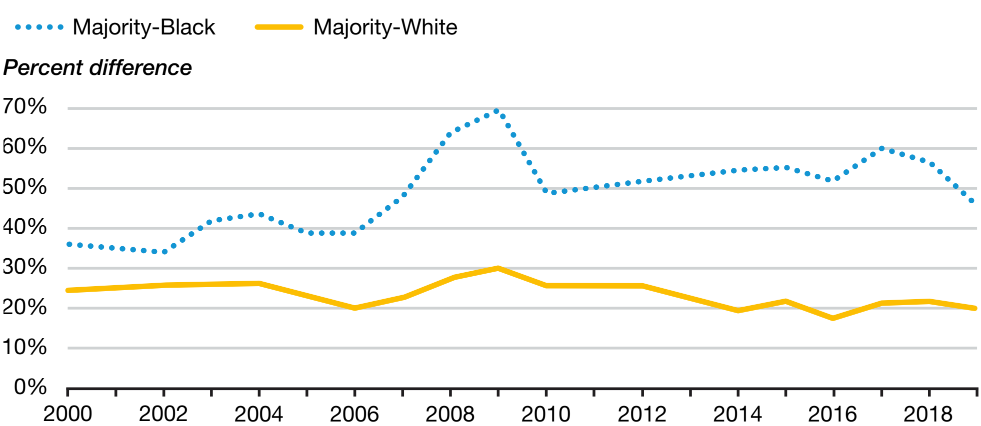
Exhibit 3

Percentage Magnitude of Automated Valuation Model Inaccuracy, by Majority Race in Neighborhood

Atlanta-Sandy Springs-Roswell, GA



Memphis, TN-MS-AR



Sources: Authors' calculations of property records data; 2014–18 American Community Survey data

These findings illustrate that AVMs could both undervalue and overvalue sales prices, both of which can be harmful. Undervaluation can limit wealth gains for homeowners seeking to refinance or sell their homes, and overvaluation may result in credit risk holders underestimating risk and may speed up irrational inflation of property values, potentially resulting in a future home price correction (PAVE, 2022). In addition, lower home values in majority-Black neighborhoods, partly reflecting historic discrimination, increase the risk of AVM error. Although the research team did not find systematic undervaluation bias in AVMs, the team observed that its AVM produced a racially disparate outcome in the form of a greater percentage magnitude of AVM error in majority-Black neighborhoods than in majority-White neighborhoods.

New Data on Property Condition

This research used the exterior condition rating (ECR) measure to capture property-level condition. The property intelligence firm CAPE Analytics provided the team with property-level ECRs. CAPE Analytics creates and applies computer vision algorithms to high-resolution images captured from airplanes to create structured data that include the ECR. The ECR covers all of a parcel's visible external features, including roofs, yards, driveways, and debris. Exhibit 4 provides the five-point scale definitions.

Exhibit 4

CAPE Analytics Exterior Condition Rating Scale Definitions

Rating	Definition
Excellent	Parcel condition falls within the best 5% of parcels
Good	Parcel condition falls within the best 20% but not the best 5% of parcels
Fair	Parcel condition is average (50% of parcels)
Poor	Parcel condition falls within the worst 23% but not the worst 2% of parcels
Severe	Parcel condition falls within the worst 2% of parcels
Unknown	Parcel could be assigned a property condition

Source: CAPE Analytics

In this analysis, the research team matched property records data for Atlanta and Memphis metropolitan areas with the ECRs from CAPE Analytics based on property latitudes and longitudes, parcel lot assessor parcel numbers, and transaction dates. The match rates are 98 percent for Atlanta and 90 percent for Memphis. The small share of unmatched properties was proportionately distributed between majority-Black and majority-White neighborhoods and, thus, did not skew the overall distribution.

For this analysis, the team collapsed the five-point ECR scale from CAPE Analytics into three categories: poor (includes poor and severe), fair, and good (includes good and excellent). Exhibit 5 presents the ECR distributions based on the grouped categories for the matched sample within the Atlanta and Memphis CBSAs.

Exhibit 5

ECR Distribution in the Atlanta and Memphis CBSAs

CBSA	ECR	Majority-Black Neighborhoods	Majority-White Neighborhoods
Atlanta-Sandy Springs-Roswell, GA	Good	9%	13%
Atlanta-Sandy Springs-Roswell, GA	Fair	45%	52%
Atlanta-Sandy Springs-Roswell, GA	Poor	46%	34%
Memphis, TN-MS-AR	Good	10%	14%
Memphis, TN-MS-AR	Fair	46%	52%
Memphis, TN-MS-AR	Poor	44%	34%

CBSA = core-based statistical area. ECR = exterior condition rating.
Source: Authors' calculations using data from the American Community Survey; CAPE Analytics; and a major property records provider

In the Atlanta and Memphis CBSAs, single-family properties in majority-Black neighborhoods are more likely to have a poor rating and are less likely to have a fair or good rating than those in majority-White neighborhoods (exhibit 5). In Atlanta, 46 percent of single-family properties in majority-Black neighborhoods had a poor rating in 2018 compared with 34 percent in majority-White neighborhoods. In Memphis, 44 percent of single-family properties in majority-Black neighborhoods had a poor rating compared with 34 percent in majority-White neighborhoods.

Intuitively, a property condition assessment reflects both external and internal adequacy (Neal, Choi, and Walsh, 2020). Before examining the impact of the ECR measure on the percentage magnitude of AVM error, the team first established that external property condition is a reasonable proxy for the property condition overall, both inside and out. To do so, the team calculated the polychoric correlation—the correlation between two categorical variables—between exterior property conditions and interior structural conditions using American Housing Survey (AHS) data.[5,6]

The AHS is a recognized source of information on property condition, albeit with a limited suite of variables and geographic granularity. The team used the survey's information on roofs and outside walls across owner-occupied homes nationwide to assess exterior conditions and used its information on fundamental or structural problems, such as floors, windows, foundations, and peeling paint, to assess interior conditions. The team found a polychoric correlation of 0.67 between exterior and interior conditions.

This polychoric correlation should be regarded as a lower-bound estimate of the true strength of the correlation because of the AHS's limited variables to capture a property's exterior condition. The ECRs in this analysis cover all of a parcel's visible external features—including roofs, yards, driveways, and debris—compared with AHS variables that cover only roofs and outside walls. Because the ECR variable in this analysis is a more comprehensive measure of exterior condition, its correlation with interior condition would likely be greater than 0.67, suggesting that it should be a reasonable proxy for the property condition overall.

[5] The exterior condition categorical variable is a score variable derived by adding five dummy variables from the AHS—roofhole, roofshin, roofsag, wallside, and wallslope. These variables flag exterior condition problems with the roof and outside walls.

[6] The interior condition categorical variable is a score variable derived by adding six dummy variables from the AHS—floorhole, fndcrumb, paintpeel, wallcrack, winboard, and winbroke. These variables flag interior condition problems with floors, foundation, ceilings, windows, and interior paints.

Methodology

The Ordinary Least Squares (OLS) Approach

To determine how the ECR contributes to the percentage magnitude of AVM appraisal inaccuracy in the Atlanta and Memphis CBSAs, the research team first conducted the OLS regressions with 2018 as the analysis period and focused only on single-family home purchases. The team followed the model specification from previous research and controlled for key neighborhood characteristics affecting the percentage magnitude of AVM inaccuracy, as equation 1 shows (Neal et al., 2020).

$$Pct_{Diffi,2018} = \alpha_0 + \beta ECR_{i,2018} + \gamma Black_{n,2018} + \delta HP_{i,2018} + \theta NC_{n,2018} + \varepsilon_{i,2018} \qquad (1)$$

$Pct_{Diffi,2018}$ is the percentage magnitude of automated valuation model inaccuracy measured by the absolute difference between the sales price and the AVM value divided by the sales price.

$ECR_{i,2018}$ captures property condition in the year 2018. $Black_{n,2018}$ represents the racial composition of the tract in which individual property i locates. This variable is a dummy, with 1 equal to majority-Black neighborhoods in which the share of Black households is greater than 50 percent and 0 equal to majority-White neighborhoods in which the share of White households is greater than 50 percent.[7] $HP_{i,2018}$ is the home value of property i. $NC_{n,2018}$ controls for key neighborhood characteristics affecting the percentage magnitude of AVM inaccuracy. These neighborhood characteristics are grouped along three dimensions: differences in properties within a neighborhood, neighborhood conditions, and turnover rates. Exhibit 6 presents summary statistics of those variables.

Exhibit 6

Summary Statistics

Variable	Majority-Black Neighborhood		Majority-White Neighborhood	
	Mean	SD	Mean	SD
Percent magnitude of AVM inaccuracy	36.3%	58.0%	13.8%	24.5%
Home value	127,756	80,969	329,443	204,256
Property age	46.4	24.3	37.5	22.1
Standard deviation of neighborhood property ages	14.0	7.1	12.5	6.5
Percentage deviation of neighborhood property values	43.2%	14.8%	34.4%	11.3%
Gentrified neighborhood	7.5%	26.3%	2.1%	14.3%
Share of neighborhood distressed home sales	15.7%	20.8%	5.0%	13.7%
Neighborhood median household income	46,198	16,657	92,312	30,955
Neighborhood number of households	2,320	1,328	2,630	1,214
Turnover rate at neighborhood level	8.8%	4.1%	7.5%	3.3%
Exterior Condition Rating				
Good	9.0%	29.0%	13.4%	34.0%
Fair	45.2%	50.0%	52.0%	50.0%
Poor	45.8%	50.0%	34.0%	47.0%

AVM = automated valuation model. SD = standard deviation.

Source: Authors' calculations using data from the American Community Survey, CAPE Analytics, and a major property records provider

[7] Census tracts in which the share of Black or White households is less than 50 percent of the total household population are excluded from the data. In the Atlanta, Memphis, and Washington, D.C. CBSAs, these tracts account for 63, 74, and 56 percent of their respective populations.

To capture property differences within neighborhoods, the research team constructed two variables: the standard deviation of neighborhood property ages and the percentage deviation of neighborhood property values.[8] Standard and percentage deviations measure the dispersion of properties by age and home value, respectively. Jiang and Zhang (2022) show a greater degree of house price dispersion in Black-dominant ZIP Codes.

To capture neighborhood conditions, the team included measures for gentrified neighborhoods, the share of distressed neighborhood sales, the neighborhoods' median household income, and the number of households within a neighborhood. Majority-Black neighborhoods are more likely to experience gentrification, which generally causes permanent and rapid home price increases as land values increase. AVMs cannot quickly pick up these house price shocks, contributing to greater AVM errors in gentrifying neighborhoods. The team considered a neighborhood to be gentrified if it met two criteria (Ellen and O'Regan, 2008): the tract-level income is less than 70 percent of the income in the metropolitan statistical area (MSA) and the neighborhood (identified at the census tract level) experienced at least a 10-percentage-point increase in the ratio of tract-level income to MSA-level income during the year. Under this definition, 7.3 percent of majority-Black neighborhoods in the United States were gentrified in 2018, which is almost five times the share of majority-White neighborhoods that were gentrified. In addition, majority-Black neighborhoods experienced significantly more distressed sales. Forced home sales, such as foreclosures, more often occur among low-price homes than among high-price homes (Campbell, Giglio, and Pathak, 2011). AVM accuracy will likely be compromised if a distressed sale results in a lower price than similar homes in the neighborhood. Among all home sales nationally in majority-Black neighborhoods in 2018, 16.0 percent were distressed home sales, almost four times the rate in majority-White neighborhoods (4.4 percent). The average household income in majority-Black neighborhoods is nearly one-half of that in majority-White neighborhoods. Lower incomes in majority-Black neighborhoods partly explain lower sales prices in these neighborhoods, which can increase the percentage magnitude of inaccuracy (Neal, Choi, and Walsh, 2020).

This analysis defines turnover rate as the number of home sales per year divided by the number of homes. Because AVM algorithms are based on comparable sales, greater turnover rates would provide a larger sample of comparable sales for AVM algorithms to provide more accurate estimates. The turnover rates in majority-Black neighborhoods are slightly higher than those of majority-White neighborhoods.

The OLS regressions usually make several assumptions about the underlying data: (1) a linear relationship between the dependent variable and the independent variables; (2) normality of the residuals—that is, the residual errors are assumed to be normally distributed; (3) homoscedasticity—that is, the residuals are assumed to have a constant variance; and (4) independence of residuals error terms. Given the complexity of the underlying data and the high dimensionality of the independent variables, the data structure in this analysis may not meet those linear assumptions, and thus, the OLS regression may not be the best audit approach to investigate the AVM error.

[8] For each property value, the percentage deviation of neighborhood home values subtracts that known value from the mean property value and divides the result by the property's value.

The LightGBM Approach

To address this issue, the team employed a nonparametric supervised ML approach, the LightGBM—a gradient boosting framework that uses tree-based learning algorithms. Nonparametric supervised ML is a highly innovative and effective vein in predictive data analysis and has several advantages over traditional linear parametric methods, such as OLS. First, ML methods fully use the available historical data. By repeatedly validating the model through training and prediction sets derived from existing data, the methods can map new data entries into specific dependent variables based on relevant independent variables used to train the model. Second, ML methods possess great capacities and effectiveness in handling interrelated variables (for example, collinearity; Aggarwal, 2015), thus boosting prediction accuracy from traditional regression methods. Third, ML methods do not assume linearity and can handle complex datasets that do not fulfill the requirements of traditional regression models.

LightGBM is among the most recent and efficient ML prediction algorithms (Ke et al., 2017). It provides more regularized model formalization and better overfitting control (Ashari, Paryudi, and Tjoa, 2013). It is also an algorithm that assumes no linearity, providing more appropriate handling to the complex dataset in this analysis. Thus, the team chose LightGBM as a nonparametric, tree-based machine learning counterpart to the OLS model, which helped the team explore the broader question of whether and how sophisticated artificial intelligence tools improve the analysis of automated systems.

The team first partitioned the entire dataset into a training set (70 percent) and a testing set (30 percent), then set up cross-validation through a stratified k-fold (k = 5) process. Next, the team entered all relevant independent variables into the LightGBM model as predictors and entered the outcome variable—the percentage magnitude of AVM inaccuracy—as the prediction target. Finally, the team employed a Bayesian optimization procedure to obtain the model parameters supporting the most accurate predictions of the target variable. The following discusses the detailed methodology.

Data Partitioning and Model Validation

This study divided the processed dataset for Memphis and Atlanta into two portions—the training and testing sets—to regulate the efficiency of the ML procedures. The LightGBM model is trained using only the training set and tested using only the testing set. This split is vital to demonstrate and tune the model's response to new data being processed for the first time. For the robustness of the division, the research team put 70 percent of the data into the training portion and the remaining 30 percent into the testing portion.

To enhance the model's validity, accuracy, and robustness, the team also employed a five-fold cross-validation procedure on the training set and adopted the k-fold (k = 5) cross-validation because of its efficiency and smoothness during the validation. Each dataset was randomly separated into k numbers of folds; k-1 folds were used for training purposes, and the remaining fold was simultaneously used for testing. The results over the k-testing folds were averaged at the end. Exhibit 7 lists the summary statistics on several key variables to show that the partition process does not distort the distribution in either the training or the testing datasets.

Exhibit 7

Summary Statistics: Training and Test Data

	Training Data (Mean)	Test Data (Mean)
Share of majority-Black neighborhoods	42.2%	42.3%
Home value	244,067	244,839
Exterior Condition Rating		
Good	11.6%	11.5%
Fair	49.2%	49.4%
Poor	39.3%	39.1%

Source: Authors

Model Parameters

To tune the hyperparameters of the LightGBM model in conjunction with the k-fold cross-validation procedure, the team employed a Bayesian optimization procedure to obtain the model parameters that would best predict the regression outcome. The parameter optimization boundaries are in exhibit 8. With those parameters, the team obtained its optimized LightGBM prediction model based on the 70-percent training set.

Exhibit 8

Model Parameters in the LightGBM Model

The parameter optimization boundaries:	
Learning rate	0–1
Number of leaves	5–40
Minimum gain to split	0–10
Minimum sum of hessian in leaf	0–20
The parameter value in the final optimized LightGBM model:	
Number of threads	6
Number of leaves	25
Learning rate	0.468
Minimum gain to split	1.823
Minimum sum of hessian in leaf	9.517

Source: Authors

Evaluation of Model Accuracy

Root mean square error (RMSE) is the standard deviation of the residuals (predicted errors) and is used to measure model prediction accuracy. The research team took advantage of its strong interpretability because it has the same unit as the regression target variable. The team tested the RMSE for the LightGBM prediction model and compared it with the RMSE for the OLS model to test whether the LightGBM model made more accurate predictions.

Identification of AVM Racial Disparity: Feature Importance

Shapley Additive Explanations (SHAP) is a novel way of computing feature contribution toward the prediction while preserving the sum of contributions being equal to the final outcome. It is especially well suited for tree-based models. SHAP values calculate a feature's importance by comparing what a model predicts with and without the feature. Given that the order in which a model sees a feature can affect its predictions, SHAP values account for all possible orders to make sure all features are fairly compared.

The team calculated the SHAP values for each predictor to determine the predictors' relative importance and effect on the model outcome. Their SHAP values allowed the team to delve deeper into the predictive model's complexity, partially unveil the ML black box, and evaluate the effect of neighborhood race and ECR on predicted AVM error.

Quantification of Feature Importance: Synthetic Control Method

Although the SHAP value could provide evidence of a specific feature's importance, it does not quantify the magnitude of the impact. To quantify the impact, the research team employed a synthetic control method to examine identified racial disparity in the AVM valuations, the ECR's impact, and the effect of the intersection of neighborhood majority race and the ECR.

The team extracted all properties in majority-Black neighborhoods from the test set and treated this selected group as the benchmark dataset. First, the team predicted the AVM error for the benchmark dataset using its constructed LightGBM model. Second, to create the corresponding synthetic datasets, the team changed the neighborhood majority race variable from Black to White, holding everything else constant. The team created 10 synthetic datasets and altered the neighborhood race in the share of properties, starting from 10 to 100 percent. For the tenth synthetic dataset, all properties' neighborhood majority race was randomly switched from Black to White. Third, the team predicted the AVM error for each synthetic dataset using its constructed LightGBM model and then obtained 10 corresponding predicted mean values of AVM error. Finally, the team compared the predicted target variable from the synthetic datasets with the benchmark dataset. This difference between the two predicted values measured the racial disparity. The properties in the synthetic groups were the same as those in the benchmark group, except that the neighborhood majority race was different. The results were expected to shed light on whether systemic racism is a key factor behind the AVM error.

Results

OLS Regressions

Exhibit 9 shows the OLS regression results, indicating that an ECR rating worse than good would raise the percentage magnitude of AVM error. Relative to an otherwise similar property with a good rating, a property with a fair rating would increase the AVM's percentage magnitude of error by 2.72 percentage points. Similarly, relative to a property with a good rating, a property with a poor rating would further increase AVM inaccuracy, increasing the percentage magnitude of error by 4.35 percentage points. In this case, the magnitude of the coefficient means that for a home with an average sales price of $250,000, having a poor rating is associated with a $10,875 greater percentage AVM error than a property with a good rating, holding all other attributes constant. Controlling for the ECR slightly reduces the magnitude of this Black neighborhood coefficient from 3.593 percentage points in column four to 3.499 percentage points in column five. This result indicates that even when controlling for property condition, location in a majority-Black neighborhood rather than a majority-White one still raises the percentage magnitude of error by 3.499 percentage points. The difference is a $4,549 greater percentage AVM error for a home with an average sales price of $130,000 in a majority-Black neighborhood compared with a property with the same attributes and sales price in a majority-White neighborhood.

Exhibit 9

Ordinary Lease Squares Regression Results

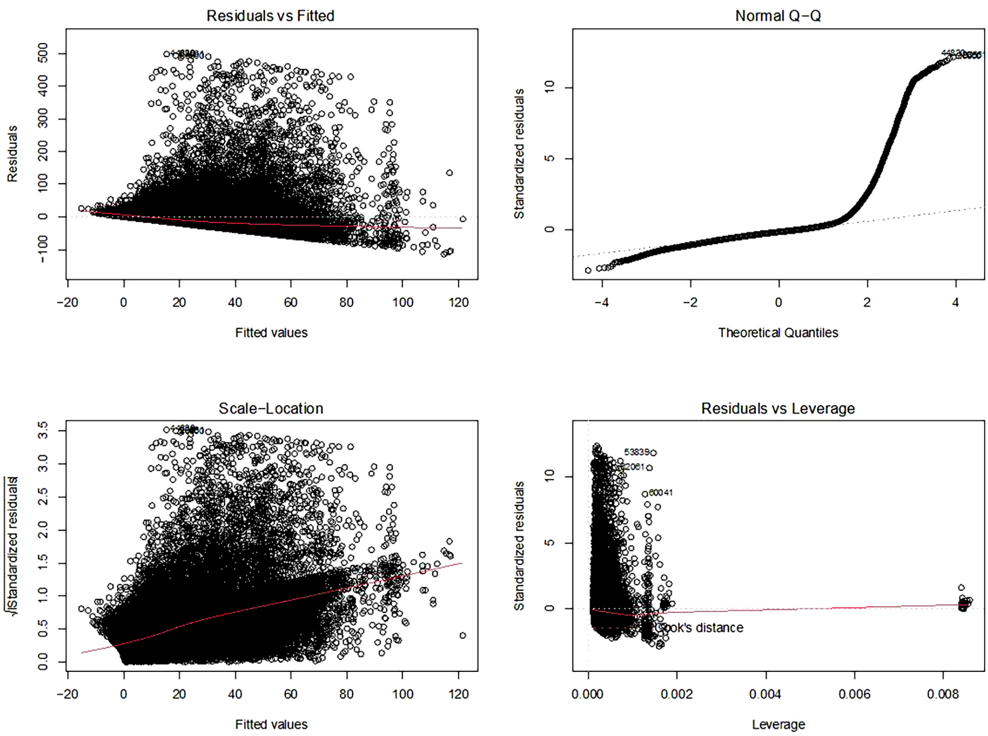
	Dependent Variable: Percentage Magnitude of Automated Valuation Model Inaccuracy				
	(1)	(2)	(3)	(4)	(5)
Black neighborhood	21.024*** (0.393)	4.816*** (0.504)	4.040*** (0.499)	3.593*** (0.542)	3.499*** (0.542)
Log (Home value)		– 15.785*** (0.316)	– 12.535*** (0.328)	– 10.358*** (0.402)	– 10.075*** (0.403)
Standard deviation of neighborhood property ages			0.155*** (0.028)	0.058** (0.029)	0.059** (0.028)
Percentage deviation of neighborhood property values (%)			0.453*** (0.014)	0.422*** (0.014)	0.422*** (0.014)
Share of neighborhood distressed home sales (%)				– 0.005 (0.010)	– 0.006 (0.010)
Gentrified neighborhood				2.155*** (0.817)	2.174*** (0.817)
Log (Neighborhood median household income)				– 4.116*** (0.652)	– 4.153*** (0.652)
Log (Number of households in neighborhood)				– 5.107*** (0.377)	– 5.016*** (0.377)
Neighborhood-level turnover rate (%)				– 0.246*** (0.049)	– 0.230*** (0.049)
Exterior condition rating (ECR): Fair					2.718*** (0.530)
ECR: Poor					4.350*** (0.546)
Constant	13.860*** (0.752)	213.240*** (4.063)	156.708*** (4.336)	219.975*** (6.708)	213.170*** (6.765)
County fixed effects	Yes	Yes	Yes	Yes	Yes
Observations	62,609	62,609	62,606	62,606	62,606
R^2	0.086	0.121	0.138	0.142	0.143
Adjusted R^2	0.086	0.121	0.138	0.142	0.143
Residual standard error	41.587 (df = 62602)	40.784 (df = 62601)	40.383 (df = 62596)	40.297 (df = 62591)	40.276 (df = 62589)
F-statistics	981.818*** (df = 6; 62606)	1,230.664*** (df = 7; 62601)	1,115.744*** (df = 9; 62596)	739.866*** (df = 14; 62591)	652.256*** (df = 16; 62589)

***p < 0.05. ***p < 0.01. df = degrees of freedom.*
Note: The dependent variable is the percentage magnitude of automated valuation model error.
Source: Authors

The research team ran several diagnostic tests to confirm whether an OLS regression is the best approach for examining property condition's effect on AVM accuracy (exhibit 10). The residuals-versus-fitted plot indicates that the randomness of the error term was not met. The Normal Q–Q plot shows that the residuals from the OLS regressions (column five) are not normally distributed. In addition, the scale-location plot shows a severe heteroscedasticity problem. All these results suggest that OLS regression may not be the best approach.

Exhibit 10

Diagnostic Tests for Linear Regression Accuracy



Source: Authors

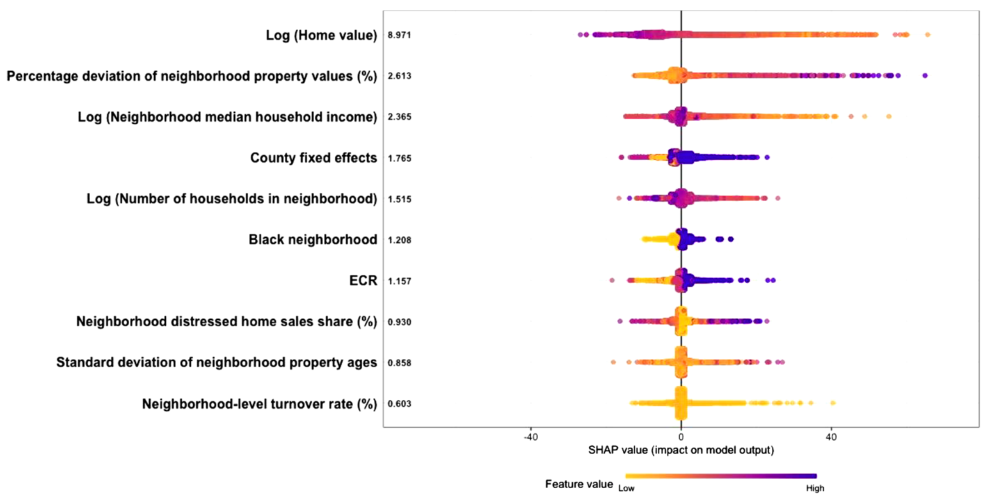
LightGBM

LightGBM has a greater predictive power than the OLS regressions. After completing data partitioning, model validation, and parameter tuning, the optimized LightGBM model produced an RMSE of 40.4. Applying the same data partitioning procedure to the OLS regression produced an RMSE of 46.2. This result suggests that LightGBM produces a 5.8-percentage-point improvement in the model's fit to explain the AVM inaccuracy. The magnitude of RMSE improvement does not differ between majority-Black and majority-White neighborhoods, with only around a 0.05-percentage-point difference. These results validate the team's selection of LightGBM over OLS regressions with respect to evaluating the identified AVM racial disparity. By relaxing the linear assumptions, this nonparametric, tree-based ML approach provides more appropriate handling of the complex dataset.

Majority-Black neighborhoods are associated with greater predicted AVM inaccuracy. Exhibit 11 illustrates the SHAP values for each feature. The y-axis displays the feature name in order of importance from top to bottom. The value next to the variable name is the mean SHAP value. The x-axis is the SHAP value. Each point represents a row from the training dataset. The gradient color represents that variable's original value. Continuous numerical variables, such as the log of home values, can contain the whole color spectrum. Dummy variables, such as majority-Black neighborhood, can take only two colors.

Exhibit 11

Shapley Additive Explanations Values



ECR = exterior condition rating. SHAP = Shapley Additive Explanations.
Source: Authors

For example, the percentage deviation of neighborhood property values is a key feature contributing to the AVM percentage error prediction, because it ranks as the second feature after the home values in log form. The percentage deviation of neighborhood property values is associated with high and positive values on the target (exhibit 11). The color distribution shows a high value in the color bar at the bottom of the figure, with purple indicating a high percentage deviation of neighborhood property values. The preponderance of purple observations indicates a positive value on the right side of the zero vertical line, signaling the SHAP value is above zero. Given that it is a continuous variable, its color scheme contains the full-color spectrum of the feature value from low (yellow) to high (purple). This result suggests that greater heterogeneity with respect to property values in a neighborhood contributes to a greater predicted percentage magnitude of AVM error.

Exhibit 11 illustrates that the majority-Black neighborhood variable is associated with high and positive values on the target. The value is high because of the figure value color bar and is positive from the x-axis SHAP value. Because this is a dummy variable, its color scheme contains only two colors for low (yellow) and high (purple). This high and positive relationship suggests that compared with majority-White neighborhoods, AVM error in majority-Black neighborhoods is greater.

Similarly, the ECR is associated with high and positive values on the target. The ratings are coded as 1 (good), 2 (fair), and 3 (poor), so high ECR values mean poor property conditions, indicating that properties in poor condition are associated with greater AVM inaccuracy. In addition, when looking at the order of importance from top to bottom, the majority-Black neighborhood variable ranks higher than the exterior condition rating, distressed sales share, and turnover rate variables. The SHAP values for neighborhood majority race combined with its ranking align with what was found in the OLS regressions, suggesting that even though an AVM algorithm does not have disparate input such as race, it still can produce racial disparities.

The Role of Historic Racism in AVM Estimates

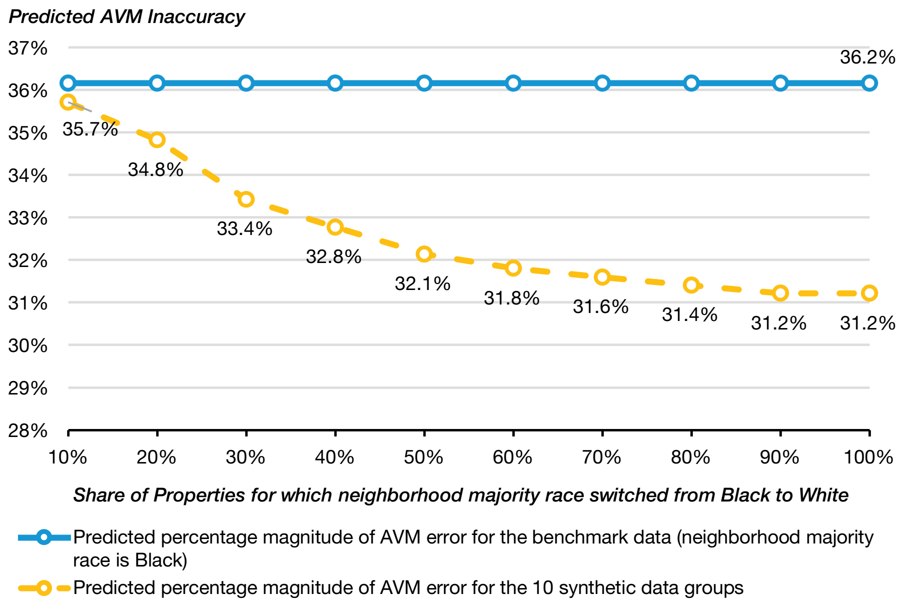
The blue line in exhibit 12 illustrates the predicted percentage magnitude of AVM error for the benchmark data, and the yellow line represents the predicted percentage magnitude of AVM error for the 10 synthetic data groups.[9] The x-axis represents the share of properties for which neighborhood majority race switched from Black to White. As the share increases, the gap between the two lines widens. This result indicates that, for example, if 60 percent of properties currently in majority-Black neighborhoods “move” to majority-White neighborhoods while keeping all other attributes constant, their associated predicted percentage magnitude of AVM error could decline from 36.2 to 31.8 percent, a 4.4-percentage-point difference. Further, if all properties currently in majority-Black neighborhoods “move” to majority-White neighborhoods, the predicted AVM error could decline by 5.0 percentage points, which is an upper-bound estimate of the racial disparity in AVMs. Such results suggest that historic racism could be a key factor behind greater AVM error in majority-Black neighborhoods.

A similar synthetic control approach was applied to examine the ECR’s impact. The team used all properties with a poor ECR from the prediction set as the benchmark dataset (exhibit 13), then created 10 corresponding synthetic datasets by changing the ECR from poor to good, holding everything else constant. Again, the share that switched from poor to good ECR increased from 10 to 100 percent. For the tenth synthetic dataset, all properties’ ECRs changed from poor to good. The team calculated the predicted AVM error for the benchmark group (the blue solid line) and the synthetic data groups (the yellow dotted line). Exhibit 13 demonstrates that if all properties currently in poor condition are upgraded to good condition, all else constant, their associated AVM error could decline from 26.1 to 21.8 percent, a 4.3 percentage-point difference. This decline in AVM error provides further evidence that policies to improve housing adequacy could reduce the adverse effect of the percentage magnitude of AVM error.

[9] The team ran the model 10 times against the same dataset, which is the 30-percent test set. Because the shares of properties in the test set were randomly selected using a with-replacement approach, results would differ slightly every time these 10 synthetic datasets were created. The 70–30 percent train-test split and the fivefold cross-validation ensured that the overall monotonically decreasing trend in exhibit 3 was statistically robust. In addition, the with-replacement selection caused slightly different results every time these 10 synthetic datasets were created.

Exhibit 12

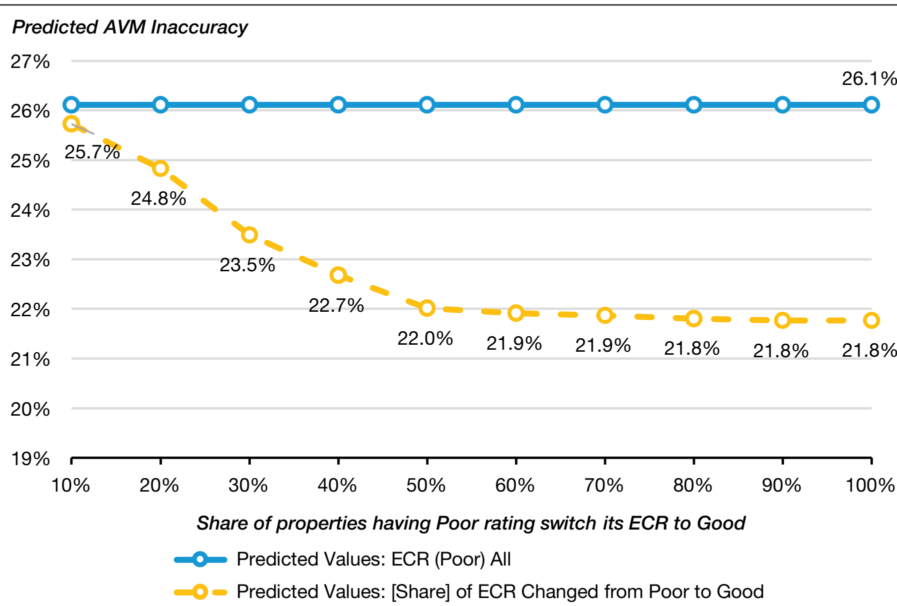
Racial Disparity: Neighborhood Majority Race



AVM = automated valuation model.
Source: Authors

Exhibit 13

Impact of ECR

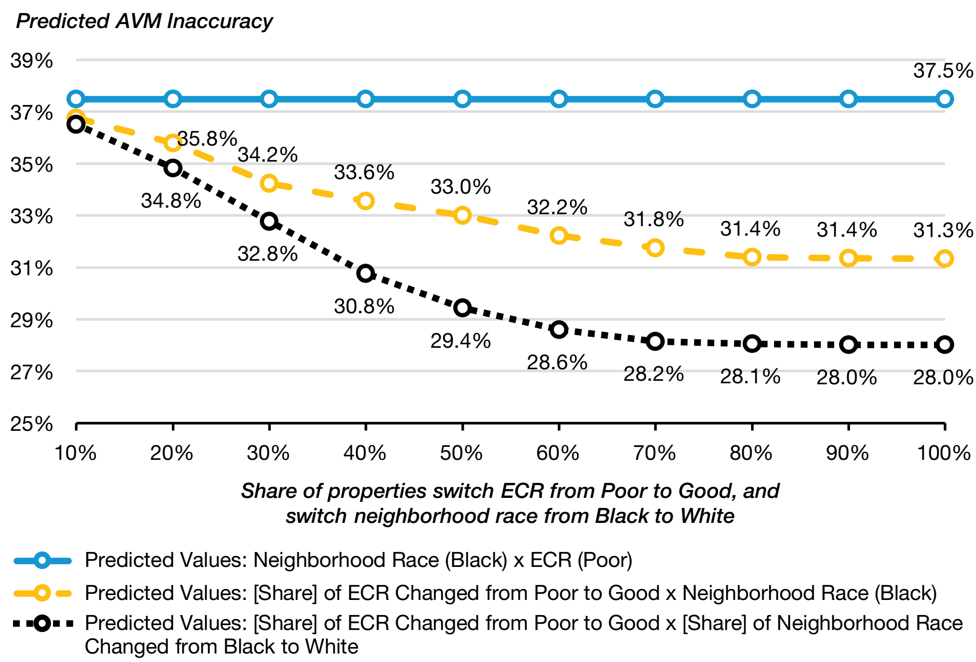


AVM = automated valuation model. ECR = exterior condition rating.
Source: Authors

Finally, the team examined the intersection of neighborhood majority race and ECR. The benchmark data consisted of properties in majority-Black neighborhoods rated as poor (the blue solid line in exhibit 14). The team changed the benchmark data by first altering the ECR from poor to good while keeping the neighborhood majority race and all other attributes constant. As before, for each synthetic data group, the team increased the share of change by 10 percent, ranging from 10 to 100 percent. Next, the team changed the neighborhood majority race from Black to White. When changing all the properties in the benchmark group from a poor to a good ECR, the predicted percentage magnitude of automated valuation model error fell from 37.5 to 31.3 percent (exhibit 14). The AVM error further declined to 28 percent once the neighborhood majority race flipped from Black to White. The gap between the two dash lines represents the effect of the intersection of neighborhood majority race and ECR. In other words, for two identical properties that have improved their ECR from poor to good, the home in a majority-Black neighborhood still experiences a 3.4-percentage-point greater percentage magnitude of AVM error, further suggesting that racial differences at the neighborhood level, which can reflect the effect of systemic discrimination, can play a role in producing percentage magnitude of AVM error.

Exhibit 14

Effect of the Intersection of Neighborhood Majority Race and ECR



AVM = automated valuation model. ECR = exterior condition rating.
Source: Authors

Conclusions and Implications

After including a direct control for property condition and employing more sophisticated ML techniques to examine the role of data omission and model selection, the research team finds that data on property condition and more sophisticated ML techniques can help more accurately assess the percentage magnitude of AVM error and its underlying contributors. Properties with poorer property conditions and in neighborhoods with more heterogeneous properties and a greater share of distressed sales are associated with greater predicted AVM error. In addition, despite data improvement and ML, evidence still shows that the percentage magnitude of AVM error is greater in majority-Black neighborhoods.

The evidence from this research contributes to the policy debate on appraisal bias and provides a quantifiable measure for auditing the performance of an AVM in majority-Black neighborhoods compared with majority-White ones. It serves as a starting point for developing a range of indicators that guard against a disproportionate effect on protected classes. If mortgage and housing providers are interested in determining the underlying contributors to the percentage magnitude of AVM error, Ordinary Least Squares regression is helpful, based on the findings of this research. However, if the goal is to assess the shortcomings of AVM models and their underlying contributors more accurately using large and complex data, the housing industry should consider exploring algorithms the AI community has developed. AI tools could enhance understanding of the complexity of predictive algorithms and partially unveil the AVM black box. The nonlinear regression results demonstrate that using a LightGBM model that includes ECR data could produce a 5.8-percentage-point improvement in the model's fit to assess the percentage magnitude of AVM error. Such results suggest that this nonparametric machine learning model more accurately copes with the complexity of variables of multiple dimensions. In addition, the SHAP values and the synthetic control approach shed light on the shortcomings of AVM algorithms that often get hidden in the black box.

Furthermore, the continued significance of a neighborhood's majority race suggests the need for integrating racial equity into the design of AVM algorithms. Researchers cannot yet reject the role of historic racism, which has persistently penetrated through home values, property conditions, and neighborhood conditions. Inequities in each of these dimensions can produce lower home values, less adequate housing, and lower household incomes across majority-Black neighborhoods. Encouraging regulatory oversight of AVMs and ML models will help ensure AVMs do not rely on biased data that could reinforce past discrimination.

This research illustrates the potential racial differences in one type of property valuation method, helping to shed light on current and standard valuation practices across the industry. However, the history of racial discrimination suggests conditions whereby the sale price may not be equivalent to a property's true value. Future research building relaxing assumptions embedded in current industry practice may represent steps toward a more fundamental assessment of differences between property values in Black communities compared with those in White communities.

Acknowledgments

The authors thank or acknowledge the Urban Institute for their financial support of this body of research, and particularly staff of the Housing Finance Policy Center, including Laurie Goodman and Janneke Ratcliffe for their input over the development of this research. The authors would also like to thank Cape Analytics, especially Sean Begley and Shane Lee (formerly of Cape Analytics) for their use of data and expert input. The authors would also like to thank the provider of their AVM data, who chooses to remain anonymous. This article was also enriched by many conversations over the years with subject matter experts around the housing industry. Finally, the authors would like to acknowledge the peer reviewers and the U.S. Department of Housing and Urban Development for their general support of policy research.

Authors

Linna Zhu is a senior research associate in the Housing Finance Policy Center at the Urban Institute. Michael Neal is a senior fellow in the Housing Finance Policy Center at the Urban Institute. Caitlin Young is a J.D. candidate at Yale Law School.

References

Aggarwal, Charu C. 2015. *Data Mining: The Textbook*, Vol. 1. New York: Springer.

Ambrose, Brent W., James Conklin, N. Edward Coulson, Moussa Diop, and Luis A. Lopez. 2021. *Do Appraiser and Borrower Race Affect Valuation?* https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3951587.

Ashari, Ahmad, Iman Paryudi, and A. Min Tjoa. 2013. "Performance Comparison between Naïve Bayes, Decision Tree and k-Nearest Neighbor in Searching Alternative Design in an Energy Simulation Tool," *International Journal of Advanced Computer Science and Applications* 4 (11).

Bogin, Alexander N., and Jessica Shui. 2020. "Appraisal Accuracy and Automated Valuation Models in Rural Areas," *The Journal of Real Estate Finance and Economics* 60 (1–2): 40–52.

Campbell, John Y., Stefano Giglio, and Parag Pathak. 2011. "Forced Sales and House Prices," *American Economic Review* 101 (5): 2108–31.

Ellen, Ingrid Gould, and Katherine O'Regan. 2008. "Reversal of Fortunes? Lower-Income Urban Neighbourhoods in the US in the 1990s," *Urban Studies* 45 (4): 845–69.

Engler, Alex. 2022. "The AI Bill of Rights Makes Uneven Progress on Algorithmic Protections." Washington, DC: Brookings Institution. https://www.brookings.edu/articles/the-ai-bill-of-rights-makes-uneven-progress-on-algorithmic-protections/.

Fout, Hamilton, Nuno Mota, and Eric Rosenblatt. 2022. "When Appraisers Go Low, Contracts Go Lower: The Impact of Expert Opinions on Transaction Prices," *The Journal of Real Estate Finance and Economics* 65 (3): 451–491.

Freddie Mac. 2021. *Racial and Ethnic Valuation Gaps in Home Purchase Appraisals*. Freddie Mac Economic and Housing Research Note. McLean, VA: Freddie Mac. https://www.freddiemac.com/research/insight/20210920-home-appraisals.

Interagency Task Force on Property Appraisal and Valuation Equity (PAVE). 2022. *Action Plan to Advance Property Appraisal and Valuation Equity: Closing the Racial Wealth Gap by Addressing Mis-valuations for Families and Communities of Color*. Washington, DC: PAVE. https://pave.hud.gov/sites/pave.hud.gov/files/documents/PAVEActionPlan.pdf.

Jiang, Erica X., and Anthony L. Zhang. 2022. *Collateral Value Uncertainty and Mortgage Credit Provision*. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4015452.

Ke, Guolin, Qi Meng, Thomas Finley, Taifeng Wang, Wei Chen, Weidong Ma, Qiwei Ye, and Tie-Yan Liu. 2017. "LightGBM: A Highly Efficient Gradient Boosting Decision Tree." In *Advances in Neural Information Processing Systems 30*, edited by Isabelle Guyon, Ulrike von Luxburg, Samy Bengio, Hanna M. Wallach, Rob Fergus, S.V.N. Viswanathan, and Roman Garnett. Long Beach, CA: Neural Information Processing Systems: 3149–57.

Neal, Michael, Jung Hyun Choi, and John Walsh. 2020. *Before the Pandemic, Homeowners of Color Faced Structural Barriers to the Benefits of Homeownership*. Washington, DC: Urban Institute. https://www.urban.org/sites/default/files/publication/102781/before-the-pandemic-homeowners-of-color-faced-structural-barriers-to-the-benefits-of-homeownership.pdf.

Neal, Michael, Sarah Strochak, Linna Zhu, and Caitlin Young. 2020. *How Automated Valuation Models Can Disproportionately Affect Majority-Black Neighborhoods*. Washington, DC: Urban Institute. https://www.urban.org/sites/default/files/publication/103429/how-automated-valuation-models-can-disproportionately-affect-majority-black-neighborhoods_1.pdf.

Perry, Andre, Jonathon Rothwell, and David Harshbarger. 2018. *The Devaluation of Assets in Black Neighborhoods: The Case of Residential Property*. Washington, DC: Brookings Institution.

Williamson, Jake, and Mark Palim. 2022. *Appraising the Appraisal: A Closer Look at Divergent Appraisal Values for Black and White Borrowers Refinancing Their Home*. Washington, DC: Fannie Mae. https://www.fanniemae.com/media/42541/display.

Additional Reading

Akinwumi, Michael, Lisa Rice, and Snigdha Sharma. 2022. *Purpose, Process, and Monitoring: A New Framework for Auditing Algorithmic Bias in Housing and Lending*. Washington, DC: National Fair Housing Alliance. https://nationalfairhousing.org/wp-content/uploads/2022/02/PPM_Framework_02_17_2022.pdf.

Akinwumi, Michael, John Merrill, Lisa Rice, Kareem Saleh, and Maureen Yap. 2021. *An AI Lending Policy Agenda for the Federal Financial Regulators*. Washington, DC: Brookings Institution. https://www.brookings.edu/articles/an-ai-fair-lending-policy-agenda-for-the-federal-financial-regulators/.

Axelrod, Judah, Alena Stern, Michael Neal, and Linna Zhu. 2022. "Comment Letter on NIST Initial Draft of Artificial Intelligence Risk Management Framework." Washington, DC: Urban

Institute. https://www.urban.org/sites/default/files/2022-05/Comment%20Letter%20on%20NIST%20 Initial%20Draft%20of%20Artificial%20Intelligence%20Risk%20Management%20Framework_0.pdf.

Buchak, Greg, Gregor Matvos, Tomasz Piskorski, and Amit Seru. 2020. *iBuyers: Liquidity in Real Estate Markets?* National University of Singapore, Institute of Real Estate and Urban Studies. https://ireus.nus.edu.sg/wp-content/uploads/2020/10/iBuyers-Liquidity-in-Real-Estate-Markets-by-Tomasz-Piskorski-.pdf.

Conklin, James, N. Edward Coulson, and Moussa Diop. 2022. *Distressed Comps*. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4008585.

Fout, Hamilton, and Vincent Yao. 2016. *Housing Market Effects of Appraising Below Contract*. Washington, DC: Fannie Mae; Atlanta: Georgia State University. https://www.researchgate.net/profile/Vincent-Yao-2/publication/298807852_Housing_Market_Effects_of_Appraising_Below_Contract/links/57c04c7508ae2f5eb3321d07/Housing-Market-Effects-of-Appraising-Below-Contract.pdf.

Howell, Junia, and Elizabeth Korver-Glenn. 2018. "Neighborhoods, Race, and the Twenty-First Century Housing Appraisal Industry," *Sociology of Race and Ethnicity* 4 (4): 473–90. https://doi.org/10.1177/2332649218755178.

Narragon, Melissa, Danny Wiley, Doug McManus, Vivian Li, Kangli Li, Xue Wu, and Kadiri Karamon. 2021. *Racial and Ethnic Valuation Gaps in Home Purchase Appraisals*. Freddie Mac Economic and Housing Research Note. McLean, VA: Freddie Mac. https://www.freddiemac.com/research/insight/20210920-home-appraisals.

Neal, Michael, Jung Hyun Choi, Kathryn Reynolds, Joe Schilling, Gideon Berger, Elizabeth Champion, and Caitlin Young. 2021. *Why Do Households of Color Own Only a Quarter of the Nation's Housing Wealth When They Compose a Third of the Nation's Households?* Washington, DC: Urban Institute. https://www.urban.org/sites/default/files/publication/105020/why-do-households-of-color-own-only-a-quarter-of-the-nations-housing-wealth-when-they-compose-a-third-of-the-nations-households.pdf.

Salkind, Neil J., ed. 2010. *Encyclopedia of Research Design*. New York: Sage Publications.

Shrestha, Noora. 2020. "Detecting Multicollinearity in Regression Analysis," *American Journal of Applied Mathematics and Statistics* 8 (2): 39–42.

Local Landscapes of Assisted Housing: Reconciling Layered and Imprecise Administrative Data for Research Purposes

Shiloh Deitz
Will B. Payne
Eric Seymour
Kathe Newman
Lauren Nolan
Rutgers University

Abstract

Understanding the stock of rental housing affordable to lower-income households is a crucial task for local governments aiming to meet rising demand and inform policy priorities. However, enumerating the number of units with public housing, Project Based Section 8, and Low-Income Housing Tax Credit (LIHTC) assistance and identifying precisely where those units are located is deceptively challenging. Although federal datasets with that information are easily accessible, development and building location information may be unavailable or imprecise. Critically, identifying units that receive more than one form of assistance is hard, especially units with LIHTC. To address these challenges in New Jersey, the authors developed a largely automated and replicable process for precisely placing subsidized housing units into tax parcels. Doing so enables linking units across federal programs and with state and local data and to more accurately aggregate counts to integrate with decennial census and American Community Survey (ACS) data from the U.S. Census Bureau. Within New Jersey, the research team re-geocoded records in three datasets using two commercial geocoding services, assigned them confidence scores, designated records for manual handling, and then assigned them to parcels. Following those steps, they identified more than 15,000 units statewide with overlapping federal subsidies, which would lead to a 12-percent overcount of subsidized units in the state if the three datasets were used as given (and up to a 40-percent overcount in individual municipalities). By reusing and reconciling those datasets at the parcel level, researchers can more accurately enumerate rental units associated with different levels of subsidy depth and duration, a crucial task for identifying housing needs within and beyond the assisted rental stock.

Introduction

Housing researchers now have access to many detailed datasets about federally assisted rental housing, including for the most widely used housing subsidy programs such as public housing, Project-Based Section 8, and Low-Income Housing Tax Credits (LIHTC). Individually, each dataset offers rich, program-specific information. To view the entire landscape of assisted housing in a given location, including units with more than one form of subsidy, the datasets must be integrated, but that task presents multiple challenges.[1] Those challenges include differences in development names and addresses across datasets, available fields, and levels of aggregation, with some information available at the contract level and other information available at the building or development level. Furthermore, some rental units appear in multiple datasets because they receive project-based subsidies from more than one federal program, an administrative data challenge that has increased as LIHTC is used to renovate or redevelop public and other federally assisted housing.

As part of the New Jersey State of Affordable Rental Housing (NJSOARH) project, a large-scale research program on rental housing in New Jersey, the research team sought to understand the landscape of federally assisted and otherwise affordable rental housing across the entire state. They collected the most granular public data on federally assisted housing and assigned developments[2] to parcels, both to precisely locate them in space and to identify overlapping subsidies. That process made it possible to estimate the share of rental housing with select federal assistance; to link assisted housing developments to other administrative data, including census data; and to visualize them within communities.

Given the scale of analysis and recent improvements in automated approaches for cleaning and linking heterogeneous datasets, the authors developed a largely automated process for placing assisted developments and buildings in tax parcels using address information contained in the source data. They then used those parcel assignments to identify potentially layered subsidies that were subsequently manually verified. The central innovation of the approach is the use of two independent geocoding engines (from Esri and Google) to triangulate the accuracy of parcel matches for federally assisted housing developments and buildings, allowing the authors to focus manual verification work on records that did not return high-quality matches to the same parcel from both geocoding engines. Recent work has highlighted the promise of using widely available geocoding engines to improve the accuracy and precision of the geographic information contained in some assisted housing datasets (Wilson et al., 2023). Although that geographic information may be suitable for the program administration purposes those datasets were designed to serve, researchers needing to identify the precise location of projects on the ground often need to make adjustments to those data.

Although the use of a single geocoding engine has been demonstrated to yield improvements from the perspective of researchers seeking to place projects in parcels, even the results reported

[1] Some project-based units also receive project- or tenant-based federal housing vouchers. These data are only available aggregated to the census tract. Although the research team was aware that overlaps exist, they were unable to untangle them using these datasets, which limits some integration with American Community Survey (ACS) and census data rental unit counts.

[2] The Low-Income Tax Credit dataset uses *developments*, whereas the Project-Based Section 8 dataset uses *projects* to refer to the same thing. Throughout the article, the authors refer to *projects* or *developments* as appropriate.

as highly accurate exhibit some degree of uncertainty due to the different underlying data that different geocoders are drawing from and the internal logics of the geocoding engines (Prener and Fox, 2021). Triangulating the results of two independent commercial geocoders and their associated metadata gave the authors an ability to see where both services agreed on the location of a given street address, allowing them to cleanly and efficiently divide data into automated and manual workflows. Specifically, the authors identified which records were easily matched to parcels based on their street addresses and which records returned less confident results that required further location verification. They supplemented that output with data fields from the corresponding tax parcels, such as property ownership, which enabled them to design a process that yielded highly accurate automated results in more than 70 percent of the input records and allowed them to focus their attention on manual checking the remaining 30 percent. The rest of this article is organized as follows: the authors describe the historical context and existing work upon which their methods are built; they then provide an overview of the stages of their method and what they revealed; finally, they discuss the importance of doing this work and the implications of the overlapping subsidies they found in New Jersey's affordable housing stock.

Existing Data Correction and Integration Efforts

The authors are not the first to link federal housing datasets across subsidy programs. Their process draws inspiration from a set of efforts to integrate federally and other assisted housing datasets nationally and at the scale of specific states and cities, sometimes with different objectives and using different data sources. The research team also draws from recent related efforts to improve the accuracy and precision of geocoding processes and work that uses those improvements to locate assisted housing more accurately for research purposes. First, the authors discuss those existing assisted housing data integration efforts, which serve as the substantive stream of influence; second, they discuss the methodological work influencing their process.

HUD's Picture of Subsidized Households (PSH), produced by the Office of Policy Development and Research (PD&R), was first published in the 1990s, with annual releases most years since 1996. It was one of the first efforts to compile and make data from multiple HUD programs publicly available (Taghavi, 2008). This widely used data source makes information about major forms of project-based assistance available down to the level of individual projects, making it easy to combine them with other demographic datasets and create community maps. However, the PSH does not include LIHTC units, a critical source of support that is increasingly used to redevelop and preserve existing public and federally assisted housing. Simply adding LIHTC data to the PSH counts would produce an overcount of federally assisted units, because LIHTC is often used alongside other subsidy programs. The PSH also lacks data on individual building locations for scattered-site projects, which could lead to counting some housing units in the wrong place.

In addition, several related projects synthesize multiple forms of federal and other housing assistance, primarily to identify preservation needs for housing with time-delimited assistance. Those projects include the Assisted Housing Inventory (AHI) developed by the Shimberg Center at the University of Florida, the Subsidized Housing Inventory Program (SHIP) through the Furman Center at New York University, and the National Housing Preservation Database (NHPD) created by the Public and Affordable Housing Research Corporation (PAHRC) and the National Low Income Housing Coalition (NLIHC) (Reina and Williams, 2012). Both the AHI and SHIP integrate information about privately owned housing developments that receive government assistance. The NHPD also integrates information about public housing. Although the NHPD offers address-level information for New Jersey, the addresses are used as given in input datasets, not definitively placed in a geocoding and mapping workflow that can identify when different street addresses correspond to the same tax parcel. In the authors' case, organizing the data at the parcel level helped with incorporating LIHTC data supplied directly by the New Jersey Housing Finance and Mortgage Agency.

In addition to those projects directly oriented toward integrating assisted housing data, the authors draw on recent methodological work in precisely geocoding street addresses. Previous work has sought to correct the spatial coordinate information for individual assisted housing datasets to improve the quality of research involving distance-based measurements. Numerous areas of inquiry link the location of assisted housing developments to a range of localized outcomes, including property values (An et al., forthcoming; Deng, 2011) and eviction (Harrison et al., 2021; Lens et al., 2020). The precise placement of properties in space is needed to produce the best estimates of the association between assisted housing and these other contextual or distance-based processes.

Motivated by those concerns, Wilson et al. (2023) found that nearly 50 percent of the HUD-provided spatial coordinates for California LIHTC projects are outside the boundaries of the true tax parcel in which a project is located, compared with an accuracy level of 80 percent when geocoding the input address with Google's geocoding service. However, Wilson et al. (2023) also found that when Google results were incorrect, they were farther from the correct location than HUD's coordinates (derived from the freely available United States Postal Service geocoder), which, although often outside the boundary of the true parcel location, were typically near the true location.

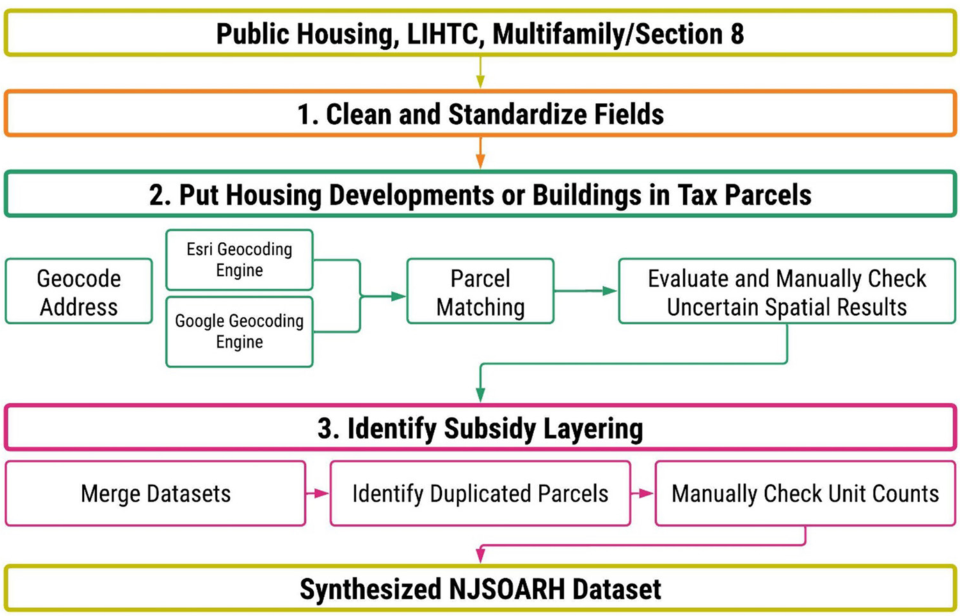
Beyond the specific application of improving the spatial accuracy of assisted housing, researchers have been developing more sophisticated methods for improving geocoding through the use of multiple services. Prener and Fox (2021) created a suite of tools for similar work in Saint Louis creating a custom composite geocoder—in their case, with access to authoritative local government positional data as a backstop. Researchers now have access to a suite of services for geocoding or converting address information to point coordinates. Although those services are generally quite accurate, the underlying reference data, normalization process, and process for matching inputs with coordinates are opaque (see Teske, 2014). The value of a good composite geocoding workflow is the triangulation between multiple sources. If geocoding results correspond across services (which have different underlying data and logics for converting address strings to spatial coordinates), they are significantly less likely to be false. Having surveyed existing efforts to integrate federally assisted housing data and the methodological literature on geocoding addresses, the authors crafted a largely automated process for integrating federally assisted housing at the parcel level using multiple geocoders in conjunction with tax parcel data. They placed assisted housing developments into parcels, similar to the SHIP project mentioned previously. Doing so enabled them to do four important things: (1) count units within the census geographies where they were actually located; (2) identify units that receive subsidies from more than one of the federal programs they are studying; (3) develop accurate narratives about housing security and neighborhood change; and (4) link together federal data, such as federal tenant-based vouchers, and local administrative data in the future.

Putting Federally Subsidized Housing in Parcels

To integrate our selected federally assisted housing datasets, the authors developed a three-step process to assign housing buildings or developments to parcels (see exhibit 1). First, they cleaned and standardized address fields from each input dataset. Second, they placed standardized addresses for buildings and projects in tax parcel polygons. Third, they used the results of the previous steps to identify developments that appeared in more than one of the assisted housing datasets and created a new composite unit field to avoid counting units more than once. Through those operations, the research team created a parcel-level dataset enabling them to integrate data across the major federal project-based subsidy programs and link it to local data sources.

Exhibit 1

Process of Integrating Federally Assisted Datasets



LIHTC = Low-Income Housing Tax Credit. NJSOARH = New Jersey State of Affordable Rental Housing.
Source: Authors

The input datasets included a set of federal and state datasets from 2022 with information about housing developments and buildings. HUD's public housing buildings dataset records the location and tenant characteristics of public housing buildings, which may constitute one of several within a single development (HUD 2022c). HUD's multifamily assistance and Section 8 database records information about development location, size, contract origination, and contract expiration dates. It does not provide building-level information or service dates (HUD 2022b).[3] The New Jersey

[3] This dataset included a variety of programs, including the 811 Rental Assistance Demonstration (RAD), Other Section 8 (S8) New, Other S8 Rehab, Project Rental Assistance Contract (PRAC) 202/811, Pension Fund, S8 FmHA, S8 Loan Management, S8 and Section 202, S8 Preservation, S8 Property Disp., S8 RAD Mod Rehab Conversion, S8 Public Housing Conversion, S8 RAD RS/RAP Conv, S8 State Agency, and Section 202.

Housing Mortgage Finance Agency (NJHMFA) prepared a dataset of all LIHTC projects ever subsidized in New Jersey that recorded their location, number of units, and year placed in service (NJHMFA, 2022) but did not identify the location of individual buildings.[4] The research team used publicly available integrated New Jersey tax assessor and parcel location data to triangulate results using property location and ownership information (New Jersey Office of GIS, 2022).

In the next section, the authors walk through their iterative, semi-automated process to describe what they did, how they did it, and what happened as a result. They discuss the challenges and how they addressed them, where possible.

Cleaning and Standardizing Fields

The first step was to clean and standardize the address fields in each of the input datasets. The research team combined addresses into a single field and employed string substitution and related forms of text cleaning to clean and standardize, for instance, by changing all road type abbreviations to their full name (e.g., “St” to “Street”). The authors spelled out those abbreviations to improve geocoding and string matching.[5] In the case of fields containing an address range (e.g., 10–20 Chestnut St.), they split the record into two separate addresses (one for each end of the range) to geocode each and determine whether they were placed in the same parcel further in the process. Preparing addresses for geocoding revealed two main challenges to pinpointing assisted housing locations: first, that they could be scattered site (e.g., address range and multiple addresses within the same input field); and second, that addresses may refer to administrative office mailing addresses rather than actual residence locations (e.g., a P.O. Box). They flagged records with incomplete, unusual (duplicate or multiple addresses and those with unusual characters), or missing addresses.[6] Across all three datasets were 278 of those locations, or 6 percent of all records.

Putting Housing Developments or Buildings in Tax Parcels

With a set of cleaned and standardized addresses, in second step, the authors employed a three-step process to assign addresses from the input data to parcels (exhibit 2). First, they geocoded

[4] HUD also provides an LIHTC building-level dataset that can help locate individual buildings in scattered-site developments. It does not include information about the unit count in each building, so the research team did not use the dataset. The dataset provided by NJHMFA included all LIHTC projects ever contracted in New Jersey. One hundred ten of these projects were put in service more than 30 years ago, and the authors could not match HUD contracts to them (HUD 2022a). Those 110 developments account for 2,183 units, which are not part of the unit tabulation. Twenty-three LIHTC projects in the NJHMFA data put in service less than 30 years ago could not be matched to a contract; the authors determined that those projects likely exist, and included them. Across those 23 developments were 2,206 units, of which 1,352 had layered subsidies (LIHTC and public housing or multifamily subsidies). Accounting for that fact, including those developments may have led to an LIHTC unit overcount of up to 854 units.

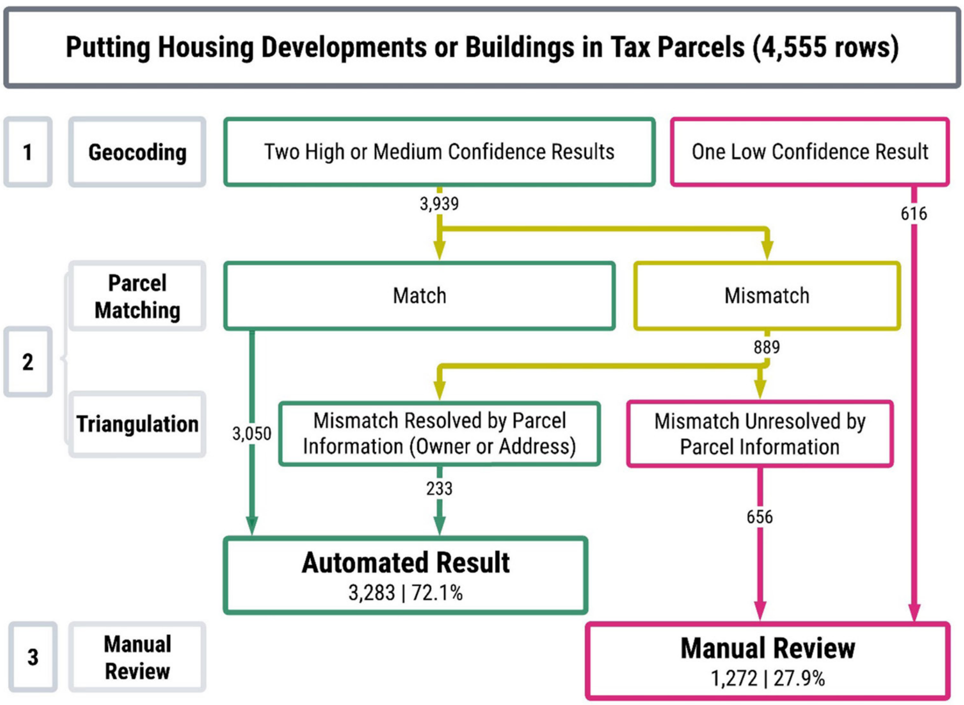
[5] The authors created their own standardization process using Python code based on common address inconsistencies in New Jersey. Standardization included the following steps: (1) capitalize the entire string; (2) fill ZIP Codes with leading zeros so that they are all 5 digits long (New Jersey ZIP Codes start with 0 and are frequently imported as integers); (3) standardize abbreviations for Avenue, Boulevard, Circle, Drive, Highway, Parkway, Place, Road, Street, Terrace, Lane, Court, 1st–10th, North, South, East, and West; and (4) remove unit or apartment numbers from addresses.

[6] These challenging addresses were more prevalent in the multifamily/Section 8 dataset than the LIHTC data. In the multifamily/Section 8 dataset, multiple addresses were often found on one line in varying ways (e.g., separated by an & symbol or a comma). This variance was not present in the LIHTC data, possibly because they were curated and cleaned by one entity: the HMFA.

addresses with two independent geocoding engines to assign point locations to developments or buildings. Second, they matched the geocode results to parcel polygons and triangulated with the parcel address and owner. Third, they manually checked the results with uncertain outcomes.

Exhibit 2

Putting Housing Developments or Buildings in Tax Parcels



Source: Authors

Geocoding

The first step in the process to automate the placement of buildings and developments in parcels was to geocode the input data using cleaned and standardized address fields (see exhibit 2, step 1). The research team ran the cleaned addresses through Google and Esri geocoding services using each company's application programming interfaces (APIs) using Python (Esri, n.d.; Google Maps Platform, n.d.). The geocoding services provided metadata about each result, including the result address, where found, and the match type (e.g., rooftop, street, intersection, or centroid of a larger geographic area). Based on the match type and the similarity between the match address and input address,[7] the authors assigned geocode results as having high, medium, or low confidence (see exhibit 3). If either geocoder returned a low confidence result, the authors manually checked the record. This category accounted for 616 records, or 13.5 percent of the total input addresses. Those 616 records included

[7] Esri provided a score from 0 to 100 grading this correspondence. The authors calculated their own score for the Google results using the FuzzyWuzzy Python package (Cohen, 2020) to measure how closely the input address corresponded with the returned address field and its parts.

246 records from the public housing buildings dataset that lacked a street address and only specified the municipality; the remaining 370 low confidence results displayed no obvious pattern. The remaining 3,939 records moved to the next automated step, parcel matching.

Exhibit 3

Geocode Confidence Metrics

Confidence	Description	Esri	Google
High	A rooftop/building-level result that the authors believe to be trustworthy and think will match to the correct parcel without issues	A record that was a sub-address (an individual suite or unit in a building) *OR* A record that was a point address (building or rooftop), and the match score is at least 97.7[a]	Every record with a matching house number and ZIP Code and a full address match greater than or equal to 90 *AND* In the “geometric center” (center of a feature) and “premise” (a named location, such as a building) result *or* “rooftop” (result is accurate to the street address) *or* “range interpolated” (approximate place on street segment)
Medium	A rooftop-level result with slightly less certainty or an interpolated street result (i.e., road centerline)	All other “point address” results not labeled high confidence *OR* “Street address” with a match score of at least 99.75 *OR* “Street address ext” (match out of range based on house number) and located in New Jersey[b]	No criteria of “low” or “high” confidence geocodes met
Low	Result that matched the wrong geography (town, ZCTA, street, or intersection), no result, or result outside New Jersey	Multiple addresses (usually identified by “/” or “&”) *OR* No criteria of “high” and “medium” confidence met	Result that was “approximate,” *or* a “geometric center” and was not a “premise”-level result *OR* House numbers of source data and output that do not match *OR* Street address match score less than 80

ZCTA = ZIP Code Tabulation Area.

[a]The authors selected all match score cutoffs after sampling the output and choosing values that led to no false positives (defined as geocodes flagged as correct that were in the incorrect parcel for the given input address). The suitability of those thresholds for their purposes is confirmed by the data presented in table 4.

[b]Although “street address ext” is less precise than “street address,” the research team’s explorations of the output showed that it was very uncommon for an address to be labeled that way. The few “street address ext” results in New Jersey were similar to “street address” results and fit the “medium” confidence category.

Source: Authors

Parcel Matching

Next, the authors assigned the 3,939 high- or medium-quality geocode results to parcel polygon locations (see exhibit 2, step 2). This process added six new fields (three for each geocoder) for each record based on the results of spatially joining the Esri and Google geocodes to parcels: the parcel identifiers for each set of geocoded points, the distance between the returned point and the nearest parcel,[8] and a flag for whether the result matched to multiple parcels for each service.[9]

For 3,050 of the 3,939 records that satisfied the earlier requirement of two high- or medium-confidence results, both geocoding services returned points that fell within the same parcel. The authors accepted those parcel results without additional manual steps, leaving 889 records that matched to different parcels. For those 889 records, the research team triangulated between parcel information and the input data to determine whether either of the parcels was likely the correct match. Using the FuzzyWuzzy string matching package for Python, they evaluated whether the parcel address in New Jersey's MOD-IV parcel database corresponded to the development or building address. For the public housing data, they also checked whether a housing authority owned the parcel by looking at the owner's name and property class fields. That process resolved 233 of the 889 high- or medium-quality geocode records that matched to different parcels. Combining those 233 records with the 3,050 accepted previously, the authors determined that 72 percent of their total records did not require further attention. They manually checked the rest as described in the following section.

Manual Process

The research team's process for automating parcel assignment flagged records with missing or un-geocodable address information and low-confidence geocoding results, producing a subset of 1,272 records for manual processing. To manually check results, the authors drew on additional resources, including the New Jersey Department of Community Affairs affordable housing dataset (NJ DCA 2022); Rowan University School of Earth and Environment's interactive map of New Jersey tax parcels (2023); Google Street View imagery for house numbers, development signs, and correspondence between housing appearance and public housing or developer websites; and other online sources (housing authorities, developers, investors, news articles, property, and planning documents and websites). In many cases, Google Street View enabled the research team to find the parcel. For the remaining records, the authors used many, if not all, of the sources to triangulate and place units in the correct parcel. When they located a property and assigned it to a parcel(s), they made notes about how they made their decisions and recorded the Internet links and documents that aided their decisionmaking. In fewer than 10 cases, the authors contacted someone with local knowledge, such as a public housing authority staff member. Ultimately, they placed all but 15 projects in parcels. They assigned 5 of the remaining properties to a block or mega-parcel (set of adjacent parcels) and 11 to a municipality.

Identifying Subsidy Layering

Having placed buildings and developments into parcels, the research team used those assignments to identify parcels that appeared in more than one of the source datasets. However, shared

[8] For most records this value was zero because the geocode point location was within the parcel.

[9] Some point locations matched to multiple parcels. This was most common in condominium developments where multiple parcels are layered on top of each other. The research team flagged those cases as multiple parcel matches, dropped the duplicates, and determined the primary parcel for those developments, an underlying parcel polygon encompassing the individual condominium units, manually.

parcels are not an unambiguous indicator of shared subsidies. Multiple projects can be located on individual parcels; furthermore, individual assisted buildings or developments can include different sets of units subsidized by nonoverlapping government programs. Whereas some projects contain units carrying multiple forms of assistance, others assist different sets of units with distinct forms of assistance, so no overlap occurs at the unit level.

The authors checked records with shared parcels using the resources from the manual parcel matching process described previously. However, this time, first they ensured that two separate developments with different subsidies were not on the same parcel; and second, they identified where subsidies applied to the same units. The authors triangulated that information in the assisted and total unit fields in each dataset, along with the manual sources, as part of a qualitative process to make a determination. Using that process, they estimated that 12 percent, or 15,529, of the total units across the datasets appeared in more than one dataset (see exhibit 4). Many of the overlaps are the result of LIHTC being used to renovate public housing or Multifamily/Section 8 developments. For example, LIHTC was used alongside federal multifamily assistance in 13,549 units. If the authors had not carried out the process previously described to reconcile the three input datasets and simply aggregated the unit counts across all three without looking for subsidy overlaps, they would have overcounted subsidized units in the state by 12 percent.

Exhibit 4

Total Units by Federal Program Before and After Identifying Layered Subsidies

Unit Count	Units in Source Data[a]	NJSOARH Output Data	Difference[b]
Total Subsidized	144,411	128,882	–15,529 (12.0%)
Only LIHTC Affordable	61,269	45,740	–15,529 (34.0%)
Only Public Housing	29,651	27,671	–1,980 (7.2%)
Only Multifamily/Section 8 (MF)	53,491	39,942	–13,549 (33.9%)
MF & LIHTC Affordable	0	13,549	+13,549 (100%)
Public Housing & LIHTC Affordable	0	1,980	+1,980 (100%)

LIHTC = Low-Income Housing Tax Credit. MF = multifamily. NJSOARH = New Jersey State of Affordable Rental Housing.

[a]In the process of cleaning the data, the research team sometimes modified unit counts for reasons other than subsidy overlap. For example, they found that 502 units of public housing had been demolished and 2,366 units of LIHTC housing were more than 30 years old and missing a contract. Those units are excluded from the totals.

[b]These values and percentages are the number of units that would have been overcounted, meaning the difference between the two counts divided by the cleaned data count.

Source: Authors

Through this process, the authors noticed that the extent of housing subsidy overlap varies widely by geography, in large part because of the clustering of multifamily developments in space. Thus, the impact of overcounting units due to subsidy layering depends on the prevalence of LIHTC in different communities and, in some communities, would affect a far greater proportion of units than the statewide average of 12 percent. Without this process, the research team would have double-counted 4,624 units of federally subsidized housing in Newark, or 22 percent of the city's total subsidized stock (see exhibit 5). An overcount of 25 percent or more would have occurred in Atlantic City, East Orange, Elizabeth, Pennsauken, and Trenton. Many of those municipalities have large numbers of older assisted housing that is being redeveloped using LIHTC. Those cities are also some of the most populous in the state, with acute housing needs and entrenched social disparities; working from accurate understandings of the existing affordable housing landscape in these places is crucial.

Exhibit 5

Top 10 New Jersey Municipalities for Most Overcounted Units[10]

Municipality (County)	Units in Source Data	NJSOARH Output Data	Difference[a]
Newark (Essex)	25,686	21,062	4,624 (22.0%)
Trenton (Mercer)	6,888	5,481	1,407 (25.7%)
East Orange (Essex)	4,567	3,357	1,210 (36.0%)
Atlantic City (Atlantic)	5,587	4,408	1,179 (26.7%)
Jersey City (Hudson)	10,319	9,319	1,000 (10.7%)
Elizabeth (Union)	3,749	2,770	979 (35.3%)
Camden (Camden)	7,276	6,375	901 (14.1%)
Paterson (Passaic)	5,166	4,284	882 (20.6%)
Orange (Essex)	2,961	2,429	532 (21.9%)
Pennsauken (Camden)	1,065	762	303 (39.8%)

NJSOARH = New Jersey State of Affordable Rental Housing.

[a]These values and percentages are the number of units that would have been overcounted, meaning the difference between the two counts divided by the cleaned data count.

Source: Authors

Process Evaluation

At the end of the parcel assignment and deduplication process, the research team assessed the results of their workflow to better understand the impact of key components of the methodology and thresholds that were set along the way. They explored two questions. First, was the automated process effective at placing buildings or developments in the correct parcels? Second, were the criteria too conservative, and could the authors have automated more of the process without sacrificing accuracy at the parcel level?

Results of Manual Validation of Select Automated Results

In this section, the authors evaluate whether the automated process was effective at placing buildings or projects in the correct parcels. They conducted this evaluation by comparing the parcel match results from their automatic process to the final results that they later determined manually. This sample included the 898 records (29 percent of automated results) that were accepted automatically but also manually checked, either because they were flagged for possible subsidy overlap or they were located in one of seven communities in which the authors were doing more in-depth research.

Nearly all of the automated results (98 percent) either correctly matched to only one parcel (86 percent) or matched to one of the correct parcels in a multiparcel development (12 percent, see exhibit 6. For example, the LIHTC and multifamily datasets are aggregated at the development

[10] Future work could attempt to better understand the contextual and demographic drivers of redevelopment in those communities.

level and typically provide only a single address per development, even in cases in which developments contain multiple buildings that straddle several tax parcels. Because this sample includes all records likely having layered subsidies, it contains a large share of developments likely sitting on multiple parcels, typically because of redevelopment.

Even when the result of the automated process was not in the right parcel (about 2 percent of the cases), it was for reasons largely out of the research team's control that no automated process could hope to address. For about one-half of those cases, the input address provided in the dataset was not the actual location of the housing, and for the second half, the result was clearly wrong (e.g., the address points to the middle of a field or a shopping mall), but the authors could not ascertain where the housing was even after extensive manual research. In those cases, the address may have been inputted incorrectly by the initial compiler of the administrative datasets. That result underscores the robustness of our conservative process for automating results. Based on those results, the thresholds for accepting results automatically were sufficiently stringent.

Exhibit 6

Results of Manual Validation of Select Automated Results

Result	%	Count
Automated and manual checks correspond	98.0	880
– Matched to a unique parcel	86.3	775
– Matched to one parcel in a multiparcel development	11.7	105
Automated and manual checks do not correspond	2.0	18
– Input address from administrative dataset is incorrect	0.9	8
– Result is wrong; housing location still unclear	1.1	10
Total	**100**	**898**

Source: Authors

Results of Manual Validation of Low-Quality Matches

In this section, the authors evaluate whether they could have accepted more geocoding results programmatically and thus manually checked fewer records. They conducted this evaluation using a sample of 994 records that had valid, geocodable addresses[11] but were manually checked because the research team could not ascertain their location through geocoding and parcel matching. Using that sample, the authors examined the relationship between the different possible sources of uncertainty that led to manual review (see exhibit 7, which shows low-quality matches from one or both geocoders or conflicting parcel matches between geocoders) and the ultimate accuracy of each of the two geocoders after manual validation. That examination enabled the authors to see whether, for instance, an address that returned a low-quality Esri geocoding match but a medium- or high-quality Google match could have been assigned to the Google result with confidence. If all or nearly all records would have been assigned to the parcel that the authors ultimately accepted on the basis of a single medium- or high-quality match, then the process of requiring parcel matches from both geocoders or independent verification from tax parcel data would have been unnecessarily restrictive. However, if the research team loosened the process to allow a

[11] Of the total records, 278 had missing or multiple addresses.

single medium- or high-quality geocoding match to determine the correct parcel, and a nontrivial number of those records would have led them to accept inaccurate parcel assignment, that outcome would have contravened the objective of deliberately minimizing false positives, ensuring reliable results and an efficient use of the team's time for manually checking disputed results.

Exhibit 7

Results of Manual Validation of Low-Quality Matches

Reason for Manual Check	Records by Reason		Correct Results After Manual Check					
			Google		Esri		Neither	
	#	%	#	%	#	%	#	%
Both Low Quality	47	4.7	14	29.8	11	23.4	29	**61.7**
Esri Low Quality	143	14.4	112	**78.3**	41	28.7	28	19.6
Google Low Quality	175	17.6	85	48.6	123	**70.3**	39	22.3
Parcel Mismatch	629	63.3	368	**58.5**	137	21.8	124	19.7
Total	994	100	579	**58.2**	312	31.4	220	22.1

Note: The bolded value in each row shows the highest percentage of ultimate accuracy for each type of low-quality geocode: Google, Esri, or neither.
Source: Authors

The results of this manual validation align with the goals and expectations the research team had when setting up their process. Looking across the possible reasons for manually checking and the parcel-level accuracy of each geocoder, no clear patterns emerge that might have been used as rules to modify the automated process. In cases in which one geocoder has low-quality results, the other geocoder is more likely to be correct but not at a high enough rate to simply accept them automatically: 78 percent of Google results were correct when only Esri had a low-quality match, and 70 percent of Esri results were correct when only Google had a low-quality match. The majority of the results that the authors manually checked were checked because both geocodes matched to different parcels despite both geocode results exhibiting high or medium quality (629, or 63 percent of the cases). In those cases, the Google geocode corresponded with the ultimate decision more than twice as often as the Esri one (368, or 59 percent, compared with 137, or 22 percent). Because parcel mismatches account for the majority of the records in this sample, Google has the overall edge in accuracy. Across all reasons for manually checking, the Google geocode result was in the parcel that the authors determined to be true 58 percent of the time (compared with Esri 31 percent of the time). But the 532 records (42 percent) that the Google geocoder incorrectly located are dispersed across all the reasons for manually checking in exhibit 7 and show no clear pattern that would have allowed the authors to identify them and handle more records procedurally without resulting in false positives.

The authors explored one other potential avenue for additional efficiency gains in the process: low-confidence results that still led to parcel matches. In 147 cases, the Google and Esri results matched to the same parcel, but at least one of the two geocoders had low match quality; 117 of those results ultimately corresponded with the manual decision (79.6 percent). However, given the goal of having parcel-level certainty whenever possible, having no way to know which 20 percent of the results were incorrect was not acceptable. The research team concluded that they could not wring additional gains out of the automated portion of the workflow.

Discussion

HUD publishes a wealth of easily accessible and richly detailed administrative housing datasets. Improving capacity to link those datasets to each other and to local data holds much promise for performing increasingly sophisticated housing analysis to aid public policy decisions. The effort outlined in this article sought to link federal datasets at the parcel level both because the authors needed to identify units with layered federal subsidies and because they wanted to situate those data more precisely in their community context—for instance, by placing properties in parcel maps, which better reveals their community presence. This activity is particularly important with the presence of processes such as gentrification and redevelopment, which often occur block by block rather than within neat categories such as ZIP Codes or census tracts. To achieve that goal, the authors developed a largely automated process that linked federal housing datasets and the state LIHTC dataset. The process was efficient and effective. It gave them high confidence in their automated results and focused their energy on necessary manual checks.

As large administrative datasets and the tools to analyze them are increasingly available to wider audiences, not taking care to understand the limitations of those datasets may introduce biased results, generating considerable error at scale. Although the results presented here have limitations, the authors developed a robust process that employed a conservative approach to accepting geocoding matches that gave them high confidence in the automated matches. Rather than accepting results for which either of the geocoding results were high when the other was low, the authors accepted only results for which both geocodes were high or medium quality to severely limit false positives.

This methodological conservatism increased the amount of manual work, but the manual work helped the research team to better understand both the data and the choices communities are making to renovate, preserve, and create new affordable housing with the subsidies. In fact, for the authors purposes, the manual work was invaluable for reasons beyond confirming or fixing the results of the automated workflow. It revealed the extent of redevelopment and renovation of the older stock of public and federally assisted housing. It also highlighted that many LIHTC and federally assisted housing projects include multiple buildings that are not revealed through address point matching. And by manually checking data, the authors were able to see the extensive redevelopment of existing public and federally assisted housing occurring throughout New Jersey. Although many people have a general understanding of redevelopment initiatives—including Housing Opportunities for People Everywhere (HOPE) VI, Choice Neighborhoods, and the Rental Assistance Demonstration (RAD)—seeing how and where they are implemented is important to grasp the implications. The manual process gave the authors insight into those trends.

Limitations of the Present Work

Although this process enabled the research team to do much of what they intended, there were several limitations. First, only the public housing dataset included building-level data; the other two sources provided development-level information. Thus, the authors were not able to place all buildings in the right location; they may not have correctly placed some sprawling and scattered-site developments. In addition, because voucher information is publicly available only at the tract level, the authors did not explore project- or tenant-based housing choice vouchers in this project, which they expect to overlap with many project-based units. Finally, the ability to replicate this process is dependent on access to two private geocoding services, which may charge rates for the use of their service that some data users would find prohibitive and which may cease to be supported in the future if the companies who created and maintain them change their business objectives, unlike public alternatives such as the United States Postal Service geocoder used by HUD in its public housing buildings dataset.

Conclusion

By reusing and reconciling subsidized housing datasets, researchers can more accurately enumerate rental units associated with particular levels of subsidy depth and duration, which are crucial for identifying housing needs within and beyond the assisted rental stock. For the authors' purposes, understanding what exists in subsidized housing projects and how the subsidies are layered to make housing affordable allowed them to bring those data into conversation with other datasets and to more accurately understand the landscape of affordable rental housing in New Jersey. The research team's aim is that other researchers and practitioners can apply the methods and lessons learned in their process to better understand the federally subsidized housing landscape in other parts of the country and to consider some of the tradeoffs between efficiency and accuracy inherent in using off-the-shelf geospatial information in administrative datasets for research purposes.

Acknowledgments

Support for this article was provided by the Robert Wood Johnson Foundation with the assistance of the New Jersey Housing Mortgage Finance Agency, which provided the LIHTC dataset. The views expressed here do not necessarily reflect the views of the Foundation.

Authors

Shiloh Deitz is a postdoctoral associate in the Edward J. Bloustein School of Planning and Public Policy at Rutgers University. Will B. Payne is an assistant professor in the Edward J. Bloustein School of Planning and Public Policy at Rutgers University. Eric Seymour is an assistant professor in the Edward J. Bloustein School of Planning and Public Policy at Rutgers University. Kathe Newman is a professor in the Edward J. Bloustein School of Planning and Public Policy and director of the Ralph W. Voorhees Center for Civic Engagement at Rutgers University. Lauren Nolan is a Ph.D. candidate in the Edward J. Bloustein School of Planning and Public Policy at Rutgers University.

Appendix A: Findings From Research in Seven Communities

The research team is also conducting in-depth research in seven communities and have manually checked and mapped every federally assisted development with more care. Through this research, they have a further understanding of the possible limitations of our workflow. Following are their findings from those seven communities.

- In **Asbury Park**, one development—Asbury Park Village—was correctly placed in the wrong address. The public housing buildings data put it in Trenton, but it was an error in input data, and no amount of automation could have identified it. The research team also found that one development (Vita Gardens) covered two parcels rather than one.

- In **Millville**, Maurice View Plaza is a development, but the name was also used for several scattered sites. That fact poses no problem to locating the units in parcels because the data were at the building level. Holly Berry Court was on a nearby parcel but not the location of the housing units (as verified on Google Street View). The address given, 1100 Holly Berry Lane, is the location where the geocoders initially placed the units, but it was not the correct place. It is also a match in the parcel data. This discrepancy, like Asbury Park Village, is a case of a correct geocoder on an incorrect address.

- In **Montclair**, one sprawling (multiparcel) development had already been identified in the research team's previous processes.

- In **Passaic**, the research team found that one development (Chestnut Homes) covered two parcels rather than one.

- In **Phillipsburg**, what was called Heckman House was actually a number of developments with different names. This fact had no effect on placing the units in parcels because they were public housing and building-level data, but it would cause problems if someone wanted to identify overlap on the basis of the development name.

- In **Salem**, the research team thinks that **Salem Historic Homes** is actually a scattered-site project. The geocoding results are accurate based on the address provided.

- In **West New York**, one development (Horizon Heights) seems to be both in West New York and Union City on adjacent parcels; the research team had identified only the large parcel in West New York.

None of those issues affected the statewide overlap count. The sites were not examples of subsidy overlap, or the overlap was identified despite the inaccuracies.

References

An, Brian, Andrew Jakabovics, Jing Liu, Anthony W. Orlando, Seva Rodnyansky, Richard Voith, Sean Zielenbach, and Raphael W. Bostic. Forthcoming. "Factors Affecting Spillover Impacts of LIHTC Developments: An Analysis of Los Angeles," *Cityscape*.

Cohen, Adam. 2020. "FuzzyWuzzy: Fuzzy String Matching in Python." Python. https://chairnerd.seatgeek.com/fuzzywuzzy-fuzzy-string-matching-in-python/.

Deng, Lan. 2011. "The External Neighborhood Effects of Low-Income Housing Tax Credit Projects Built by Three Sectors," *Journal of Urban Affairs* 33 (2): 143–66.

Esri. n.d. "Geocoding and Geosearch." ArcGIS Online Help. https://doc.arcgis.com/en/arcgis-online/reference/geocode.htm.

Google Maps Platform. n.d. "Geocoding Request and Response | Geocoding API." Google for Developers. https://developers.google.com/maps/documentation/geocoding/requests-geocoding.

Harrison, Austin, Dan Immergluck, Jeff Ernsthausen, and Stephanie Earl. 2021. "Housing Stability, Evictions, and Subsidized Rental Properties: Evidence from Metro Atlanta, Georgia," *Housing Policy Debate* 31 (3–5): 411–24.

Lens, Michael C., Kyle Nelson, Ashley Gromis, and Yiwen Kuai. 2020. "The Neighborhood Context of Eviction in Southern California," *City & Community* 19 (4): 912–32.

New Jersey Housing Mortgage Finance Agency (NJHMFA). 2022. Low Income Tax Credit Properties. Data file.

New Jersey Office of GIS. 2022. "Parcels and MOD-IV Composite of NJ (download)." Shapefile. New Jersey Geographic Information Network (NJGIN) Open Data. https://njogis-newjersey.opendata.arcgis.com/documents/406cf6860390467d9f328ed19daa359d/about.

Prener, Christopher G., and Branson Fox. 2021. "Creating Open Source Composite Geocoders: Pitfalls and Opportunities," *Transactions in GIS* 25 (4): 1868–87. https://doi.org/10.1111/tgis.12741.

Reina, Vincent, and Michael Williams. 2012. "The Importance of Using Layered Data to Analyze Housing: The Case of the Subsidized Housing Information Project," *Cityscape* 14 (1): 215–22.

Rowan University School of Earth and Environment. 2023. "NJ Map." Parcel Explorer. https://www.njmap2.com/parcels/.

State of New Jersey Department of Community Affairs. 2022 (NJ DCA). List of Affordable Developments by County. Data file. https://www.nj.gov/dca/codes/publications/developments.shtml.

Taghavi, Lydia B. 2008. "HUD-Assisted Housing 101: Using 'A Picture of Subsidized Households: 2000'," *Cityscape* 10 (1): 211–20.

Teske, Daniel. 2014. "Geocoder Accuracy Ranking." In *Process Design for Natural Scientists: An Agile Model-Driven Approach*, edited by Anna-Lena Lamprecht and Tiziana Margaria, 161–74. Communications in Computer and Information Science (CCIS), vol. 500. Berlin, Heidelberg, Germany: Springer. https://doi.org/10.1007/978-3-662-45006-2_13.

United States Department of Housing and Urban Development (HUD). 2022a. Low-Income Housing Tax Credit (LIHTC): Property Level Data. Data file. https://www.huduser.gov/portal/datasets/lihtc/property.html#data.

———. 2022b. Multifamily Assistance and Section 8 Database. Data file. https://www.hud.gov/program_offices/housing/mfh/exp/mfhdiscl.

———. 2022c. Public Housing Buildings. Data file. https://hudgis-hud.opendata.arcgis.com/datasets/52a6a3a2ef1e4489837f97dcedaf8e27_0/explore?location=35.877966%2C-115.070600%2C4.47.

Wilson, Nicole E., Michael Hankinson, Asya Magazinnik, and Melissa Sands. 2023. "Inaccuracies in Low Income Housing Geocodes: When and Why They Matter," *Urban Affairs Review*, March. https://doi.org/10.1177/10780874231165767.

Who Owns Our Homes? Methods to Group and Unmask Anonymous Corporate Owners

Renz Torres
University of Florida and Shimberg Center for Housing Studies

Abstract

Institutional investors have been acquiring larger shares of single-family housing markets since the financial crisis of the late 2000s. Evidence shows that institutional single-family property owners negatively impact Black homeownership and evict tenants at higher rates than small landlords.

Identifying institutional ownership in housing is a challenge because it is difficult to assess who owns what. The necessary data are often unavailable or difficult to use. When the data are available, widespread use of limited liability companies and multilevel subsidiary structures make it difficult to identify the ultimate benefiting owner.

This article presents a graph-based analytical model designed to identify sums of properties held by groups of subsidiaries. Florida tax parcel and business registry data were transformed into a graph data structure to cluster owners based on directly or indirectly shared names, addresses, corporate filings, and shared officers. Applying this method to the Jacksonville, Florida, metropolitan area allowed the author to identify the biggest owners of single-family homes, explore aggregate spatial patterns of owners by size, analyze the expansion of two institutional investors' portfolios, and identify three regulatorily concentrated census tracts.

Although this method used single-family housing as a case study, it can easily analyze ownership in other housing sectors. The model can include heterogenous datasets, such as evictions, foreclosures, and financial flows, to identify large evictors, large owners of foreclosed properties, and central financial players.

Introduction

Since the financial crisis of the late 2000s, institutional investors—entities financed through methods such as private equity, real estate investment trusts, and pensions—have been acquiring more and more single-family homes (Colburn, Walter, and Pfeiffer, 2021; Ehrlich et al., 2023).

The impacts of institutional investor acquisition are variegated, and the long-term horizon of institutional investment remains unknown.

Identifying institutional acquisition of single-family housing is a challenge because it is difficult to assess who owns what. The tax parcel data necessary to assess ownership often are unavailable or unstandardized across jurisdictions (Dawes, Cook, and Helbig, 2006). When property ownership data are available, the widespread use of anonymous limited liability corporations and complex multilevel subsidiary corporate structure makes it difficult to truly identify owners (Prechel and Morris, 2010).

This article addresses this challenge with a method that augments models used by academics and activists. Using the single-family housing market of the Jacksonville, Florida, metropolitan statistical area as a test case, Florida tax parcel and business registry data are transformed into networks of owners and properties, allowing the author to identify the biggest single-family landlords of the metropolitan area, analyze acquisition patterns of two of the largest institutional investors, and identify collective patterns among owners of similar size.

Why Study Property Ownership?

A picture of single-family rental (SFR) property ownership is necessary to understand the effects of institutional investment on housing outcomes. For example, it has been found that tenants of large SFR corporate landlords face higher rates of eviction filings compared with small landlords (Seymour and Akers, 2021), even when controlling for foreclosure, tenant, property, and neighborhood characteristics (Raymond et al., 2018). Some find that large institutional investors invest in areas with higher proportions of people of color (Ehrlich et al., 2023), which has led to weakened homeownership opportunities for Black homebuyers (An, 2023). Others find that areas where institutional investors have properties do not have more people of color than the larger metropolitan area (Goodman et al., 2023). Gurun et al. (2022) found mixed results; institutional investors do leverage market power to extract higher rents, but they also internalize costs to improve neighborhood safety.

Ultimately, SFR's long-term horizon remains unclear. From a peak in 2021, purchases of existing homes by institutional SFR investors have declined (Malone, 2023), motivated by fewer cheap foreclosed homes and more owner-occupier demand (Goodman and Zinn, 2023). SFR is beginning to seize on the build-to-rent model, long-term ownership, and rental of newly constructed homes (Brill, Raco, and Ward, 2023; Nethercote, 2020), which may reduce competition for homebuyers but still result in bad outcomes for tenants.

How to Study Ownership?

Studying SFR ownership is difficult for three reasons. First, the SFR market is highly fragmented—nationwide, owners of less than 100 properties own 96.2 percent of the SFR stock (Goodman et al., 2023). Second, tax parcel data are not freely available nationwide; when they are, standardization across jurisdictions is lacking (Dawes, Cook, and Helbig, 2006). Lastly, multilevel subsidiary structures common in corporate ownership (Prechel and Morris, 2010) and the impenetrability of the limited liability corporation (Travis, 2019) make knowing who owns what quite difficult (Ashwood et al., 2022a). Shelton and Seymour (2024), for example, found that the three largest SFR owners in the Atlanta metropolitan area used over 190 aliases and 74 addresses.

Previous methodologies have attempted to chip away at this problem. Shelton and Seymour (2024) developed an iterative process using name and address similarity matching, identification of similarly named/derivative corporations, and corporate registration matching to find three landlords' aliases and addresses. An et al. (2022) used OpenRefine to cluster similar owner names and addresses using fingerprinting and n-gram methods. For example, they collapsed a list of 6,637 owner names into 1,663 clusters with fingerprinting and another list of 5,011 rows into 669 clusters using n-gram. Anti-Eviction Mapping Project (AEMP; n.d.) and JustFix (n.d.) both relied on a graph data model to connect parcel owners with each other, business registries, and rental registries. Using their model, AEMP (n.d.) found that 66 percent of multifamily units in Concord, California, were owned by the top 10 percent of landlords by size. JustFix (n.d.) found larger landlords owned most rent-regulated buildings and were connected to higher eviction rates.

The method proposed in the following section scales and automates Shelton and Seymour's (2024) method using the data models provided by AEMP and JustFix. Like Shelton and Seymour's, this process relies on an inductive-deductive blend of methodologies, starting from a property owner and matching successively larger groups of names and addresses. However, the following proposed process scales these methodologies across every single owner in an area. The underlying data model—the mechanism that powers the scaling—and the inclusion of business registry data are motivated by AEMP and JustFix. As an example, it is unclear whether JustFix's method applies the same inclusion of both indirectly shared addresses and names as Shelton and Seymour's. The following iterative process is made explicit using a graph model.

Methods

The publicly accessible tax parcel and business registry tables are transformed into a graph to identify, group, and manually label anonymous property owners. The underlying data model is most like AEMP's and JustFix's graph theory-inspired models. A graph is a mathematical object that models connections between objects through nodes and relationships. In graph data structures, table rows are transformed into nodes, and relationships are defined to connect these nodes. The advantage of this structure is the ease of detecting indirect relationships among nodes; with tables, the same query is a convoluted series of many-to-many joins.

From the tax parcel data, parcel nodes are created using fields intrinsic to the parcel and owner nodes based on unique name-address combinations. Separate name and address nodes are then created and related to the owner nodes. These relations assume that sharing an address means that different owners are the same entity or strongly related. In theory, this alone would be enough to create the analysis, but the use of business filing corporations as mailing addresses creates paths between unrelated entities. To disqualify these addresses from pathfinding, a list of registered agents is integrated from the business registry. String matching is used to reconnect corporations based on shared officers. This process is conducted at the scale of the single-family market to understand the local market power of SFR investors.

Data Sources

Unlike much of the United States, the state of Florida makes parcel and business registry data freely available. The Florida Department of Revenue (DOR) posts standardized yearly snapshots of county property appraiser parcels, sales, and geographical data online. The Florida Division of Corporations (Sunbiz) makes quarterly snapshots of its own corporate filing database accessible online.

Historical tax parcel data were sourced from the University of Florida Shimberg Center, which ultimately sources from the DOR's Name-Address-Legal dataset (2019, 2022). Each row in the data represents a tax parcel. Relevant fields include the parcel's physical address, identifier, homestead exemption, assigned state land use code, and the parcel owner's name and mailing address. Parcel geographies were also retrieved from the DOR.

Corporate filing information was sourced directly from Sunbiz (Florida Division of Corporations, 2023). Each row in the Sunbiz table represents a corporation's filing and contains a unique identifier, a name, officers, and a registered agent for the corporation. Using normalization techniques, the data were transformed into a filing table and officer and registered agent tables with joins to the filing table.

All string data fields across tax parcel and business registry data were preprocessed by conversion to uppercase, removal of superfluous whitespace, and removal of characters that were not alphanumeric, dashes, or spaces. Owner records with empty owner names or owner addresses were maintained, but the empty field was not used to match with other owners.

Procedures

Identifying Corporate Owners

Owner names were first classified as corporate in a stepwise manner based on certain criteria and text strings (exhibit 1). Noncorporate owners were identified first based on having a name of fewer than four characters or a marker indicating trust, estate, or government ownership. Corporate owners (any private, nontrust, and nonpersonal entity, including organizations like churches and nonprofits) were then classified based on the presence of a corporate marker. Remaining owners were classified as noncorporate by default.

Exhibit 1

Corporate and Noncorporate Markers (1 of 2)

Noncorporate Markers	Corporate Markers		
No Owner Name	*Has a Digit*	MORTGAGE	HOMEOWNERS
Name <3 characters	OF, AT, or BY	RENTAL	INVESTMENT
CONFIDENTIAL	LLC	MULTI	CONDO
REFERENCE ONLY	L L C	APARTMENT	PROPERT
REF ONLY	LP	VILLAS	MANAGEMENT
UNKNOWN	L P	REAL PROP	REALTY
ESTATE	LLLP	MARKET	JOINT VENTURES
LIFE ESTATE	L L L P	EQUITY	VENTURE

Exhibit 1

Corporate and Noncorporate Markers (2 of 2)

Noncorporate Markers	Corporate Markers		
TRUST, TRUS, TRU, TR	INC	EQUITIES	BORROWER
LIVING TRUST	I N C	REALTY	CHURCH
REVOCABLE LIVING TRUST	L C	RANCH CLUB	FELLOWSHIP
TRUST COMPANY	LTD	HOUSING	CHRIST
TRUST SERVICE	LIMITED	SERVICE	METHODIST
CUSTODIAN	PARTNERSHIP	COUNTRY CLUB	BAPTIST
401K	CORPORATION	OWNERS	MINISTRIES
IRA	CO	HOLDINGS	"FIRST "
EST OF	COMPANY	RESIDENTIAL	HOLY
TRS	COMPANIES	LEASING	MISSIONARY
CITY OF	ASSOCIATION	COMMUNITY	RENT
COUNTY OF	ASSOCIATES	DEVELOPMENT	HOME
STATE OF	BANK		
HOUSING AUTHORITY			
MANAGEMENT DISTRICT			

Notes: Italics indicate true/false criteria; CAPITALS indicate search strings.

CO = company. EST = estate. INC = incorporated. IRA = individual retirement account. LC = leasing commission. LLC = limited liability company. LLLP = limited liability limited partnership. LP = limited partnership. LTD = limited. TRS = trust.

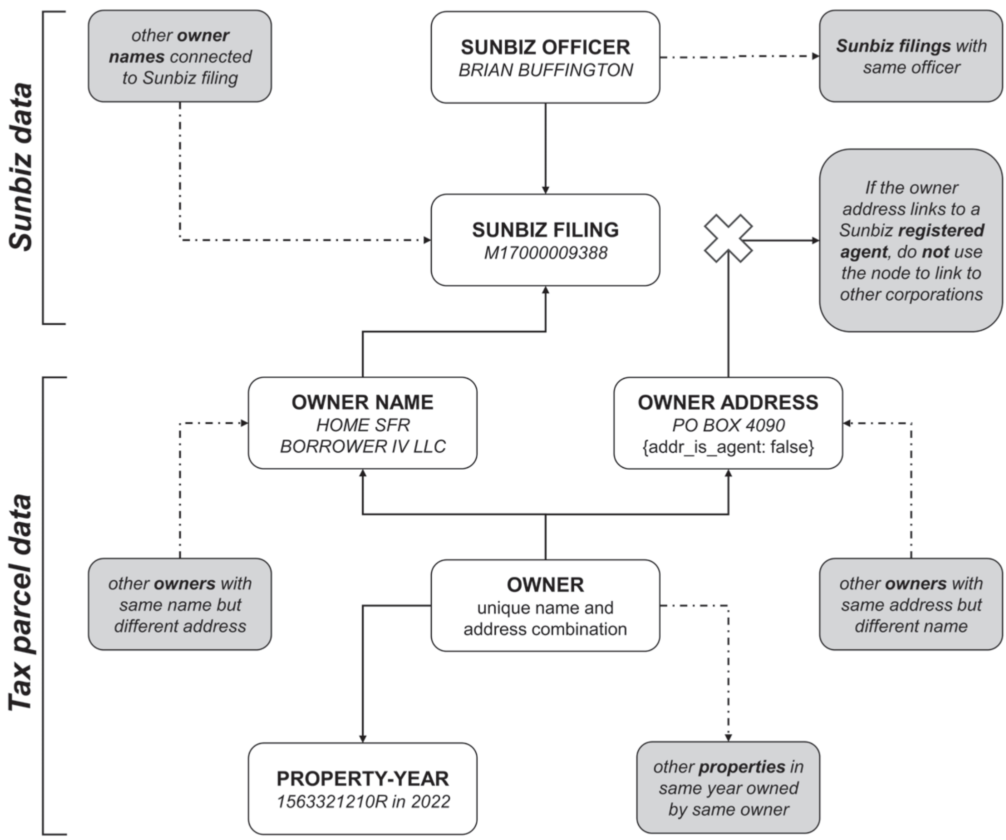
Source: Author's identification of property ownership markers

Grouping Owners and Properties

The bottom half of exhibit 2 describes transformations of the tax parcel data. The following processes were carried out in Neo4j (Neo4j, 2023), using Cypher queries to transform row-level data into relationships and store the nodes and relationships in a database. First, tax parcel data were restructured to lift parcel-specific information into parcel nodes. Combinations of a name and address were identified as "owners," and relationships were created between parcels and owners. The unique combinations were then split into separate owner name and address nodes, and relationships were created between the original combination and owner name and the combination and owner address. Owners became directly connected by sharing a name or an address and indirectly connected by tracing paths through related owners. In addition, using the Sunbiz data, each owner address was evaluated to see whether it was also a registered agent's address. Registered agents, such as filing companies and lawyers, provide administrative and tax services for unrelated companies; not accounting for registered agents results in these unrelated companies becoming connected in this analysis. Positively identified agent nodes were discounted from the analysis. Exhibit 2 provides a schematic of the graph data model.

Exhibit 2

Graph Data Model Connecting Data Sourced From Sunbiz (Top) and Tax Parcel Datasets (Bottom)



LLC = limited liability company. SFR = single-family rental.
Source: Author's depiction of proposed property ownership data model

Consider the example in exhibit 3, with eight owners, five unique owner names, and four unique addresses. Eight owner-to-name and eight owner-to-address relationships were created. Owners 1 and 2, 1 and 3, 3 and 4, 1 and 4, 4 and 5, and 5 and 6 were directly connected by shared names and addresses. Owners 1 and 6 became connected indirectly through 1 and 4's address, 4 and 5's name, and 5 and 6's address. Owner 8 was not connected to Owner 7 because their shared address was a registered agent address. Owners 1–7 form one corporate entity, and the sum of their holdings becomes the corporate entity's holdings.

Exhibit 3

Example of Unique Owner-Name Combinations

ID	Name	Address
1	A LLC	0 MAIN ST, WILMINGTON
2	A LLC	1000 US 90, JACKSONVILLE
3	B LLC	0 MAIN ST, WILMINGTON
4	C LLLP	0 MAIN ST, WILMINGTON
5	C LLLP	1 US 301, BALDWIN
6	D LP	1 US 301, BALDWIN
7	D LP	400 AGENT ST *(registered agent)*
8	E LLC	400 AGENT ST *(registered agent)*

LLC = limited liability company. LLLP = limited liability limited partnership. LP = limited partnership.
Source: Author's example using fictional names and addresses

As shown in the top half of exhibit 2, owner names were matched to corporation filing nodes by whether the owner name was a substring of or equal to the other name. Corporate filings became connected through shared officers. By tracing a path through shared officers, owners disconnected by registered agent addresses can become reconnected through the Sunbiz data nodes. For example, in exhibit 3, if Owners 7 and 8 shared an officer, they would become connected.

To understand which owners are related directly or indirectly, the Weakly Connected Components (WCC) algorithm of the Graph Data Science library was implemented (Neo4j, 2023). WCC groups all nodes in a graph based on whether a path exists between nodes, decomposing the graph into intra-connected subgraphs. The resulting connected group of owners is referred to as a "sibling group." For corporations, a sibling group corresponds to a group of subsidiaries.

After a first pass through WCC, the results were manually inspected for over-clustering—the connection of unrelated entities—and under-clustering—the disconnection of related entities (see the Discussion section for more details). For the 50 biggest owners, the connected owner results were inspected, nodes were manually connected or disconnected, and WCC was run again. This process was iterated for the new biggest 50 owners three times or until convergence.

WCC can only group owners together—neither business registry nor parcel data contain information about corporate hierarchy. Thus, WCC cannot name the ultimate "owner" of a group of owners. For the largest five landlords and select locally important landlords, the parent of their sibling groups was manually identified by cross-referencing names and addresses with news articles, company websites, and/or Electronic Data Gathering, Analysis, and Retrieval filings that linked names and addresses to parent corporate entities.

Identifying Nonowner-Occupied Housing (NOOH)

The properties under analysis are nonowner-occupied housing (NOOH). This designation includes corporate SFR, noncorporate SFR, second or vacation homes, and single properties of multiple parcels. Differentiating between the latter three through parcel data alone is difficult, so all categories are included under NOOH. The status of a property as owner-occupied, NOOH, or unknown was defined through the following stepwise criteria. The first true criterion determines the status.

1. NOOH: owned by a corporate sibling group (where at least one owner is identified as corporate), the owner state is not in the state of the analysis area, the owner city is not any city in the study area, or the owner ZIP Code does not equal the parcel ZIP Code AND the owner address is not a post office box.
2. Owner-occupied: owner address is a substring of or equal to the physical address (or vice versa), the property is homesteaded, or the owner address and physical address have the same physical location (this occurs frequently when one street may have multiple representations, e.g., in Duval County, Florida: Beaver Street and U.S. 90).
3. Unknown: owner address is a post office box.
4. NOOH: Levenshtein string distance between the owner and physical addresses is greater than 7, or the entity owns four or more properties across Duval County and the property is not homesteaded.
5. Unknown: by default.

Quantifying Neighborhood Concentration

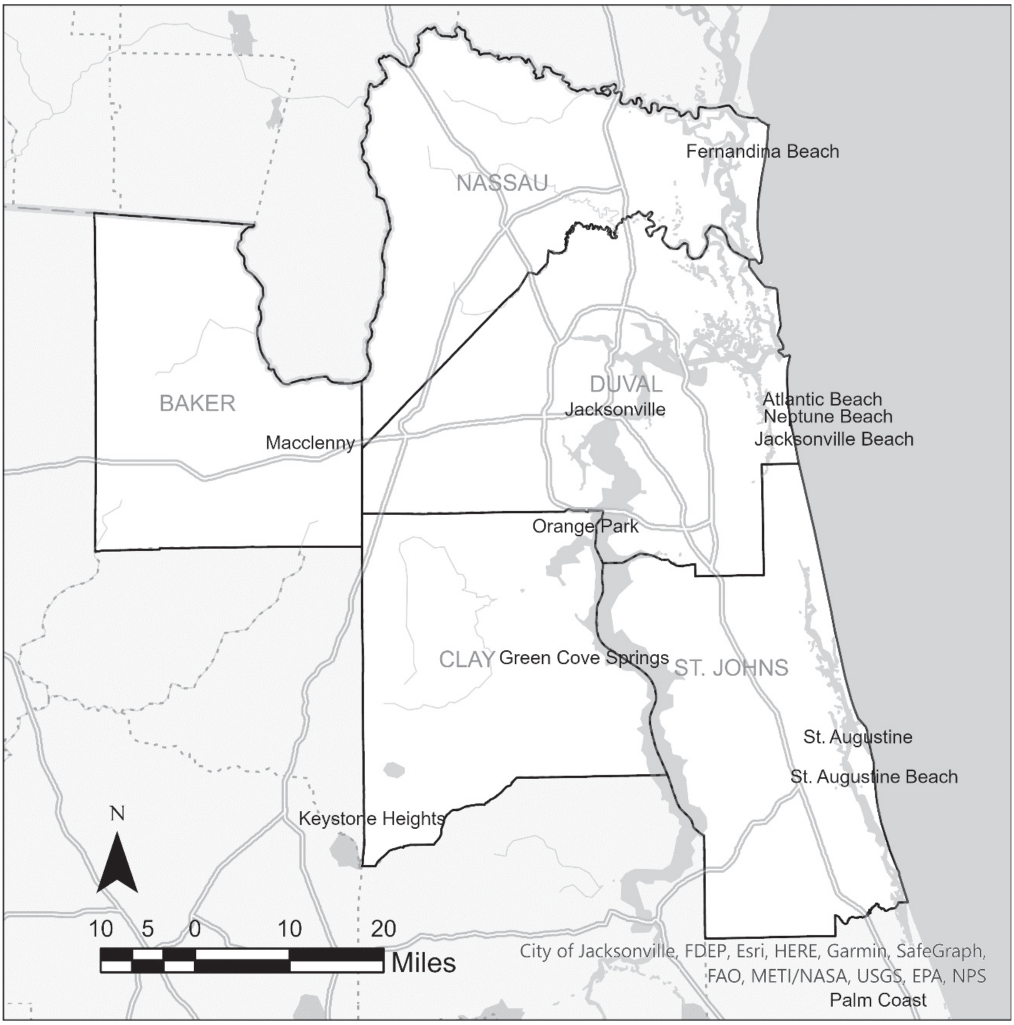
Census tract-level concentration was calculated using the Herfindahl-Hirschman Index (HHI) across NOOH properties only. The Index is defined by the formula $\sum_{i=1}^{n} s_i^2$, where s_i is the owner's NOOH stock share as a percentage of the total NOOH stock and n is the total number of owners (Rhoades, 1993). The HHI ranges from 0 (an infinitely competitive market) to 10,000 (a monopoly). The resultant HHI was classified according to U.S. Department of Justice regulatory guidelines on concentration (Tapp and Peiser, 2023); an HHI of 2,500 or higher indicates highly concentrated markets, and an HHI of 1,500–2,499 indicates moderate concentration.

Case Study: Jacksonville, Florida

This analysis uses the Jacksonville, Florida, single-family residential market (Baker, Clay, Duval, Nassau, and St. Johns Counties; exhibit 4) in the years 2019 and 2022 as a case study. Like other Sunbelt metropolitan areas after the foreclosure crisis between 2007 and 2010, Jacksonville's single-family market has had intense investor encroachment and high growth of single-family rental and build-to-rent markets (City of Jacksonville, 2022; Fields and Vergerio, 2022; Hughes and *Action News Jax*, 2021; Immergluck, 2018; JAX Rental Housing Project, n.d.).

Exhibit 4

Map of Five-County Study Area and Major Cities



Source: Author's depiction of Jacksonville metropolitan area, Esri basemap

The following sections explore findings informed by a comprehensive who-owns-what analysis: the geographical distribution of NOOH as a share of all single-family homes, an identification of the biggest single-family owners, aggregate distributions of owner activity by size of owner, geographical patterns of expansion of two large SFR landlords, and NOOH ownership concentration by tract.

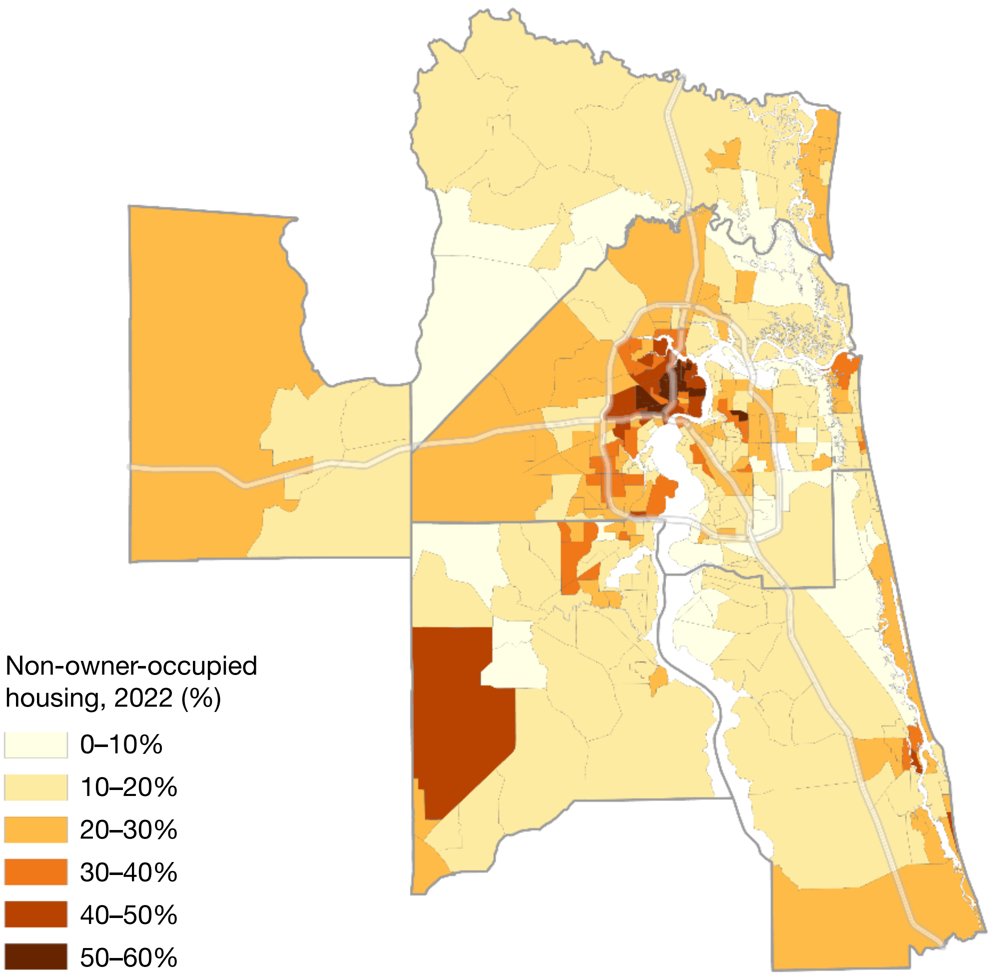
Where Is Nonowner-Occupied Housing?

In 2022, there were 468,733 single-family parcels across the Jacksonville metropolitan area. Of those parcels, 368,826 (78.7 percent) were classified as owner-occupied housing, 96,137 (20.5 percent) as nonowner-occupied housing (NOOH), and 3,770 (0.8 percent) of unknown status.

When aggregated by census tract, the areas with the highest percentages of NOOH were in the urban core of Jacksonville, with some tracts as high as 60 percent NOOH (exhibit 5). Medium percentages of NOOH prevailed across southwest and west Jacksonville, Orange Park, St. Augustine, and beach areas in Nassau and St. Johns Counties. Lower proportions of NOOH prevailed across Jacksonville's Southside neighborhood and most other counties' inland areas.

Exhibit 5

Map of Nonowner-Occupied Housing in 2022 by Percentage and Census Tract



Source: Author's tabulation of tax parcel and business registry data

Who Were the Biggest Owners Identified by WCC?

In 2022, 429,133 unique owners owned 468,733 single-family homes. Among these owners, there were 424,731 unique owner names and 404,531 unique owner addresses. The Weakly Connected Components (WCC) algorithm condensed these owners into 408,615 sibling groups, and 9,281 were identified as corporate. Before WCC, the largest number of properties owned by a single owner was 1,205. After WCC clustering, this number increased to 4,863. About 13.8 percent of 2022 owner addresses were discounted as agents.

The biggest owners by number of properties were Progress Residential, AMH (previously American Homes 4 Rent), Invitation Homes, FirstKey Homes, and Amherst Group. These owners were represented by myriad names and addresses (exhibit 6). WCC only clusters owners together; the algorithm cannot identify the ultimate owner of the subsidiaries. Therefore, the groups in exhibit 6 were manually labeled using information from 10-K filings for publicly traded AMH and Invitation Homes and corporate websites for the balance.

Exhibit 6

Weakly Connected Components Algorithm Summary for the Top Five Largest Owners by Properties Owned, 2022

Parent	Properties	Owners	Names	Addresses	Filings
Progress Residential	4,863	151	89	35	64
AMH	2,965	126	62	39	17
Invitation Homes	1,999	92	52	19	8
FirstKey Homes	1,688	162	62	43	14
Amherst Group	1,424	152	67	28	19

Note: The term Owners *refers to unique owner name-address combinations.*
Source: Author's tabulation of tax parcel and business registry data

Cross Validation

The estimates for AMH and Invitation Homes concurred with their 10-K filings within 4 percent. For AMH, there were an estimated 2,965 properties as of October 2022; AMH's December 2022 10-K showed 2,891 Jacksonville metropolitan area properties (American Homes 4 Rent, 2022). For Invitation Homes, there were an estimated 1,999 properties; Invitation's 2022 10-K listed 1,928. (Invitation Homes, 2022). Possible causes of the differences include false positives in the data model and property sales in the lag between October and December.

Do Investors Target Different Areas Based on Their Scale?

Sibling groups were classified by portfolio size using a classification system adapted from CoreLogic, Inc. (Goodman et al., 2023): extra small owners as 1–3 properties owned, small as 4–9, medium as 10–99, large as 100–999, and mega as 1,000 and more. Extra small owners continue to dominate single-family NOOH, but the large presence of mega investors still looms at 14.9-percent ownership of all NOOH, almost the same number as all medium and large investors' holdings combined (exhibit 7).

Exhibit 7

Descriptions of Size Classifications for Investors and Summaries of Properties Owned, 2022

Category	Properties Owned by Single Owner	Sibling Groups	Total Class 2022 Properties	% Total 2022 NOOH Stock
Extra Small	1–3	47,809	54,545	56.7
Small	4–9	2,522	12,473	13.0
Medium	10–99	438	9,284	9.7
Large	100–999	23	5,556	5.8
Mega	1000+	6	14,279	14.9
Total		*50,798*	*96,137*	*100*

NOOH = nonowner-occupied housing.

Source: Author's definition of size classifications, tabulation of tax parcel and business registry data

Between 2019 and 2022, there was a dramatic shift in NOOH ownership from extra small owners to mega investors. Collectively, mega investors gained ownership of 9.8 percent of the NOOH stock, whereas extra small entities lost 16.1 percent of the same stock in the same period (exhibit 8). Small, medium, and large owners saw less appreciable gains.

Exhibit 8

Changes in Percent of Market Owned in Aggregate by Size of Investor, 2019–22

Category	Properties Owned by Single Owner	% Total 2019 NOOH Stock	% Total 2022 NOOH Stock	% Change 2019 to 2022
Extra Small	1–3	72.8	56.7	–16.1
Small	4–9	10.9	13.0	+2.1
Medium	10–99	7.7	9.7	+2.0
Large	100–999	3.5	5.8	+2.3
Mega	1000+	5.1	14.9	+9.8

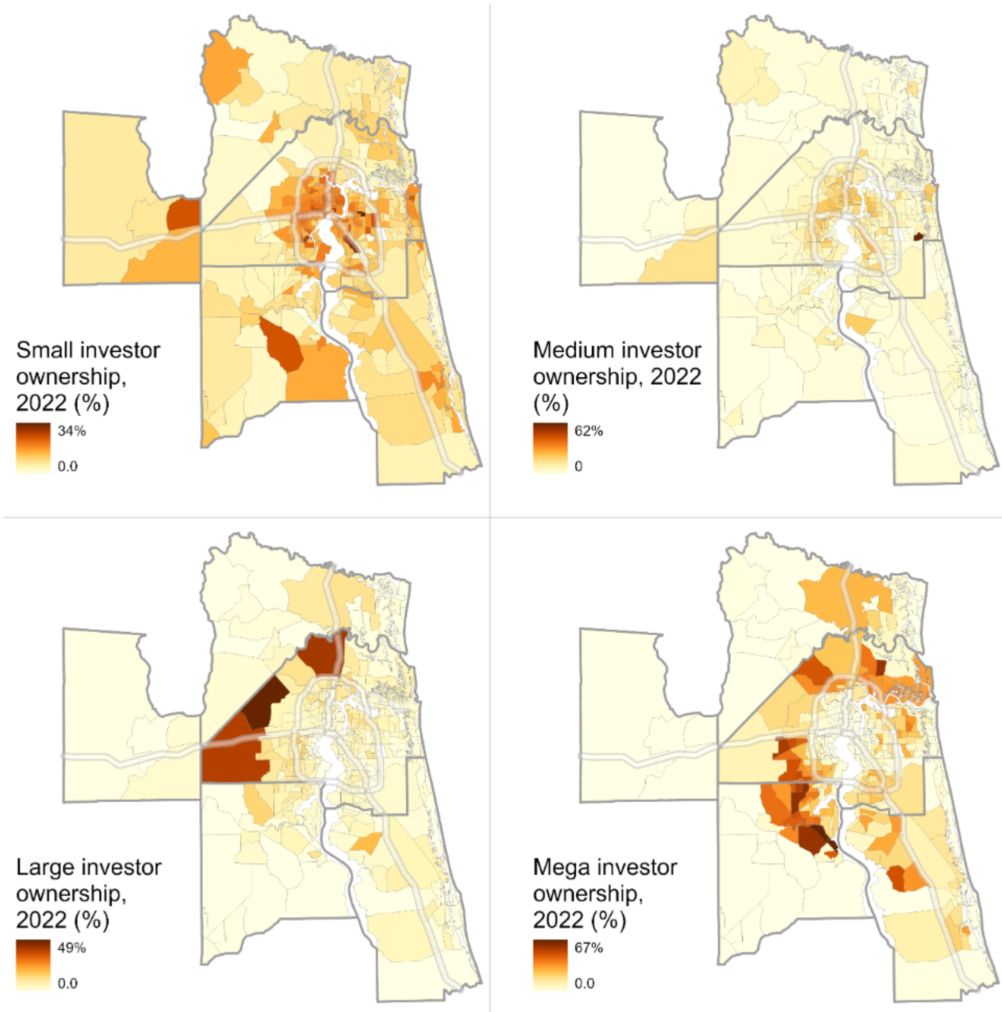
NOOH = nonowner-occupied housing.

Source: Author's tabulation of tax parcel and business registry data

Investors' geographical distribution depended on their size. Small investors clustered around Jacksonville's urban core and inner-ring suburbs, with minor pockets around Jacksonville beaches and St. Augustine (exhibit 9, top left). Medium investors also targeted the urban core, but that activity was overshadowed by intense ownership in a southeast Duval census tract (exhibit 9, top right). Large and mega investors formed a ring outside of Interstate 295 in north, west, and southwest Duval County, northeast Clay County surrounding Orange Park, and inland north St. Johns County (exhibit 9, bottom left and right, respectively).

Exhibit 9

Ownership of Nonowner-Occupied Housing Properties by Size Category of Investor, 2022



Source: Author's tabulation of tax parcel and business registry data

By weighted average, small, medium, and large investors were active in census tracts[1] with higher proportions of people of color (not "White and Non-Hispanic"), as tabulated by the 2021 American Community Survey (U.S. Census Bureau, 2021). In contrast, mega investors were active in tracts with roughly the same proportion of people of color as the whole NOOH market (exhibit 10).

[1] Note the offset between the year of the American Community Survey data and the year of the parcel dataset, which is due to unavailability of 2022 American Community Survey data at time of writing.

Exhibit 10

Weighted Percent Average of Census Tracts With Investor Activity by Size of Investor, 2022 Nonowner-Occupied Housing and 2021 American Community Survey Data

Category	Property Range	% People of Color in Census Tract (2021), Weighted Average	% People of Color in Census Tract (2021), Standard Deviation	% Difference From Whole 2022 NOOH Market
Extra Small	1–3	40.6	27.4	–4.3
Small	4–9	51.9	27.6	+7.0
Medium	10–99	60.0	25.8	+15.1
Large	100–999	50.7	23.4	+5.8
Mega	1000+	43.6	19.7	–1.3
All NOOH		*44.9*		

NOOH = nonowner-occupied housing.

Source: Author's tabulation of tax parcel, business registry, and 2021 American Community Survey data

Investor Acquisition Patterns

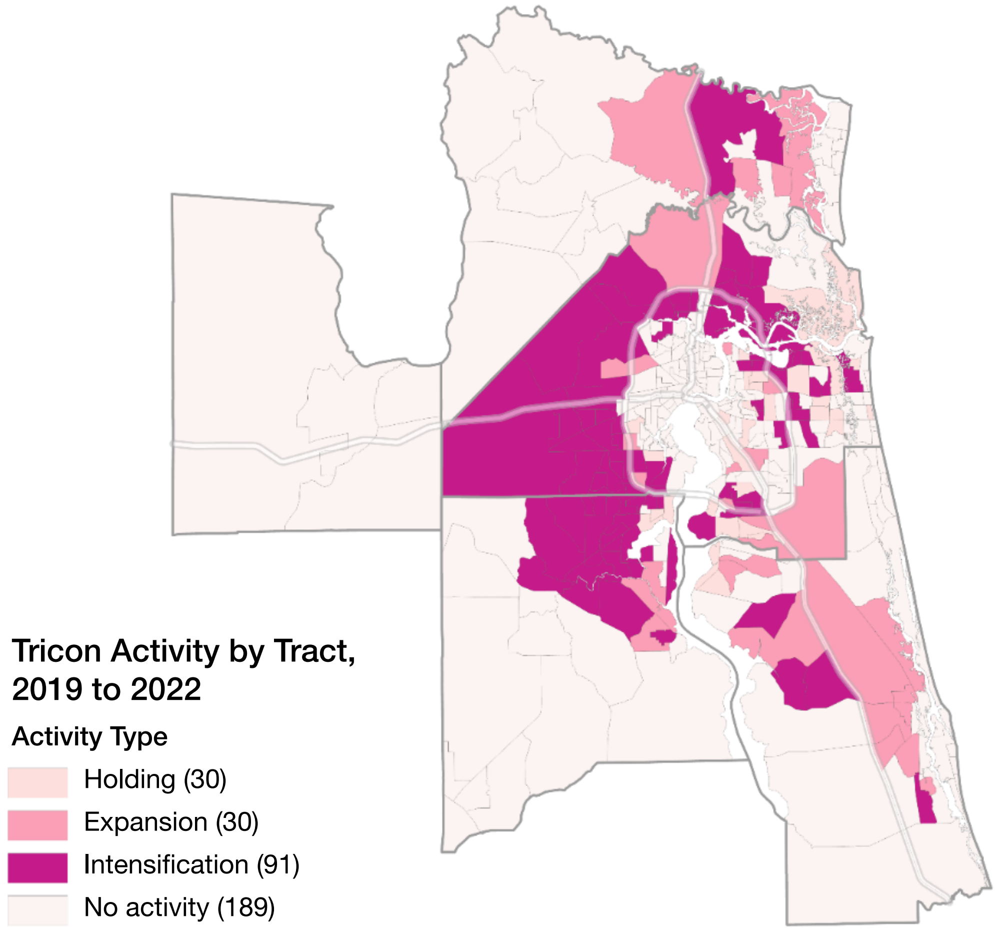
Using this analysis, one can also investigate the behavior of individual investors, such as their geographical distribution and change over time and from whom they acquire properties. This section explores two nationwide SFR operators that rapidly increased their Jacksonville area portfolios between 2019 and 2022: Tricon Residential and Progress Residential.

Tricon Residential was the fifth largest owner of Jacksonville metropolitan area single-family homes in 2019 and sixth largest in 2022. Despite the drop in ranking, Tricon increased its Jacksonville portfolio from 549 to 1,340 homes in 2022. It sold 3 homes and gained 794. Of its acquisitions, 53 percent were previously owner-occupied in 2019, 26 percent were NOOH, and 21 percent were built since 2019. About 80 percent of its NOOH acquisitions were from owners of fewer than four properties.

Tricon generally intensified its presence in neighborhoods rather than expanding to other census tracts. Of the 151 census tracts where Tricon had properties in 2022, 91 already had properties in 2019, and Tricon added more (exhibit 11; intensification). Only 30 census tracts, mostly in outlying Duval County and suburban counties, were new areas (expansion).

Exhibit 11

Changes in Tricon Residential Activity by Census Tract, 2019 to 2022



Notes: "Holding" means no changes in property numbers. "Expansion" means census tracts with parcels acquired where none existed in 2019. "Intensification" means census tracts with preexisting acquisitions to which more properties were added.

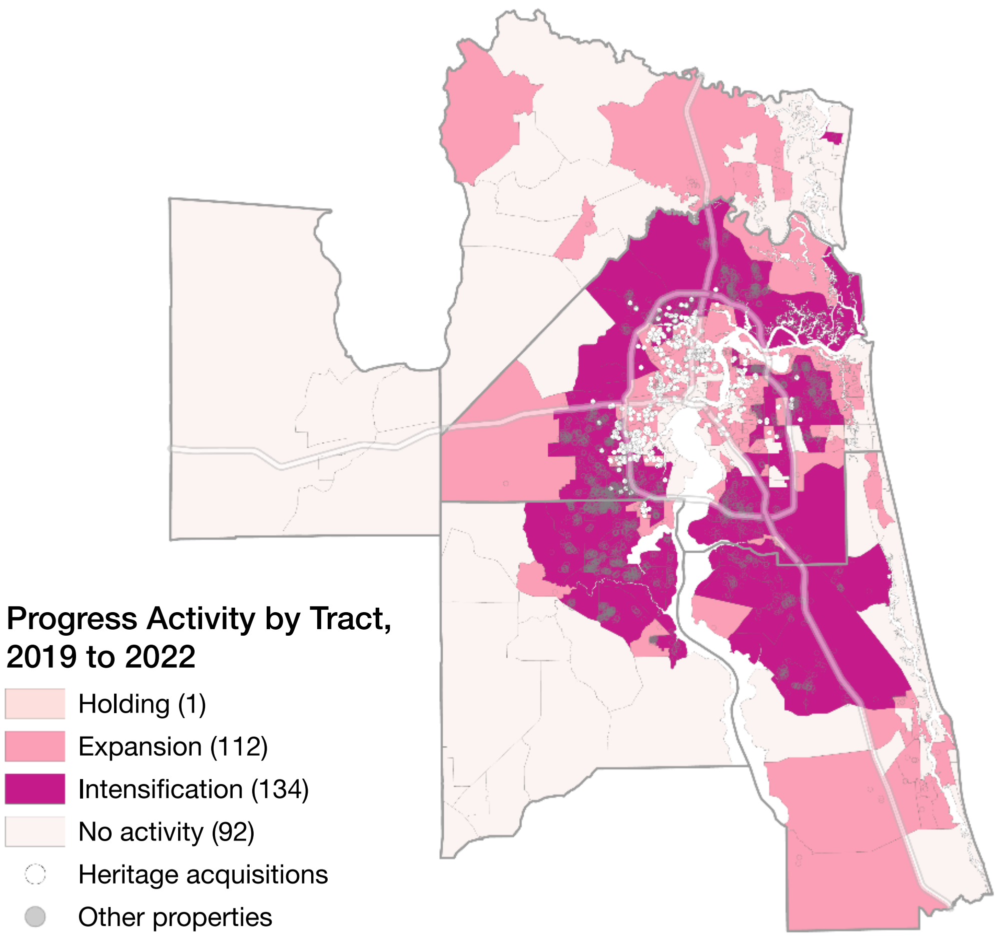
Source: Author's tabulation of tax parcel and business registry data

Progress Residential was the biggest owner of Jacksonville metropolitan area single-family homes in 2022, growing from third largest in 2019. Progress owned 1,845 homes in 2019; it added 3,021 homes and sold only 3 for a total of 4,863 homes in their 2022 Jacksonville portfolio. Forty-nine percent of their acquisitions were owner-occupied in 2019, 47 percent were NOOH in 2019, and 4 percent were housing built since 2019. Compared with Tricon, only about 32 percent of the NOOH acquisitions by Progress were from owners of fewer than four properties; 527 acquisitions of previously built housing were from Heritage Holdings, a once locally large SFR player focused on Jacksonville's urban core, as a portfolio acquisition; and 184 were from Havenbrook Homes as part of another merger.

Unlike Tricon, Progress greatly expanded its presence to new areas. Of the 247 census tracts it was invested in during 2022, 112 (47 percent) were new tracts to Progress (exhibit 12). Along with the outward expansion by Tricon, Progress expanded inward into Jacksonville's urban core, atypical compared with other mega investors. The acquisition of Heritage properties allowed Progress to springboard into a new racial and geographical market, Jacksonville's mostly Black inner city.

Exhibit 12

Changes in Progress Residential Activity, by Census Tract, 2019 to 2022



Notes: "Holding" means no changes in property numbers. "Expansion" means census tracts with parcels acquired where none existed in 2019. "Intensification" means census tracts with preexisting acquisitions to which more properties were added. Opaque dots indicate portfolio acquisitions from Heritage Holdings.

Source: Author's tabulation of tax parcel and business registry data

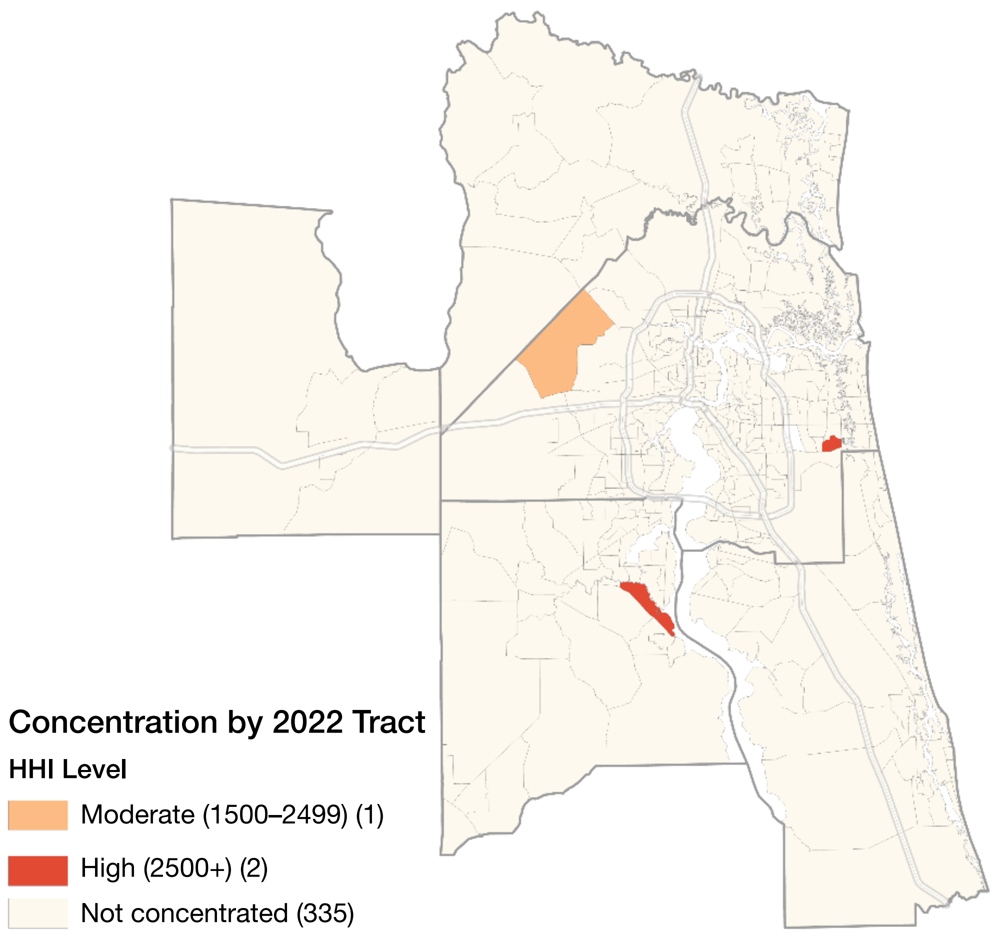
Are There Locally Concentrated Single-Family Rental Markets?

In 2022, as measured by the Herfindahl-Hirschman Index, two tracts met regulatory definitions of highly concentrated NOOH markets, one in southeast Duval County and one in northeast Clay

County (exhibit 13). The Duval County tract had 105 NOOH properties, 62 percent of which were owned by Cypress Garden Homes, LLC. This tract is also the one with high medium investor ownership in exhibit 9 (top right). Cypress Garden is a subsidiary of Healthpeak Properties, Inc. (Healthpeak Properties, Inc., 2023), an investor in medical real estate. The Clay County tract had 143 NOOH properties, 61 percent of which were owned by AMH, and generally showed high ownership by mega investors (exhibit 9, bottom right). One tract in northwest Duval County met regulatory definitions of moderate concentration. That tract had high large investor ownership (exhibit 9, bottom left) and 190 single-family NOOH properties, 89 (47 percent) of which were owned by Upward America Southeast Property Owner, a build-to-rent company formed by homebuilder Lennar Corporation (Lennar Corporation, 2021). AMH owned another 31 (16 percent).

Exhibit 13

Single-Family Nonowner-Occupied Housing Market Concentration Levels by Tract, 2022



HHI = Herfindahl-Hirschman Index.

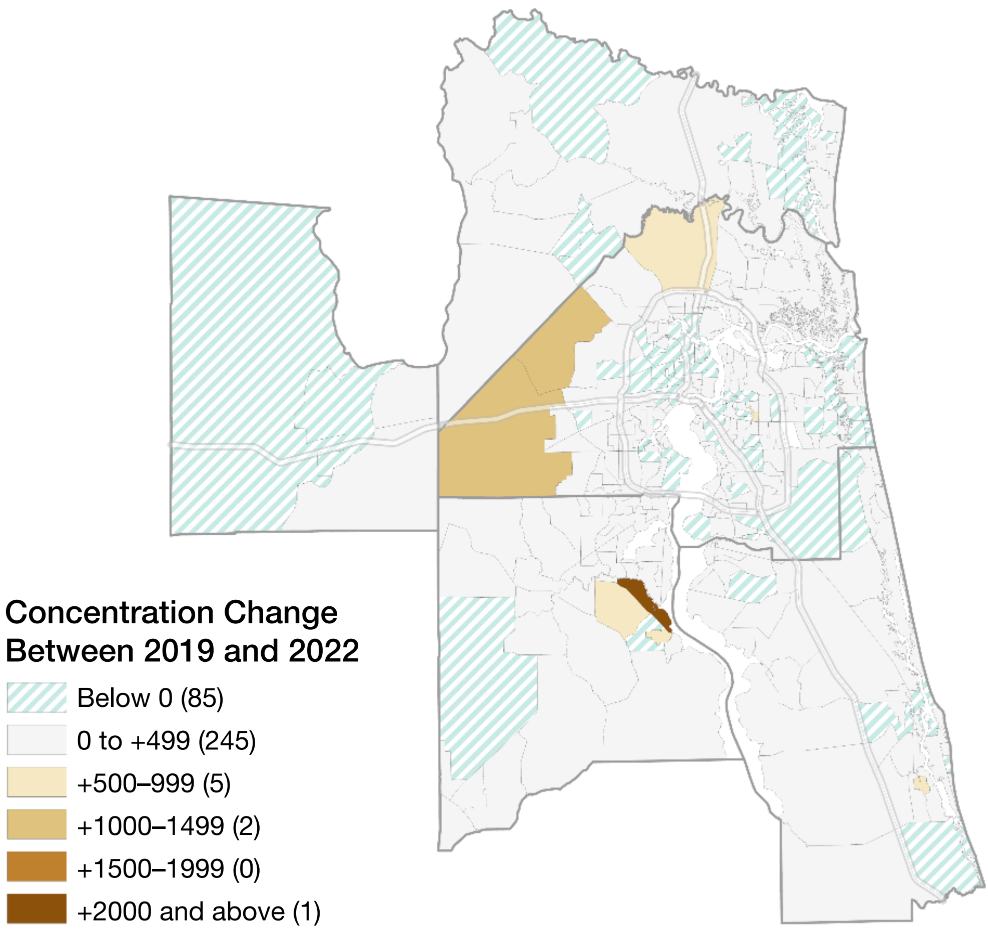
Source: Author's tabulation of tax parcel and business registry data

In all three tracts, most properties owned by the dominant owners were developed at the same time and intended for rent. In the first case, Cypress Garden's homes were built for rent around 1995. Many of those properties entered owner-occupied markets but have since returned to Cypress Garden for long-term rental. In contrast, AMH and Upward America's properties were developed as recently as 2020 solely for rent, and those properties are truer build-to-rent exemplars.

Between 2019 and 2022, a handful of tracts in northeast Clay and west Duval Counties saw rises in HHI Index values of 1,000 or more (exhibit 14). The highest increase was in the Clay County tract mentioned previously, primarily due to the construction of AMH build-to-rent community Black Creek Village in Green Cove Springs. The second highest jump in HHI Index values was in the Upward America Duval County tract mentioned previously. About one-fourth of all tracts, mostly in the urban core, saw their HHI values decrease.

Exhibit 14

Changes in Single-Family Nonowner-Occupied Housing Market Concentration Levels by Tract, 2019 to 2022



Source: Author's tabulation of tax parcel and business registry data

Discussion

The graph-based method demonstrated in this article allowed exploration of property ownership through a transformation of tax parcel and business registry data. Graph paths were used through shared owner names, owner addresses, and business officers to group owners and their properties. With these results, several metrics and methods were derived to study investor activity in aggregate and by entity and to identify market concentration. These metrics are comparable with similar literature findings.

Although single-family nonowner-occupied housing (NOOH) properties were clustered in Jacksonville's urban core, large and mega investor activity was highest in Jacksonville's outer-ring suburbs, similar to findings in Atlanta (An, 2023). Mega investors made up 14.9 percent of the 2022 NOOH market, an estimate in line with Ehrlich et al.'s (2023) 16 percent for Jacksonville in 2022. Mega investors did not have properties where there were more people of color, as in Goodman et al.'s (2023) finding. In contrast, small, medium, and large investors were more active in Jacksonville neighborhoods of color.

Mega investors Tricon Residential and Progress Residential had differing expansion strategies; both intensified their market power and geographically expanded their portfolios outward through acquisitions from owner-occupiers and build-to-rent, but Progress used a portfolio acquisition from a local SFR institutional investor to expand inward toward the urban core.

The analysis identified three tracts with single-family NOOH markets concentrated mostly due to the impact of build-to-rent housing developments (as derived from parcel data on year of construction). It may be worthwhile to separate build-to-rent from other SFR in analyzing concentration; build-to-rent is more like multifamily than SFR converted from owner-occupied housing because prospective single-family homebuyers do not compete with build-to-rent in a market.

Limitations

Known problems include over-clustering and under-clustering. Over-clustering happens most commonly with errors such as an owner name being changed after a sale but not the address, leading most often to owner-occupiers or small landlords becoming clustered. The data model also assumes that corporations clustered at a non-agent address are in fact the same entity or at least strongly related, which may not be true. Under-clustering is generally caused by spelling errors and the agent address discounting step. For example, two spellings of the same owner name remain unconnected if they do not share an address. Discounting agent addresses separates unrelated entities using the same business agent, but related entities also become disconnected. Business registry data may not reconnect these truly related entities. Over- and under-clustering were mitigated by manually connecting nodes and discounting relationships; anyone with time available can carry out a similar manual inspection. Spelling errors can be mitigated by employing text clustering methods suggested by An et al. (2022).

Underlying model assumptions may be inapplicable to other jurisdictions because of data availability and local practices. Florida is unique in the standardization and free availability of statewide tax parcel data, so this analysis can be carried out statewide. Although less data may be available elsewhere, this fact does not render the analysis invalid but just limits its scope. To overcome this circumstance, commercial nationwide datasets can be purchased at cost from services like Regrid, and alternatives like web scraping are possible (St-Hilaire, Brunila, and Wachsmuth, 2023). Jurisdictions may differ on registration requirements for businesses to own property; these requirements have bearing on the business registry data's completeness. Florida Statute 607.1501 requires non-Floridian businesses to register with Sunbiz to rent property, and therefore, all corporate rental owners in Florida are represented in Sunbiz. Even if this regulation exists in other states, business registries may not always be available. Where this regulation does not exist, multiple states' business registries may have to be used.

Directions for Further Research

One way to expand property ownership with this methodology is exploring other real estate sectors. This article focused on SFR because it is perhaps the most fragmented of all real estate sectors, but there is no reason multifamily homes, mobile homes, and other properties cannot be analyzed in the same way. For example, single-family build-to-rent might be better analyzed with other multifamily homes rather than single-family homes.

The graph data structure, unlike tables, also lends itself well to linking with other datasets. For example, if a parcel sales dataset indicates distressed sales, one can relate these sales to the parcel to understand when, where, and how investors are acquiring previously distressed housing and how that housing became distressed. Tabular models may require many-to-many joins to explore the same relationship. Evictions, liens, code violations, and other administrative data sources can be linked to explore similar relationships.

Lastly, adding corporate family trees to the analysis can allow one to quickly identify the parents of subsidiaries. With this information, one can further attach ownership and creditor relationships to determine financial relationships between entities. For example, one could identify central shareholders and creditors in housing (Ashwood et al., 2022b) and anti-competitive practices (Tapp and Peiser, 2023). Ultimately, all these directions would be useful to tenants, policymakers, regulators, and other housing stakeholders.

Acknowledgments

I thank the Shimberg Center for Housing Studies for their work in archiving Floridian public data on housing. I also thank Anne Ray of the Shimberg Center and Dr. Renée Tapp of the University of Florida's Urban and Regional Planning department for their guidance in this research.

Author

Renz Torres is a graduate student of urban and regional planning at the University of Florida and a researcher at the Shimberg Center for Housing Studies.

References

American Homes 4 Rent. 2022. "Form 10-K February 24th, 2023." https://www.sec.gov/Archives/edgar/data/1716558/000156240123000032/amh-20221231.htm.

An, Brian Y. 2023. "The Influence of Institutional Single-Family Rental Investors on Homeownership: Who Gets Targeted and Pushed Out of the Local Market?" *Journal of Planning Education and Research*. https://doi.org/10.1177/0739456X231176072.

An, Brian, Andrew Jakabovics, Anthony W. Orlando, and Seva Rodnyansky. 2022. "Who Owns Urban America? A Methodology for Identifying Real Estate Owners," *SSRN Electronic Journal*. https://www.researchgate.net/publication/362183221_Who_Owns_Urban_America_A_Methodology_for_Identifying_Real_Estate_Owners.

Anti-Eviction Mapping Project (AEMP). n.d. "Evictorbook Methodology." https://evictorbook.com/methodology.

Ashwood, Loka, John Canfield, Madeleine Fairbairn, and Kathryn De Master. 2022a. "What Owns the Land: The Corporate Organization of Farmland Investment," *The Journal of Peasant Studies* 49 (2): 233–262. https://doi.org/10.1080/03066150.2020.1786813.

Ashwood, Loka, Andy Pilny, John Canfield, Mariyam Jamila, and Ryan Thomson. 2022b. "From Big Ag to Big Finance: A Market Network Approach to Power in Agriculture," *Agriculture and Human Values* 39 (4): 1421–1434. https://doi.org/10.1007/s10460-022-10332-3.

Brill, Frances, Mike Raco, and Callum Ward. 2023. "Anticipating Demand Shocks: Patient Capital and the Supply of Housing," *European Urban and Regional Studies* 30 (1): 50–65. https://doi.org/10.1177/09697764211069837.

City of Jacksonville. 2022. *Special Committee on Critical Quality of Life Issues*. Jacksonville, FL. http://apps2.coj.net/City_Council_Public_Notices_Repository/COJ%20CQLI%20Final%20Report%2012.16.22.pdf.

Colburn, Gregg, Rebecca J. Walter, and Deirdre Pfeiffer. 2021. "Capitalizing on Collapse: An Analysis of Institutional Single-Family Rental Investors," *Urban Affairs Review* 57 (6): 1590–1625. https://doi.org/10.1177/1078087420922910.

Dawes, S.S., M.E. Cook, and N. Helbig. 2006. "Challenges of Treating Information as a Public Resource: The Case of Parcel Data." In *Proceedings of the 39th Annual Hawaii International Conference on System Sciences* (HICSS'06), 81a–81a. Kauai, HI: IEEE. https://doi.org/10.1109/HICSS.2006.83.

Ehrlich, Cameron, Tim McDonald, L. David Vertz, and Sage Computing. 2023. "Editor's Note," *Evidence Matters*, Winter.

Fields, Desiree, and Manon Vergerio. 2022. "Corporate Landlords and Market Power: What Does the Single-Family Rental Boom Mean for Our Housing Future?" UC Berkeley. https://escholarship.org/uc/item/07d6445s.

Florida Department of Revenue (DOR). 2022. "Assessment Roll: Name-Address-Legal." Shimberg Center. https://floridarevenue.com/property/Pages/DataPortal_RequestAssessmentRollGISData.aspx.

———. 2019. "Assessment Roll: Name-Address-Legal." Shimberg Center. https://floridarevenue.com/property/Pages/DataPortal_RequestAssessmentRollGISData.aspx.

Florida Division of Corporations. 2023. "Corporate Data File." Sunbiz. https://dos.myflorida.com/sunbiz/other-services/data-downloads/.

Goodman, Laurie, and Amalie Zinn. 2023. *What Is the Build-to-Rent Sector, and Who Does It Serve?* Washington, DC: Urban Institute: https://www.urban.org/sites/default/files/2023-06/What%20Is%20the%20Build-to-Rent%20Sector%20and%20Who%20Does%20It%20Serve.pdf.

Goodman, Laurie, Amalie Zinn, Kathryn Reynolds, and Owen Noble. 2023. *A Profile of Institutional Investor-Owned Single-Family Rental Properties*. Washington, DC: Urban Institute. https://www.urban.org/sites/default/files/2023-08/A%20Profile%20of%20Institutional%20Investor%E2%80%93Owned%20Single-Family%20Rental%20Properties.pdf.

Gurun, Umit G., Jiabin Wu, Steven Chong Xiao, and Serena Wenjing Xiao. 2022. "Do Wall Street Landlords Undermine Renters' Welfare?" *The Review of Financial Studies* 36 (1): 70–121. https://doi.org/10.1093/rfs/hhac017.

Healthpeak Properties, Inc. 2023. "Form 10-K Annual Report February 8th, 2023." https://www.sec.gov/Archives/edgar/data/765880/000162828023002794/ex21112312022.htm.

Hughes, Tenika, and *Action News Jax*. 2021. "Priced out of Jax: Build-to-Rent Communities Growing in Northeast Florida," *Action News Jax*, November 16. https://www.actionnewsjax.com/news/local/duval-county/priced-out-jax-build-to-rent-communities-growing-northeast-florida/WWNTZGSJUZBU7D7AQKXNBU47VU/.

Immergluck, Dan. 2018. "Renting the Dream: The Rise of Single-Family Rentership in the Sunbelt Metropolis," *Housing Policy Debate* 28 (5): 814–829. https://doi.org/10.1080/10511482.2018.1460385.

Invitation Homes Inc. 2022. "Form 10-K Annual Report February 22nd, 2023." https://www.sec.gov/ix?doc=/Archives/edgar/data/1687229/000168722923000029/invh-20221231.htm.

JAX Rental Housing Project. n.d. "Corporate Owned Single Family Homes and Eviction Filings." https://jaxrentalhousingproject.domains.unf.edu/wp-content/uploads/2022/08/JRHP-Research-Brief-2.1-Corporate-Owned-Single-Family-Homes-and-Eviction-Filings.pdf.

JustFix, Inc. n.d. "Methodology. Who Owns What in NYC?" https://whoownswhat.justfix.org.

Lennar Corporation. 2021. "Lennar Announces $4 Billion Single Family Rental Platform With Centerbridge Partners as Lead Investor." Press release. Miami, FL: PR Newswire. https://www.prnewswire.com/news-releases/lennar-announces-4-billion-single-family-rental-platform-with-centerbridge-partners-as-lead-investor-301249490.html.

Malone, Thomas. 2023. "Single-Family Home Investor Share Remained High in Q4 2022, With Smaller Investors Driving Demand," CoreLogic®, March 3. https://www.corelogic.com/intelligence/single-family-home-investor-share-remained-high-in-q4-2022-with-smaller-investors-driving-demand/.

Neo4j. 2023. "Release Notes: Neo4j 5." https://neo4j.com/release-notes/database/neo4j-5/.

Nethercote, Megan. 2020. "Build-to-Rent and the Financialization of Rental Housing: Future Research Directions," *Housing Studies* 35 (5): 839–874. https://doi.org/10.1080/02673037.2019.1636938.

Prechel, Harland, and Theresa Morris. 2010. "The Effects of Organizational and Political Embeddedness on Financial Malfeasance in the Largest U.S. Corporations: Dependence, Incentives, and Opportunities," *American Sociological Review* 75 (3): 331–354. https://doi.org/10.1177/0003122410372229.

Raymond, Elora Lee, Richard Duckworth, Benjamin Miller, Michael Lucas, and Shiraj Pokharel. 2018. "From Foreclosure to Eviction: Housing Insecurity in Corporate-Owned Single-Family Rentals," *Cityscape* 20 (3): 157–188.

Rhoades, Stephen A. 1993. "The Herfindahl-Hirschman Index," *Federal Reserve Bulletin* 79 (3): 188–189.

Seymour, Eric, and Joshua Akers. 2021. "'Our Customer Is America': Housing Insecurity and Eviction in Las Vegas, Nevada's Postcrisis Rental Markets," *Housing Policy Debate* 31 (3–5): 516–539. https://doi.org/10.1080/10511482.2020.1822903.

Shelton, Taylor, and Eric Seymour. 2024. "Horizontal Holdings: Untangling the Networks of Corporate Landlords," *Annals of the American Association of Geographers* 1–13. https://doi.org/10.1080/24694452.2023.2278690.

St-Hilaire, Cloé, Mikael Brunila, and David Wachsmuth. 2023. "High Rises and Housing Stress: A Spatial Big Data Analysis of Rental Housing Financialization," *Journal of the American Planning Association*: 1–15. https://doi.org/10.1080/01944363.2022.2126382.

Tapp, Renee, and Richard Peiser. 2023. "An Antitrust Framework for Housing," *Environment and Planning A: Economy and Space* 55 (3): 562–582. https://doi.org/10.1177/0308518X221135612.

Travis, Adam. 2019. "The Organization of Neglect: Limited Liability Companies and Housing Disinvestment," *American Sociological Review* 84 (1): 142–170. https://doi.org/10.1177/0003122418821339.

U.S. Census Bureau. 2021. "B0032: Hispanic or Latino Origin by Race." 2021 American Community Survey 5-year Estimates.

Commentary: How Data Architects Are Crafting Equitable Housing Policy Research

Matthew Murphy
NYU Furman Center

Nationwide, tools are surfacing to bridge knowledge gaps and deepen our comprehension of the effects of historical and recently changed housing policies. However, housing researchers, along with state and municipal governments, face a key and somewhat insurmountable challenge: insufficient data infrastructure to study housing policy comprehensively. To fill the gap, researchers are developing new methods to link together fragmented datasets to tell a more comprehensive story. Where data exist, researchers are refining the use of emerging technologies to avoid past pitfalls, such as racial biases. A focus has also been placed on pivotal housing justice concepts, like understanding the extent of the shift toward corporate ownership of America's rental properties, beyond public records.

I am excited about these kinds of advancements. The three articles in this series show the hoops researchers often go through to tell a comprehensive story about important housing policy considerations. The urgency to use data and new methods to adapt and meet the needs of residents and policymakers is evident in all housing-related work. Researchers require intricate data not just on cities and neighborhoods but on individual properties, their units, and their building systems. This kind of data has myriad applications. With escalating housing shortages, it is essential to grasp the full scope of available and potential housing inventory and what it might cost the public, or even the required investment from the market, to make new housing available. As natural disasters increase in frequency, understanding where people can seek refuge during and after these events becomes paramount. In a time of surging insurance premiums, robust asset management data will be needed to challenge broad assumptions made by insurers about low-income housing in particular. As we approach a preservation era, especially concerning low-income housing tax credit projects from the 1990s, in-depth property data will be crucial to ensuring the preservation and continuation of quality affordable housing. If local governments opt for social housing development, they must view these properties not just as monitoring points but as dynamic assets responding to market shifts. Finally, if local governments choose to use either a carrot or stick approach to force property owners' buildings to reduce carbon emissions, they will need detailed information about the systems in place tomorrow and what it will take to transform and recapitalize them for the next generation.

These particular issues influence me because I have seen them up close, having spent 8 years working in New York City government on affordable housing. I have sat in frustration, looking for data that should be much easier to find than they actually were. However, in my work as the executive director of the NYU Furman Center, I have come to realize that local authorities and advocates should perceive these challenges as opportunities.

The foundation for affordable housing policy discussions must be an ability to rely on high-quality and timely data. To fill this gap, researchers scrutinize whether they can rely on emerging technology, specifically regarding the concerns over racial bias. Linna Zhu, Michael Neal, and Caitlin Young's article explores automated valuation models (AVMs) that use advanced machine learning techniques. Their description of the possibilities of emerging technology provides a compelling case study on the potential technology has to address longstanding societal challenges. The study's findings reveal a higher percentage magnitude of AVM error in majority African-American neighborhoods, even after accounting for property condition and other variables, underscoring a troubling yet vital truth: although the allure of algorithms promises objectivity, they can inadvertently perpetuate biases present in historical data. It is especially concerning given the study's evidence that properties in poorer conditions or those in neighborhoods with a greater share of distressed sales are more susceptible to AVM inaccuracies. The use of the Light Gradient Boosting Machine, or LightGBM model, which demonstrated a 5.8-percentage-point improvement in model fit, showcases the potential of artificial intelligence in navigating the complexities and human biases assumed to be inherent in property valuation. However, the persistence of racial disparities in valuation, even with such advancements, is a stark reminder of the deeply entrenched effects of historical racism on property values and conditions. As we advance into the next generation of property valuation, a subjective field to begin with, it becomes crucial not just to adopt these sophisticated tools but to do so with an unwavering commitment to understanding and addressing the systemic racial biases they might reflect. At the same time, how do we come to rely on interior or even unit-level conditions at a more reliable scale? This article relies on a correlation between exterior conditions and interior ones. In practice, researchers or practitioners are extremely far away from having that knowledge available at scale. However, researchers should all strive to have the best possible data on housing conditions in their toolkits.

In a work that is near and dear to my heart, "Local Landscapes of Assisted Housing: Reconciling Layered and Imprecise Administrative Data for Research Purposes," researchers Shiloh Deitz, Will Payne, Eric Seymour, Kathe Newman, and Lauren Nolan of Rutgers University untangle the intricate data behind affordable housing programs. It is a funny but fitting name for an article—it reads as nearly satirical. However, the work is critically important for any state in the country. Their methodology involved meticulously cross-referencing disparate subsidy datasets, aligning them with property-level outcomes—a task of paramount importance in the realm of affordable housing, given the intricacies of layered subsidies. When I was a graduate student at New York University working at the Furman Center, I worked on the creation of a database of nearly all of New York City's affordable housing projects. Researchers still use the database today to inform preservation decisions and show the public which programs created what degree of affordable housing and where. I empathize with the researchers who took on this task, but I am so happy for their contribution.

A key revelation from their study was the identification of an overcounting of approximately 17,000 housing units. On paper, maybe 17,000 units does not sound like a lot. For example, about 3.5 million units are in New York City. However, picture a world in which a hurricane has displaced 5,000 households. A governor asks where people might be able to live—what is available tonight? You need to be a lot closer than off by 17,000. Rather, picture creating a 10-year strategy and needing to budget in the appropriate amount of money to rehabilitate all publicly subsidized units for 10 years. At $100,000 per unit, an overcount of 17,000 units is a $1.7 billion mistake. How can you get your arms around an issue when total population data are shaky to begin with? The research not only underscores the inherent challenges in handling multifaceted datasets but also indicates the potential discrepancies that can emerge without vigilant oversight. Something that stood out to me was that, despite leveraging sophisticated data integration techniques, it was indispensable for the research team to engage in manual data verification to ensure the accuracy of their findings. This confluence of advanced analytics and manual validation underscores the team's commitment to precision and integrity in research.

I do not think any housing planners can get away with a worthwhile project that does not involve some amount of manual data collection. Studying housing policies and programs is complicated, and some of the most critical data are often lacking. However, understanding housing programs and policies in detail *should* be complicated; it requires going to a physical place, whether it be an actual housing development or a file room at an under-resourced government agency. Studying housing requires an examination of the intersection of how people live (or want to live) with a huge number of considerations housing policy researchers account for, such as the state of physical structures, complicated financing structures and government subsidy programs at both the tenant and development level, land use regulations, property tax policies, patterns of neighborhood change and development, climate risk, and so much more. Skipping any of these considerations risks missing out on another important story that might be going on. It is the fun and frustration of being a housing policy researcher.

However, part of the frustration also has to do with why researchers must go through these hoops to begin with. Although the researchers' dedication here is laudable, it should prompt reflection on the broader institutional responsibilities. Should such intricate data reconciliation be the purview of academic researchers alone? It is imperative for government agencies to assume a more proactive role in ensuring the accuracy and comprehensiveness of housing data. This research underscores the pressing need for housing data linkage to be at the forefront of best management practices rather than a peripheral concern. However, I am confident that this research represents a starting place that governments can use and move forward. Having someone from the outside do it can be refreshing and invigorating—it is hard to pull off these projects within government, and often, government agencies are so under-resourced and lack needed technology, much less the data, so a project like this can help create the pressure necessary to shepherd the resources government employees require to up their game.

Speaking of hoops researchers must go through, even when publicly available information is quite good, the last article in the series, "Who Owns Our Homes? Methods to Group and Unmask Anonymous Corporate Owners" by Renz Torres, reads like an FBI forensic investigation into what it takes to understand who owns property, and to what degree. Torres uses graph-based analysis to study single-family rental ownership in Jacksonville, a methodological innovation. I have learned that a graph-based approach provides a more holistic view of the interconnections between property owners, addresses, and business officers. By transforming tax parcel and business registry data into a network of relationships, the study manages to uncover nuanced patterns of ownership and investment strategies. For instance, the clustering of single-family nonowner-occupied housing (NOOH) properties in Jacksonville's urban core versus the pronounced activity of mega investors in the suburbs provides a detailed landscape of market dynamics. In an interesting spin, Torres also uses corporate disclosure forms to examine the accuracy of the methodology. I love the use of public disclosure data in the absence of clear transparency.

The results, such as the differential investment behaviors in areas populated by people of color, are noteworthy; the novelty lies in the methodology. However, it is essential to view this approach with a degree of caution. The methods used to overcome the inherent challenges of such an analysis, such as over-clustering due to data inconsistencies, again remind us of the complexities of real-world data. Furthermore, although Florida's data standardization aids this approach, its scalability in less data-rich environments would have to be further understood. Nevertheless, the article marks a promising stride in housing research, hinting at the potential of graph-based methodologies to revolutionize our understanding of property markets and inform policy decisions. The question of what to do to advance homeownership remains, especially for first-time homebuyers, and what the positive roles that single-family rentals, even corporate-owned ones, might be, especially in an America that is becoming more housing-starved. However, Torres has shown us that with commitment, fancy data work, the right questions, and some manual labor, researchers can link together information to tell important stories.

I learned a lot reading these three articles, and I hope any data-driven housing practitioner does, too. The articles mix methodological innovations with common-sense applications. As a researcher and former government policymaker, I appreciate the use cases, and I am excited that the markets examined are not New York City or San Francisco. I look forward to how this world evolves and the role researchers play as technology grows exponentially during the next few years. The tools that were on the horizon 5 years ago are now here—whether and how they can be used to transform housing policy for positive effects remains to be seen. Current research workarounds, although essential, only partially address the core data challenges impacting our understanding and improvement of housing policy. A more impactful approach would involve directing federal funds to enhance technological and data collection capabilities of local and state governments. This enhancement should focus on comprehensive monitoring of the housing market, including cost fluctuations, natural disaster risks, and the condition of rental properties over time. Simultaneously, private entities should collaborate to provide accessible, standardized rent data, aiding in the accurate assessment of housing costs and the effectiveness of various policies. In addition, government-sponsored enterprises and state housing finance agencies could significantly contribute by making more of their data publicly available, contingent upon receiving necessary resources. Amidst these potential advancements, the research community continues to innovate. The authors of these articles offered a hint of the possibilities, and it is exciting to think about where we can all push next.

Acknowledgments

The author extends heartfelt gratitude to all colleagues at the NYU Furman Center, both past and present. Special acknowledgment is given to Rich Froehlich, a titan in the field of affordable housing, who dedicated his life to improving New York City. Rich's passion for using data, analysis, and technology to ensure affordable housing for New Yorkers was unparalleled. His commitment to educating a new generation of affordable housing practitioners in the intricacies of housing policy and finance, including the author, has left an indelible impact on the community.

Author

Matthew Murphy is the executive director of the NYU Furman Center.

Departments

In this issue—

- *Affordable Design*
- *Data Shop*
- *Graphic Detail*
- *Industrial Revolution*
- *Policy Briefs*

Secretary's Award for Excellence in Historic Preservation

Sherri L. Thurston
U.S. Department of Housing and Urban Development

The ACHP/HUD Secretary's Award for Excellence in Historic Preservation is sponsored by the U.S. Department of Housing and Urban Development (HUD) in collaboration with the Advisory Council on Historic Preservation (ACHP). Annually, the Secretary of HUD acknowledges the efforts of developers, organizations, and agencies in furthering historic preservation while concurrently addressing the critical needs of affordable housing, with an emphasis on preserving affordable housing in gentrifying neighborhoods and expanding economic opportunities for low- and moderate-income families and individuals.

Winner 2022: Victory Housing, Inc., Brinshore Development, LLC, and Bank of America Community Development Company, LLC—The Appleton at Spring Flats

Winner 2023: The Community Builders, Inc.—Commodore Place Apartments

Members of the jury:

- Lauren McNamara, Office of Community Planning and Development (CPD), HUD
- Lynn Rakos, Office of Community Planning and Development (CPD), HUD
- Richard Duckworth, Office of Policy Development and Research (PD&R), HUD
- Jordan Tannenbaum, Expert Member, Advisory Council on Historic Preservation (ACHP)

- Luke A. Nichter, Expert Member, Advisory Council on Historic Preservation (ACHP)
- Kristopher King, Expert Member, Advisory Council on Historic Preservation (ACHP)

The winners were selected by a jury of six judges convened on December 5, 2022, and an award ceremony was held on July 27, 2023, at HUD Headquarters in Washington, D.C. In attendance were The Honorable Eleanor Holmes Norton, U.S. Representative of the District of Columbia; The Honorable Shontel Brown, U.S. Representative of Ohio; The Honorable Adrianne Todman, Deputy Secretary, HUD; The Honorable Sara C. Bronin, ACHP Chair; Anita Cozart, Director, District of Columbia Office of Planning; and Cleveland Mayor Justin Bibb by way of video message. Due to COVID-19 restrictions last year, the event presented awards to the winners from 2022 and 2023.

The 2022 award was presented to The Appleton at Spring Flats in Washington, D.C., and the 2023 award was presented to Commodore Place Apartments in Cleveland, Ohio.

The 2022 Winner: The Appleton at Spring Flats in Washington, D.C.

Front facade of The Appleton at Spring Flats after renovation. Photo credit: ©Judy Davis, Architectural Photographer

Located in the Petworth neighborhood of Washington, D.C., The Appleton at Spring Flats renovation was completed in December 2021. The Appleton brought much-needed affordable housing stock to this transit-oriented community. It is a $29.8 million adaptive reuse of the historic Hebrew Home for the Aged, constructed in 1925 to serve the local Jewish community; it was added to the DC Inventory of Historic Sites in 2012 and the National Register of Historic Places in 2014. When the Hebrew Home moved to a new site in 1968, the building was sold to the Washington, D.C., government, which used it as a mental health facility for years before it was vacated in 2009. The building remained vacant for more than a decade, with boarded-up windows and an overgrown lawn. The Appleton is part of a redevelopment plan to build intergenerational, mixed-income rental and homeownership opportunities for families on a 3.3-acre parcel of land secured by the development team through a land disposition agreement with the District of Columbia.

The restoration work included cleaning and repairing the exterior masonry, two original cornerstones, and terra cotta and restoring original ornamental windows and other original details, such as the Star of David and other Jewish iconography. The renovated Appleton created 88 affordable, age-restricted apartments for seniors 62 and older and 14 Permanent Supportive Housing units for people experiencing homelessness. The amenities for the Appleton include a business center, wellness room for visiting healthcare providers, media center/game room, and fitness center. The apartments include ENERGY STAR appliances, a dishwasher, a garbage disposal, ENERGY STAR LED light fixtures, and cable readiness.

Dwelling unit at The Appleton at Spring Flats. Photo credit: ©Judy Davis, Architectural Photographer

The development team included an extensive community engagement process that helped ensure that the new development would meet the needs of the community. The team held a series of community charettes to gather input, in which members expressed their desire for affordable senior housing, homeownership opportunities, and sustainable public-space improvements that activate the streetscape. The overall redevelopment included a new streetscape along 10th Street NW adjacent to the site, public bike racks, a grant for local workforce development efforts and the restoration of the adjacent pocket park, and public art installations, including the sculpture by local artist Jackie Braitman, commissioned and installed to enhance the reflective nature of the space.

Leila Finucane, president and chief executive officer of Victory Housing, accepted the award on behalf of The Appleton at Spring Flats. She commented, "This endeavor [was] to restore the historic but long-vacant Hebrew Home for the Aged into The Appleton, an affordable senior community. Revitalizing this once-cherished community asset while fulfilling our mission to provide affordable senior housing has been a labor of love for Victory Housing, Brinshore, Bank of America CDC, Wiencek + Associates Architects + Planners, and Hamel Builders."

Juror Lauren McNamara commented that she appreciated the community involvement, supportive housing, and Leadership in Energy and Environmental Design certification. Juror Jordan Tannenbaum was impressed with the intergenerational aspect of the larger community development and that the team conducted a traffic study.

Congresswoman Eleanor Holmes Norton commented, "Historic preservation is important to the nation's capital … where such rich history has been made since our nation's founding. However, affordable housing is even more critical to the daily lives of D.C. residents."

The 2023 Winner: Commodore Place Apartments, Cleveland, OH

Front facade of Commodore Place Apartments after renovation. Photo credit: The Community Builders

Living and kitchen space at Commodore Place Apartments. Photo credit: The Community Builders

Commodore Place Apartments is located at the corner of Euclid Avenue and Ford Drive in the heart of Cleveland's University Circle district—one of the best Arts Districts in the country, as voted by the 2023 *USA Today* 10Best Readers' Choice Awards.[1] Completed in October 2021, the development is a mixed-income complex open to extremely low-income residents, long-time East Side residents, market-rate residents, students, and families. Commodore Place provides residents of all incomes access to shopping, dining, social services, the arts, and cultural opportunities. Originally opened in 1924 as a hotel and converted to apartments in 1964, the building retained many of its original features, including red brick and stone ornamentation around windows and entryways, characteristic of the Tudor Revival style. The restoration also retained the original corridor layout on the upper floors.

The new development consists of studios and one- and two-bedroom apartments with open floorplans, all-inclusive utilities, central heating and cooling, and energy-efficient kitchens. The building has a rectangular footprint, but the upper floors have irregular U-shaped floorplans; courtyards are on the rear and side. The ground floor has approximately 12,000 square feet of commercial space, and the upper 12 floors comprise 198 residential units, including 144 housing choice voucher and 54 market-rate units. Amenities include a community room, a fitness center, controlled building access, laundry rooms on every floor, and bicycle storage. However, Commodore Place's most exemplary facet is its history of housing East Side Cleveland residents in a first-rate location.

Juror Richard Duckworth noted that he was impressed by the number of units and subsidized units. He was also impressed with the number of HUD assistance contracts in a high-resource area because of intense pressure to go to the luxury market.

[1] See the complete list at https://10best.usatoday.com/awards/travel/best-arts-district-2023/.

Congresswoman Shontel Brown stated, "The redevelopment of Commodore Place was years in the making. I want to commend everyone with Community Builders in the city of Cleveland. … nothing of value gets built without an investment—an investment of capital and investment of time and investment of care. These beautiful and affordable apartments at Commodore Place are there because people believed in Cleveland." She continued, "Congresswoman Holmes Norton and I share Secretary Fudge's dedication to affordable fair and healthy housing for all Americans. Commodore Place shows that we can reinvest in ways that are also inclusive and affordable so the entire Community can benefit."

Author

Sherri L. Thurston, a Certified Government Meeting Professional (CGMP), is a program analyst/ meeting planner in the Research Utilization Division at HUD.

Data Shop

Data Shop, a department of Cityscape*, presents short articles or notes on the uses of data in housing and urban research. Through this department, the Office of Policy Development and Research introduces readers to new and overlooked data sources and to improved techniques in using well-known data. The emphasis is on sources and methods that analysts can use in their own work. Researchers often run into knotty data problems involving data interpretation or manipulation that must be solved before a project can proceed, but they seldom get to focus in detail on the solutions to such problems. If you have an idea for an applied, data-centric note of no more than 3,000 words, please send a one-paragraph abstract to datashop@hud.gov for consideration.*

Mapping Gentrification: A Methodology for Measuring Neighborhood Change

Serena Smith
Owais Gilani
Vanessa Massaro
Caroline McGann
Gavin Moore
Bucknell University

Michael Kane
Yale University

Abstract

The effects of gentrification are well studied, with varied findings. Studies debating and nuancing gentrification's effects have subsequently entailed variation on how the phenomena should be defined. The variance in definitions can create different calculations and potentially muddy findings on its effects. Having a well-defined methodology for calculating gentrification is essential to ensuring a deeper understanding of the phenomena and its effects. This article seeks to establish such a methodology that relies exclusively on publicly available data. This article overviews the definitions used in several peer-reviewed articles to identify 12 different methods for calculating gentrification. The authors created an interactive tool that classifies census tracts as gentrifying (https://ogilani.shinyapps.io/Gentrification/), nongentrifying, and nongentrifiable in metropolitan areas in the United States. Through a case study of

Abstract (continued)

Pittsburgh, the authors offer insights into which definition of gentrification best fits a qualitative understanding of the city. This article leaves readers with a methodology and tool for defining and mapping gentrification across the United States, making it easy to compare the results across different definitions. This tool and application offer a way for researchers, activists, and policymakers to compare various definitions in a particular geography to ensure consistent findings in studies across the United States.

Introduction

Gentrification has become an integral concept for understanding neighborhood change and exclusion in the United States. Although a general agreement on gentrification exists conceptually, how it is measured varies. Glass (1989) defined gentrification as elevating the economic status of the neighborhood's population and changing the "social character" or culture found within that neighborhood. Since then, researchers have debated how gentrification should be measured, but the same general idea is consistent throughout most definitions of gentrification. Conceptually, gentrification is understood as an influx of (White) residents with a higher socioeconomic status than the incumbent residents reshaping low-income, typically central city, neighborhoods (Barton, 2016; Mujahid et al., 2019; Zuk et al., 2018). Currently, several variations of the definition of gentrification center on economic status, whereas others argue for a focus on the influx of a cultural, creative class. Although they all funnel down to the same idea, the variance can have methodological implications for study outcomes that are not well understood.

The debate on the qualitative definition of gentrification has led to different measures and quantitative definitions of gentrification. Furthermore, the operationalization of different concepts, such as an increase in home values, is not agreed on, which is problematic. It is difficult to compare events linked to gentrification in cities and the potential effects of gentrification if the methodology for classification is not consistent. These differences in classification have caused a large variation in the number and location of areas classified as gentrifying, with minimal overlap between methods of classification (Barton, 2016; Mujahid et al., 2019). In other words, different methods for defining and operationalizing gentrification can lead to different results, not only in how many neighborhoods in a particular area are deemed to be gentrifying but also in the neighborhoods where gentrification is believed to occur.

This study aims to further investigate this issue by looking at the resulting areas of Pittsburgh, Pennsylvania, that are deemed gentrifying based on different variations of the most commonly used method for categorizing gentrification, the Freeman method (Freeman, 2005). In 2019, the National Community Reinvestment Coalition (NCRC) released a study conducted from 2000 to 2013. Through the 13-year period, 20 percent of the tracts in Pittsburgh gentrified, making it the eighth most gentrified city in the country as of 2013 (Richardson, Mitchell, and Franco, 2019). This article investigates the varying definitions of gentrification and applies them to the city of Pittsburgh. It is exemplary of a mid-sized city with little formal research on gentrification.

Gentrification is both an economic and cultural phenomenon, and how much to emphasize income in the calculation relative to educational attainment and other, less numeric, more subjective markers of cultural class contribute to this variation. Although the Freeman method is generally agreed on, few works take a larger examination of the methodology of calculating the qualifiers. This article seeks to remedy that gap by offering different variables for each of Freeman's categories and an interactive tool that researchers, activists, and policymakers can use to compare various definitions in a particular geography to ensure consistent findings in studies across the United States.

Measuring Gentrification in the Social Sciences

Gentrification is a phenomenon that tracks with a returning middle-class interest in inner city spaces that were previously disinvested via processes of deindustrialization and White flight (Clay, 1979). Despite the variance in defining and calculating gentrification, it is widely agreed that the prevalence of gentrification in the United States is rising because low rents attract the middle class (Ellen and Ding, 2016). Hwang and Lin (2016) note that changes in public policy, technology, demographics, socioeconomic factors, and housing supply all contribute to the growing prevalence.

Importance of Accurately Detecting Gentrification

Both retail and commercial changes can be indicators that low-income neighborhoods are undergoing change, and oftentimes, these same signs are what draw middle-class individuals to become residents in these neighborhoods (Brown-Saracino, 2004). These new amenities correlate with an increase in property values (including rent prices), White residents, and middle-class community members, which can make it difficult for the incumbent community members and businesses to stay in the neighborhood (Cunningham and Houston, 2012; Zukin, 2009). Koebel (2002) found that large amounts of change in neighborhoods were tied to both property and location characteristics. Gentrification leads to the revitalization of communities through reinvestment. However, as new residents move into the neighborhood, the changes can reinforce already existing patterns of segregation and inequality.

Gentrification is also linked to historical patterns of residential segregation (Freeman, 2005; Pattillo, 2008; Powell and Spencer, 2002; Wilson and Taub, 2006). Residential segregation often ties to redlining, which is the spatial discrimination against residents of a center city (Zenou and Boccard, 2000). When the cost for workers to maintain employment in center city areas is high, it results in segregated areas within the city by race (Zenou and Boccard, 2000). When the cost for workers to maintain employment in center city areas is low, it results in a city that is segregated by employment status (Zenou and Boccard, 2000). Between 1980 and 2010, the proportion of higher-income households in higher-income tracts doubled from 9 to 18 percent, whereas the proportion of lower-income households in lower-income tracts went from 23 to 25 percent during the same period (Fry and Taylor, 2012). In metropolitan areas, income inequality affected neighborhood income segregation the most, where the top quantile experienced much higher compensation, and the bottom quantile experienced an inadequate number of job opportunities (Reardon and Bischoff, 2011; Watson, 2009). Not only are jobs affected, but schools can be affected as well.

Schools in the United States continue to be highly segregated in nature, primarily caused by residential segregation along lines of race and income (Rivkin, 1994).

Gentrification has also been linked to the displacement of lower income residents (Grier and Grier, 1978). Residents moving into gentrifying neighborhoods have higher educational attainment, higher incomes, and are more likely to be White than the residents that have historically lived in the area. Consequently, characteristics of residents moving from gentrifying neighborhoods are more commonly non-White renters with lower income (Zuk et al., 2018). However, studies contradict each other when debating whether or not gentrification causes displacement (Zuk et al., 2018.). Vigdor, Massey, and Rivlin (2002) found that the housing turnover rates were larger in neighborhoods that were classified as gentrifying. Freeman and Braconi (2004) found that poor households in gentrifying neighborhoods were less likely to move than poor households in nongentrifying neighborhoods (Freeman 2005; Freeman and Braconi, 2004). Some of this discrepancy could be due to variations in operationalizing a definition of gentrification.

When displacement occurs, it affects minority groups the most. Based on data from 2000 to 2013, the NCRC reported that for gentrifying tracts in the United States, displacement of ethnic minorities was commonplace (Richardson, Mitchell, and Franco, 2019). Studies found there was less movement of poor households in gentrifying areas but the pressure became too much for the households, forcing them to move (Zuk et al., 2018).

If and when displacement occurs, it ultimately means the poorest do not benefit from the gentrification-driven neighborhood revitalization. The displacement also produces negative outcomes. Danley and Weaver (2018) found that in gentrifying cities, exclusion in daily activities and spaces are precursors to the displacement that coincides with gentrification. In addition, the fear of displacement that results from gentrification can lead to resistance to development that could affect the neighborhood in a positive way (Danley and Weaver, 2018). Gentrification can affect the allocation and maintenance of a community's resources and the community's social cohesion (Stanley, 2003). This decline in cohesion may occur, in part, due to the displacement of integral members of the community's support system and social networks (Mujahid et al., 2019).

Given the potential for gentrification to exacerbate inequality, drive displacement of low-income and racial minorities, and disrupt social cohesion, a clear understanding of gentrification and how to track it is essential. Researchers have been working on developing early warning systems for changes in neighborhoods, including changes associated with gentrification (Chapple and Zuk, 2016). The overarching idea is that through tracking community characteristics like investment, disinvestment, and population flow, policymakers can intervene to mitigate the changing patterns in the community before it is too late (Snow, Pettit, and Turner, 2003). In the 1980s, researchers developed early warning systems for gentrification and displacement that used different factors to determine which patterns most commonly indicated changing neighborhoods (Chapple and Zuk, 2016). Some early warning systems for community revitalization include monitoring changes in housing sales, racial shifts, and an influx of amenities to the area (Ellen and Ding, 2016). These changes can increase the level of exclusion in the neighborhood toward incumbent residents. Although the nuances and complexities are many in the relationship between displacement and gentrification, methodological consistency on calculating gentrification is essential for a clear and generalizable picture. Before being able to predict the effects of gentrification, including displacement in an area, gentrification must be defined and that definition operationalized.

Varied Methods for Classifying Gentrification

Variations and inconsistencies in calculating gentrification have practical consequences. In response to NCRC's "Shifting Neighborhoods" report (Richardson, Mitchell, and Franco, 2019), the *Pittsburgh City Paper*, a local Pittsburgh newspaper, talked about the discrepancies between the study's findings and residents' experiences who observed gentrification in the city (Deto, 2019). They specifically mentioned NCRC not classifying East Liberty as gentrified, which was surprising due to the high volume of African-American displacement and change in the area. In 2020, the *Pittsburgh City Paper* detailed the displacement of a Lawrenceville resident. The article described Lawrenceville as "what was once known as an enclave is now known for its lively restaurant scene, hip bars, and trendy boutique" (Deto, 2020). The article also discussed the displacement experienced across Lawrenceville and how most of the incumbent residents' wages are not enough to retain their homes, especially if they are on Section 8. This study's definitions of gentrification encompass Lawrenceville in its entirety (Lower, Central, and Upper), which match what those who live in the city fighting the change describe.

Given the ongoing debates in Pittsburgh, this city serves as an excellent case to better understand this tool and its policy utility. Methodologically, how scholars define gentrification can lead to subtle differences in which neighborhoods are labeled and how well that matches the lived experience of those places. Varying the methodology for labeling areas as gentrifying is key to robust studies and understandings of urban change across the United States.

This project formulates methods of measurement for gentrification throughout, built on the definitions Freeman (2005) established, which selected three criteria a neighborhood must meet to be gentrifiable (that is, has the potential for gentrification): (1) central city neighborhoods; (2) consisting of low-income households; and (3) households previously experienced disinvestment. Freeman (2005) operationalized criterion two by determining whether the median household income in a neighborhood was at or below the median household income for the metropolitan area in which the neighborhood belonged. Criterion three was met if the proportion of new housing within the past 2 decades in that neighborhood was lower than the proportion of new housing in the corresponding metropolitan area. If areas did not meet these criteria, they were nongentrifiable.

The following must occur for an area to undergo gentrification: (1) an influx in affluence associated with the neighborhood and (2) an increase in investment in the neighborhood. Criterion one was measured by looking at the educational attainment level within the neighborhood. Education is an indicator of class standing, and an increase in class is often tied to gentrification. Education was used in place of income because it highlights the difference between individuals who live within the neighborhood whose income increases and those who are moving into the neighborhood who may play a part in increasing the overall class standing of that neighborhood. Criterion one required a neighborhood to have an observed increase in educational attainment that is greater than or equal to that of the greater metropolitan area. Criterion two was measured using housing prices to represent investment in an area. Therefore, criterion two was met for a neighborhood if the real housing prices in that neighborhood increased during the period of analysis.

Ding, Hwang, and Divringi (2016) defined gentrification using American Community Survey (ACS) and U.S. Census Bureau data. Similar to the second criterion under potential for gentrification in Freeman's article (2005), Ding, Hwang, and Divringi (2016) defined a tract as *gentrifiable* if it had "a median household income at or below the citywide median at the beginning of the period of analysis." A tract was defined as *gentrifying* if it was gentrifiable at the beginning of the period of analysis and if the median percentage increase in residents with a college education and in gross rent or home value was above the citywide median percentage increase (Ding, Hwang, and Divringi, 2016). These stipulations cover criteria one and two under Freeman's (2005) process of gentrification requirements. Ding, Hwang, and Divringi (2016) chose to use home value *or* gross rent rather than home value *and* gross rent, because changes in the two different types of residencies do not necessarily coincide. However, both account for affordability. They defined a tract to be nongentrifying if the tract was determined to be gentrifiable but did not gentrify during the period of analysis.

Operationalizing a Gentrification Classification

Freeman's (2005) definition of gentrification is widely used and applied, but nuances exist in how it is operationalized. This study's definitions vary in how each of Freeman's (2005) criteria can be operationalized using publicly available data. A variety of variables are publicly available that one could use to assess Freeman's (2005) criteria for determining areas' gentrification statuses. Further, Ding, Hwang, and Divringi's (2016) analysis requires only one criterion be met rather than all. These seemingly subtle differentiations create different results. This study created 12 different definitions of gentrification that operationalize the criteria differently. They all draw from publicly available data to determine if areas within a selected metropolitan region are gentrifying.

This project specifies five variables indicative of gentrification according to the literature. It uses ACS data and defines "neighborhood" as census tracts. The variables included to operationalize Freeman's (2005) definition are (1) median household income, (2) median home value, (3) median rent, (4) vacancy rate, and (5) educational attainment. Median household income was used both as a variable for the classification of *gentrifiable* and for some of the definitions of *gentrifying*. Exhibit 1 details how each of these five variables are operationalized.

Exhibit 1

Operationalization of Each Variable to Classify Tracts

Variable	Criteria for Gentrification	Variable Name (2010 and 2019 ACS)
Median household income	This variable is used in two ways. First, it classifies tracts as *gentrifiable* or *nongentrifiable*. A tract is considered nongentrifiable if it is already above the citywide median income at the beginning of the analysis period (2010); else it is *gentrifiable*. Second, *gentrifiable* tracts are *gentrifying* if they experience a percentage increase in household income during the 10-year period (2010–19) that was more than the calculated citywide percentage increase during the same period.	S1901_C01_012E
Median home value	If a tract's percentage increase for home value is higher than the citywide median percentage increase in home value.	S2506_C01_009E
Gross rent	If a tract's percentage increase for rent cost is higher than the citywide median percentage increase in rent cost.	B25064_001E
Vacancy rate	If a tract's percentage decrease in proportion of vacant residencies is higher than the citywide median percentage decrease in vacant residencies.	B25002_003E
Educational attainment	If a tract's proportion of people aged 25 and older with a bachelor's degree or higher went from below the city average to higher during the 10-year period.	2010: S1501_C01_015E 2019: calculated using S1501_C01_015E and S1501_C01_006E

ACS = American Community Survey.
Source: Authors

The analysis period for this project was the 10 years from 2010 to 2019. The tool uses ACS 5-year estimates rather than the 3- and 1-year estimates, because the 3-year estimates had been discontinued, and the 1-year estimates were not offered for the Pittsburgh area at the tract level. The ACS data for 2010 and 2019 were retrieved using the *R* package *tidycensus* (Walker and Herman, 2023).

Each definition of gentrification uses a combination of the variables related to educational attainment, home value, rent cost, income, and vacancy. Data from 2010 were used to determine whether a tract was *gentrifiable* or not, and changes from 2010 to 2019 were used to determine whether *gentrifiable* tracts were *gentrifying* over this duration.

For every definition, a tract must first meet the requirement of being *gentrifiable*. The tract's median household income must be below the citywide median household income at the beginning of the analysis period (2010). If a tract meets this initial condition, then it is *gentrifiable*. If it is already above the citywide median income, it is considered *nongentrifiable*. Exhibit 2 outlines each definition's requirements for a *gentrifiable* census tract to subsequently be classified as *gentrifying*, and exhibit 3 provides a summary of the variables that were included in each of the 12 definitions.

Exhibit 2

Criteria for Each Definition

#	Description	Requirement
1	Freeman (2005) method with home value	A tract's percentage increase in median home value is above the percentage increase in citywide median home value at end of period, AND a tract's percentage increase in college educated residents is above the citywide percentage increase in college educated residents.
2	Freeman (2005) method with rent	A tract's percentage increase in median gross rent is above the percentage increase in citywide median gross rents, AND a tract's percentage increase in college educated residents is above the citywide percentage increase in college educated residents.
3	Freeman (2005) method with rent and home value	All criteria for definitions 1 and 2 must be met.
4	Freeman (2005) method with home value and vacancy	All criteria for definition 1 must be met, AND the tract's percentage decrease in vacancy is above the citywide percentage decrease in vacancy.
5	Freeman (2005) method with rent and vacancy	All criteria for definition 2 must be met, AND the tract's percentage decrease in vacancy is above the citywide percentage decrease in vacancy.
6	Freeman (2005) method with rent, home value, and vacancy	All criteria for definition 3 must be met, AND the tract's percentage decrease in vacancy is above the citywide percentage decrease in vacancy.
7	Freeman (2005) method with home value and income	All criteria for definition 1 must be met, AND the tract's percentage increase in median household income is above the citywide percentage increase in median household income.
8	Freeman (2005) method with rent and income	All the criteria for definition 2 must be met, AND the tract's percentage increase in median household income is above the citywide percentage increase in median household income.
9	Freeman (2005) method with rent, home value, and income	All the criteria for definition 3 must be met, AND the tract's percentage increase in median household income is above the citywide percentage increase in median household income.
10	Freeman (2005) method with home value, vacancy, and income	All the criteria for definition 4 must be met, AND the tract's percentage increase in median household income is above the citywide percentage increase in median household income.
11	Freeman (2005) method with rent, vacancy, and income	All the criteria for definition 5 must be met, AND the tract's percentage increase in median household income is above the citywide percentage increase in median household income.
12	Freeman (2005) method with rent and home value, vacancy, and income	All the criteria for definition 6 must be met, AND the tract's percentage increase in median household income is above the citywide percentage increase in median household income.

Source: Authors

Exhibit 3

Criteria Included for Each Definition

Definitions:	Gentrifiable	Home Value	Rent	Education	Vacancy	Income
1	X	X		X		
2	X		X	X		
3	X	X	X	X		
4	X	X		X	X	
5	X		X	X	X	
6	X	X	X	X	X	
7	X	X		X		X
8	X		X	X		X
9	X	X	X	X		X
10	X	X		X	X	X
11	X		X	X	X	X
12	X	X	X	X	X	X

Note: A tract is considered gentrifying if it was gentrifiable *at the beginning of the analysis period and met ALL criteria selected for that definition.*
Source: Authors

Recall that all definitions required that the proportion of residents with a bachelor's degree or higher be greater than the citywide median proportion of residents with a bachelor's degree or higher. All the definitions also required that the tracts be *gentrifiable*, that is, that the median household income for that tract in 2010 was less than the citywide median household income in 2010 (start of the analysis period). Beyond that, definitions 1 through 3 look at the percentage increases in home value, rent, and both home value and rent, respectively. Definitions 4 through 6 respectively build on definitions 1 through 3 by additionally assessing whether the percentage increase in proportion of vacant homes was less than the percentage increase in the citywide proportion of vacancy. Definitions 7 through 9 respectively build on definitions 1 through 3 by additionally examining whether the percentage increase in median household income is greater than the citywide percentage increase in median household income. Finally, definitions 10 through 12 respectively build on definitions 1 through 3 by assessing changes in both vacancy and median household income, as definitions 4 through 6 and 7 through 9 separately assess. This tool allows users to map gentrification for any metropolitan region in the United States using all 12 definitions. Each definition applies a unique combination of five publicly available variables. Users can download the data to use outside the application. The authors use this tool for Pittsburgh to better understand its implications, and the remainder of the article discusses this application.

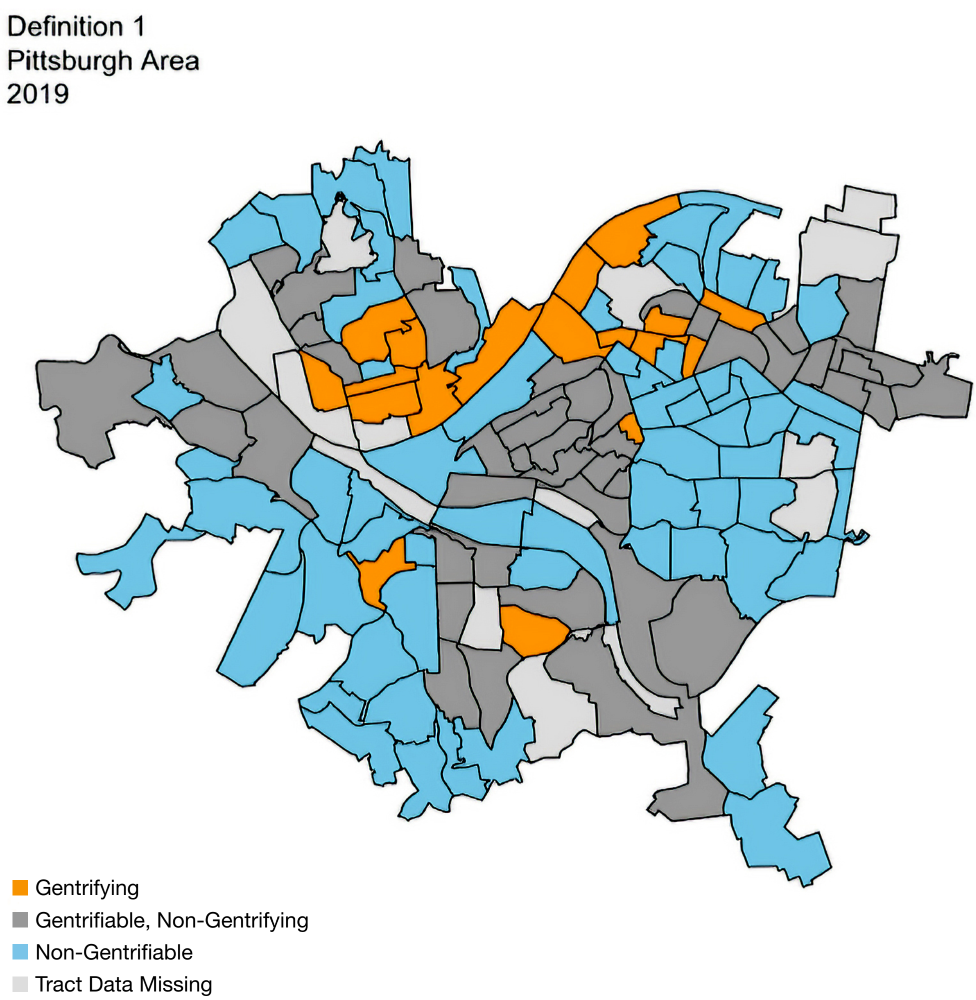
Case Study: Pittsburgh

The authors chose Pittsburgh as the case study city because it showed an interesting phenomenon when comparing the NCRC study results with conventional wisdom within the community. According to the P*ittsburgh City Paper*, NCRC lists Bloomfield, Downtown, Garfield, Lawrenceville, Polish Hill, sections of the North Side, and Mount Washington as areas of Pittsburgh that experienced gentrification. However, the study did not note East Liberty as a gentrified area of the city, which is inconsistent with popular belief among Pittsburgh residents (Deto, 2019; Richardson, Mitchell, and Franco, 2019).

The authors tested the 12 definitions for Pittsburgh to exemplify the utility of the tool for understudied but rapidly changing cities in the United States. This analysis explores the advantages and disadvantages of each definition. Exhibit 4 displays the study area.

Exhibit 4

Definition 1 Map of Pittsburgh



Source: Created using the authors' app, available here: https://ogilani.shinyapps.io/Gentrification/

Results

Exhibit 5 illustrates the breakdown of tracts considered to be gentrifying based on the various definitions. The rows correspond to all the different tracts found to be gentrifying in at least one of the definitions. The columns correspond to the different definitions. Across each row, the definitions in which the tract is identified as gentrifying have an “X” under that particular definition. The tracts found to be gentrifying based on all 12 definitions were 603, 809, 903, and 2406, corresponding to the neighborhoods of Lower Lawrenceville, Bloomfield, East Liberty, and Troy Hill, respectively. It is important to note that in some instances multiple tracts have the same neighborhood name, whereas in others, a tract is split across multiple neighborhoods. Among all tracts identified by at least one definition, tract 409 (South Oakland) appears in the fewest definitions (one). Exhibit 5 shows a more detailed breakdown of the definitions and their corresponding gentrifying tracts.

Exhibit 5

Tracts That Were Gentrifying in Pittsburgh According to Each Definition

		Definition												Frequency
Tracts	**Neighborhoods**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**of Inclusion**
404	North Oakland	X						X			X			3
406	Central Oakland		X		X	X								3
409	South Oakland		X											1
603	Lower Lawrenceville	X	X	X	X	X	X	X	X	X	X	X	X	12
807	Friendship	X	X	X				X	X	X				6
809	Bloomfield	X	X	X	X	X	X	X	X	X	X	X	X	12
901	Central Lawrenceville	X	X	X				X	X	X				6
903	Bloomfield	X	X	X	X	X	X	X	X	X	X	X	X	12
1011	Upper Lawrenceville	X	X	X				X	X	X				6
1017	Garfield	X						X			X			3
1113	East Liberty	X						X						2
1114	Garfield		X						X					2
1915	Mount Washington	X						X			X			3
2107	Manchester	X	X	X				X	X	X				6
2206	Central Northside	X	X	X				X	X	X				6
2406	Troy Hill	X	X	X	X	X	X	X	X	X	X	X	X	12
2509	Fineview	X	X	X	X	X	X				X			7
2614	Perry South	X						X						2
5617	Mount Oliver Neighborhood	X	X	X				X	X	X				6
5627	Allegheny Center, Allegheny West	X	X	X				X	X	X				6
5630	Chartiers City, Fairywood, Windgap		X						X					2
5632	East Allegheny, North Shore	X	X	X				X	X	X				6

Source: Authors

Discussion

After additional qualitative research, the authors determined definition 9 to be most representative of the reality for gentrification in Pittsburgh and consulted Nick Cotter from the Pittsburgh Neighborhood Project, a project that investigates Pittsburgh neighborhoods and the segregation (racial and economic) that still affects the city. Cotter believed that definitions 1 through 6 were good displays of market pressure but not necessarily gentrification. He also opposed the use of vacancy rates as a measure of gentrification in Pittsburgh specifically because somewhere between 70,000 and 80,000 individuals out of the population of 300,000 are students who are highly mobile by nature. Definition 9 considers both home value and rent, which were also thought to be important. This definition does not exclude any type of resident in the city because it represents both homeowners and renters. The authors chose to use home value *and* rent, as opposed to the broader home value *or* rent, because the gentrification phenomenon encompasses both home value and rent increasing rapidly. Altogether, the most important aspects of an area to investigate in defining gentrification in Pittsburgh are educational attainment, home value, rent, and income.

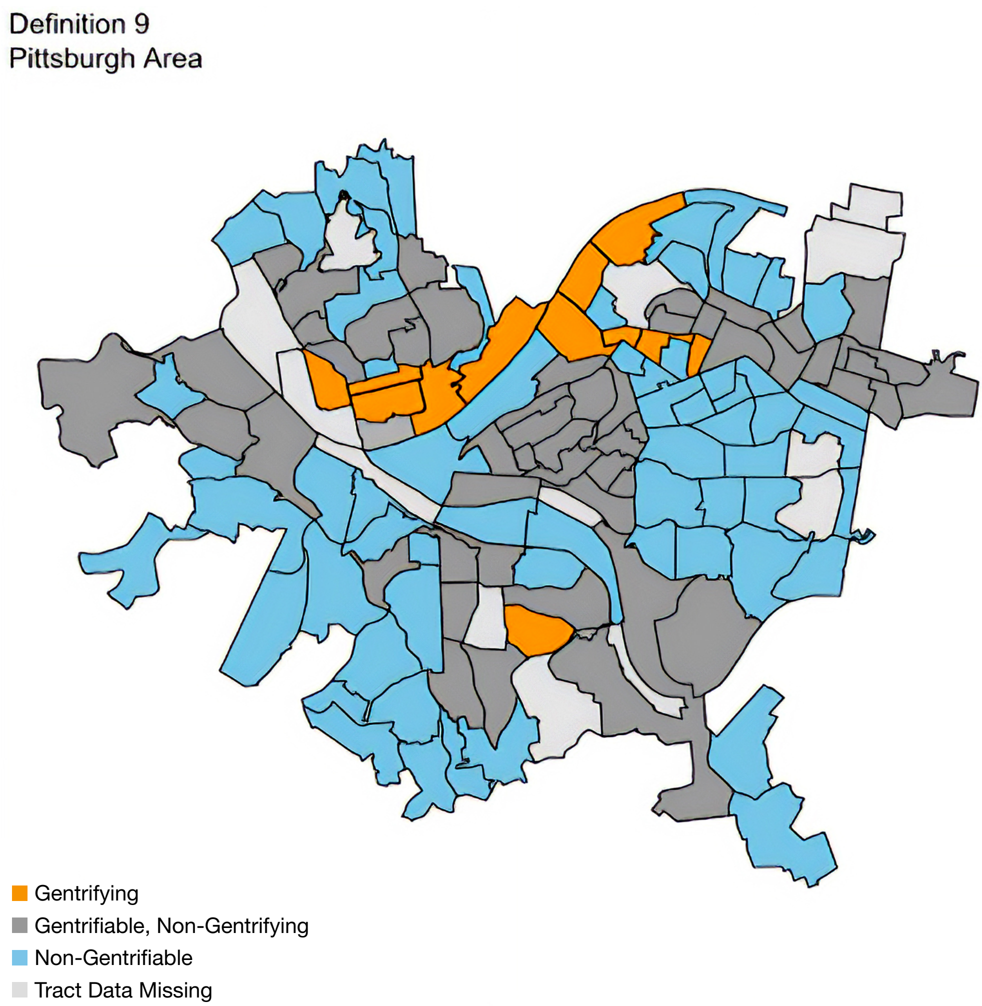
The tracts that were identified to be gentrifying by definition 9 were 603 (Lower Lawrenceville), 807 (Friendship), 809 (Bloomfield), 901 (Central Lawrenceville), 903 (Bloomfield), 1011 (Upper Lawrenceville), 2107 (Manchester), 2206 (Central Northside), 2406 (Troy Hill), 5617 (Mount Oliver), 5627 (Allegheny Center/West), 5630 (Chartiers City, Fairywood, Windgap), and 5632 (East Allegheny, North Shore). It is common that many more *gentrifiable* tracts exist than *gentrifying* tracts, because gentrification is not the rule, rather a special case. The gentrifying areas definition 9 identifies are the areas in which people who are familiar with Pittsburgh would also expect to identify as gentrifying during this period. These areas gained the most news coverage and drew the most attention with displacement or as “up and coming” or “hot spot” in Pittsburgh (exhibit 6).

In 2017, The Takeaway podcast series from WNYC Studios on the change in Pittsburgh due to gentrification titled “A Tale of Two Cities” explained that historic homes from the steel town are gone, with buildings being renovated into lofts near the river and the new and trendy eateries and high rises taking over areas (Asante, 2017). These areas were historically African-American locations like East Liberty. The segment discussed the pressure of rising housing prices and stagnant wages and how, from the viewpoint of a lifelong Pittsburgh resident, the atmosphere has changed, and the character of Pittsburgh had started to dwindle by comparison with how it used to be (Asante, 2017).

The Land & Power podcast series, released in 2020, explained changes from the perspective of the residents and a true story of displacement in Pittsburgh due to the eviction of Penn Plaza in East Liberty. It detailed the various stages of eviction due to the gentrifying area over the course of decades, and it talked about the effect that it had on an entire building of residents and the rest of the community. Uprooting many of these senior citizens left them feeling like they no longer had a support system and that their network had been dismantled (Krauss, 2020).

Exhibit 6

Definition 9 Map of Pittsburgh



Source: Created using the authors' app, available here: https://ogilani.shinyapps.io/Gentrification/

Conclusion

This analysis of Pittsburgh reveals several strengths and limitations of the tool.[1] After looking at the literature, a strength of this study is that the authors were able to establish multiple definitions of gentrification and to create dynamic maps for the Pittsburgh area based on these various definitions. The authors were able to capture what appears to be the largest areas in Pittsburgh experiencing change due to gentrification. In addition, unlike the National Community Reinvestment Coalition report, East Liberty was identified as a gentrifying neighborhood in this

[1] Available at https://ogilani.shinyapps.io/Gentrification/.

study (Richardson, Mitchell, and Franco, 2019). Furthermore, this analysis can be easily replicated for other cities in the United States using this tool.

This tool has utility in future research and also for policymakers. Users can match it to the ground reality they experience. It can help better identify the factors in specific locales that are most important for driving gentrification. In other words, this tool can help better confirm that a neighborhood is gentrifying, even if it comes up only on some of the definitions.

This study also has limitations. First, it would have been preferable to use Census Bureau data over ACS data. However, the 2020 Census uses different tract boundaries. Changing tract boundaries between the decennial census surveys restricts this type of analysis to 10-year periods. In addition, given the transient population, Pittsburgh may be unique, which would mean that the ideal definition for Pittsburgh may not apply everywhere. However, this study is easily replicable, which means that others can see which definitions apply best to a city of interest. Last, the Freeman (2005) criteria did include households previously experiencing disinvestment. The authors were unable to use this criterion because they were unable to find a publicly available data source that provided the necessary information at the census tract level.

Further research should investigate how different definitions of gentrification apply to different gentrifying areas. The tool and operationalizations of variables gave insights into Pittsburgh, but it is important to determine whether or not it is true for other gentrifying areas, especially due to the fact that every city has its own challenges and differences. For example, definition 9 did not include vacancy rates due to the high proportion of college students in the city, but it might be productive and beneficial to include vacancy rates for an area that potentially does not have such a high population of students. Here, the authors developed several variations of the same definitions that can be applied to and checked in different areas of concern. Further work could be done to investigate the dynamic progression of gentrification by looking at the change on a year-by-year basis or the overlap of years. However, gentrification is a slow process, and it is not likely that much change would be seen year to year.

Acknowledgments

This material is based upon work supported by the National Science Foundation under Grant No. 2024233 and 2024335.

Authors

Serena Smith is an Associate at The Talent Studios. Owais Gilani is an Associate Professor of Statistics in the Department of Mathematics at Bucknell University. Michael Kane is an Assistant Professor of Biostatistics at Yale University School of Public Health. Vanessa Massaro is an Associate Professor in the Department of Geography at Bucknell University. Caroline McGann is a Geography and Economics major at Bucknell University. Gavin Moore is a Mathematics major at Bucknell University.

References

Asante, John. 2017. "Gentrification in Pittsburgh: A Tale of Two Cities." *The Takeaway*, WNYC Studios. https://www.wnycstudios.org/podcasts/takeaway/segments/gentrification-pittsburgh-tale-two-cities.

Barton, Michael. 2016. "An Exploration of the Importance of the Strategy Used to Identify Gentrification," *Urban Studies* 53 (1): 92–111.

Brown-Saracino, Japonica. 2004. "Social Preservationists and the Quest for Authentic Community," *City & Community* 3 (2): 135–56. doi:10.1111/j.1535-6841.2004.00073.x.

Chapple, Karen, and Miriam Zuk. 2016. "Forewarned: The Use of Neighborhood Early Warning Systems for Gentrification and Displacement," *Cityscape* 18 (3): 109–130. http://www.jstor.org/stable/26328275.

Clay, Phillip L. 1979. *Neighborhood Renewal: Middle-Class Resettlement and Incumbent Upgrading in American Neighborhoods*. Lexington, Mass: Lexington Books.

Cunningham, Matt, and Dan Houston. 2012. "The Civic Economics of Retail: Ten Years of Studies." Civic Economics. https://nebula.wsimg.com/eb1a35cadd85dd440dcba5cb1eba005e?AccessKeyId=8E410A17553441C49302&disposition=0&alloworigin=1.

Danley, Stephen, and Rasheda Weaver. 2018. "'They're Not Building It for Us': Displacement Pressure, Unwelcomeness, and Protesting Neighborhood Investment," *Societies* 8 (3). https://www.proquest.com/scholarly-journals/they-re-not-building-us-displacement-pressure/docview/2124145210/se-2,doi:http://dx.doi.org/10.3390/soc8030074.

Deto, Ryan. 2020. "The Displacement of Anthony Hardison From His Lawrenceville Apartment Is a Microcosm of a Neighborhood Epidemic," *Pittsburgh City Paper*, 15 January. https://www.pghcitypaper.com/news/the-displacement-of-anthony-hardison-from-his-lawrenceville-apartment-is-a-microcosm-of-a-neighborhood-epidemic-16556108.

———. 2019. "Pittsburgh Is One of the Most Gentrified Cities in the U.S." *Pittsburgh City Paper*, 4 April. https://www.pghcitypaper.com/news/pittsburgh-is-one-of-the-most-gentrified-cities-in-the-us-14381722.

Ding, Lei, Jackelyn Hwang, and Eileen Divringi. 2016. "Gentrification and Residential Mobility in Philadelphia," *Regional Science and Urban Economics* 61: 38–51.

Ellen, Ingrid G., and Lei Ding. 2016. "Advancing Our Understanding of Gentrification," *Cityscape* 18 (3): 3–8. https://www.huduser.gov/portal/periodicals/cityscpe/vol18num3/guest.html.

Freeman, Lance. 2005. "Displacement or Succession? Residential Mobility in Gentrifying Neighborhoods," *Urban Affairs Review* 40 (4): 463–491.

Freeman, Lance, and Frank Braconi. 2004. "Gentrification and Displacement New York City in the 1990s," *Journal of the American Planning Association* 70 (1): 39–52.

Fry, Richard, and Paul Taylor. 2012. "The Rise of Residential Segregation by Income," *Pew Research Center's Social & Demographic Trends Project blog*, August 1. https://www.pewresearch.org/social-trends/2012/08/01/the-rise-of-residential-segregation-by-income/.

Glass, Ruth. 1989. "London: Aspects of Change." In *Clichés of Urban Doom and Other Essays*. Oxford: B. Blackwell: 132–158.

Grier, George, and Eunice Grier. 1978. *Urban Displacement: A Reconnaissance*. Washington, DC: U.S. Department of Housing and Urban Development, Office of the Secretary; Bethesda, MD: Grier Partnership.

Hwang, Jackelyn, and Jeffery Lin. 2016. "What Have We Learned About the Causes of Recent Gentrification?" *Cityscape* 18 (3): 9–26. https://www.huduser.gov/portal/periodicals/cityscpe/vol18num3/article1.html.

Koebel, C. Theodore. 2002. "Analyzing Neighborhood Retail and Service Change in Six Cities." Blacksburg, VA: Center for Housing Research, Virginia Polytechnic Institute and State University.

Krauss, Margaret J. 2020 "We Don't Do Business This Way," *Land & Power podcast*. https://www.eastliberty.org/news-land-power-podcast-dives-deep-into-the-story-of-penn-plaza-gentrification/.

Mujahid, Mahasin S., Elizabeth Kelley Sohn, Jacob Izenberg, Xing Gao, Melody E. Tulier, Matthew M. Lee, and Irene H. Yen. 2019. "Gentrification and Displacement in the San Francisco Bay Area: A Comparison of Measurement Approaches," *International Journal of Environmental Research and Public Health* 16 (12): 2246.

Pattillo, Mary. 2008. *Black on the Block: The Politics of Race and Class in the City*. Chicago: University of Chicago Press.

Powell, John A., and Marguerite L Spencer. 2002. "Giving Them the Old One-Two: Gentrification and the K.O. of Impoverished Urban Dwellers of Color," *Howard Law Journal* 46: 433.

Reardon, Sean F., and Kendra Bischoff. 2011. "Income Inequality and Income Segregation," *American Journal of Sociology* 116 (4): 1092–1153. https://doi.org/10.1086/657114.

Richardson, Jason, Bruce Mitchell, and Juan Franco. 2019. "Shifting Neighborhoods: Gentrification and Cultural Displacement in American Cities." National Community Reinvestment Coalition. https://ncrc.org/gentrification/?gclid=EAIaIQobChMIvJiG0K7sgQMVmoVaBR1GLQFWEAAYASAAEgL3WPD_BwE.

Rivkin, Steven G. 1994. "Residential Segregation and School Integration," *Sociology of Education* 67 (4): 279–292. https://doi.org/10.2307/2112817.

Snow, Christopher W., Kathryn L. Pettit, and Margery A. Turner. 2003. *Neighborhood Early Warning Systems: Four Cities' Experience and Implications for the District of Columbia*. Washington, DC: The Urban Institute Metropolitan Housing and Communities Policy Center.

Stanley, Dick. 2003. "What Do We Know about Social Cohesion: The Research Perspective of the Federal Government's Social Cohesion Research Network," *The Canadian Journal of Sociology / Cahiers Canadiens de Sociologie* 28 (1): 5–17. https://doi.org/10.2307/3341872.

Vigdor, Jacob L., Douglas S. Massey, and Alice M. Rivlin. 2002. "Does Gentrification Harm the Poor?" *Brookings-Wharton Papers on Urban Affairs* 133–182.

Walker, Kyle, and Matt Herman. 2023. "Tidycensus: Load U.S. Census Boundary and Attribute Data as 'tidyverse' and 'sf'-Ready Data Frames. R package version 1.5." https://walker-data.com/tidycensus/.

Watson, Tara. 2009. "Inequality and the Measurement of Residential Segregation by Income in American Neighborhoods," *Review of Income and Wealth* 55 (3): 820–844.

Wilson, William J., and Richard P. Taub. 2006. *There Goes the Neighborhood: Racial, Ethnic, and Class Tensions in Four Chicago Neighborhoods and Their Meaning for America*. 1st ed. New York: Knopf.

Zenou, Yves, and Nicolas Boccard. 2000. "Racial Discrimination and Redlining in Cities," *Journal of Urban Economics* 48 (2): 260–285.

Zuk, Miriam, Ariel H. Bierbaum, Karen Chapple, Karolina Gorska, and Anastasia Loukaitou-Sideris. 2018. "Gentrification, Displacement, and the Role of Public Investment," *Journal of Planning Literature* 33 (1): 31–44.

Zukin, Sharon. 2009. *Naked City: The Death and Life of Authentic Urban Places*. Oxford University Press. https://books.google.com/books/about/Naked_City.html?id=DbSK-x5Je3AC.

Graphic Detail

Geographic Information Systems (GIS) organize and clarify the patterns of human activities on the Earth's surface and their interaction with each other. GIS data, in the form of maps, can quickly and powerfully convey relationships to policymakers and the public. This department of Cityscape *includes maps that convey important housing or community development policy issues or solutions. If you have made such a map and are willing to share it in a future issue of* Cityscape*, please contact alexander.m.din@hud.gov.*

Whom Do We Serve? Refining Public Housing Agency Service Areas

Kristen Tauber
Ingrid Gould Ellen
Katherine O'Regan
NYU Wagner School and Furman Center

Abstract

Knowing public housing authority/agency (PHA) service areas is important for understanding the neighborhoods that housing choice voucher holders can most easily reach and for evaluating programmatic reforms and new laws. However, no centralized database on PHA service areas exists, and the information is not always available on PHA websites. The U.S. Department of Housing and Urban Development (HUD) has developed a set of estimated service areas, but they often differ significantly from areas reported on individual PHA websites or fail to capture a large share of voucher activity. In this article, the authors provide a new methodology for improving the coverage of HUD estimated service areas.

Background

Although the Housing Choice Voucher (HCV) program is a federal program, it is managed by more than 3,300 public housing authorities/agencies (PHAs), each with its own jurisdictional boundaries, or service areas. Although technically, voucher holders can use their vouchers to lease a home anywhere in the United States, using vouchers outside the service area of the issuing PHA can be administratively burdensome. Knowing PHA service areas is thus necessary for understanding the choices presented to HCV holders and for estimating the impacts of programmatic reforms and new laws.

One of the primary goals of the HCV program is to help participants move to higher-opportunity, lower-poverty neighborhoods. However, considerable evidence indicates that it is failing to promote this type of mobility (Ellen, 2020; Galvez, 2010; McClure, Schwartz, and Taghavi, 2015; Wood, Turnham, and Mills, 2008). Garboden (2021) points to administrative boundaries being part of this problem. He finds that voucher holders predominantly move within service areas and that cross-service area moves are extremely rare. His analysis shows that service area boundaries are more salient than either county or municipal boundaries. Knowing service area boundaries can shed light on restrictions on voucher holders' choices and show how coordination and cooperation among PHAs can overcome those barriers.

Knowing a PHA's service area is also key to understanding and evaluating the effects of policies like Small Area Fair Market Rents (SAFMRs) and source of income (SOI) discrimination laws. For example, the introduction of SAFMRs could have different implications for PHAs serving primarily low-rent or high-rent ZIP Codes (Dastrup et al., 2018). Consider also that SOI discrimination laws are enacted by counties and municipalities (not housing agencies), so identifying which counties and municipalities fall within a PHA's service area is important.

Finally, knowing PHA service areas is necessary for creating the data and maps that HUD committed to providing to PHAs and the public as part of its recently reintroduced 2015 Affirmatively Furthering Fair Housing (AFFH) rule. However, no centralized source of information on PHA service areas exists, and few PHAs clearly identify their service areas on their websites. To address this gap, HUD estimated service area boundaries for PHAs across the country,[1] relying on PHA names and the location of most of the PHA's voucher and low-rent units (total units) (HUD, 2018).[2] HUD defines the service areas of state and county PHAs as the entire state or county. Regional PHAs operating in more than one county are assigned the single county that contains most of the PHA's units, potentially leading to under-delineation of its service area. HUD's methodology can also over-delineate the service areas of county, state, or regional PHAs that do not serve areas being served by local PHAs (those that do not operate at the state or county level). HUD defines the service areas of local PHAs as either the unit of general local government (UGLG) that contains the largest share of a PHA's units or the whole county if 20 percent of units are outside the UGLG (HUD, 2018). HUD's estimated areas, which do not specifically focus on vouchers, often are inconsistent with the service areas reported on the few PHA websites that clearly describe them, likely because of the issues described previously.

The authors used administrative HCV data from 2017 to address these delineation problems by modifying HUD's estimated service areas into "revised service areas," which better capture most of the voucher activity for PHAs.

[1] The map can be accessed at https://www.arcgis.com/home/item.html?id=651cfdd0047b463f9aee56d354ad0515.

[2] State PHAs are identified using names and codes if the name contains the state name or the PHA code begins with a "9."

Methods

The authors developed two tests to check the validity of HUD's estimated service areas for PHAs that administer at least 50 vouchers.[3] The first assesses if the estimated service area omits a sizable share of voucher holder locations (so is "too small"); the second checks whether estimated service boundaries include areas the PHA does not seem to serve and that are served by another PHA (so is "too large"). The authors performed these tests and corrections on 377 PHAs with at least 80 percent of their vouchers in a metropolitan area required to adopt SAFMRs or a set of comparison metros.

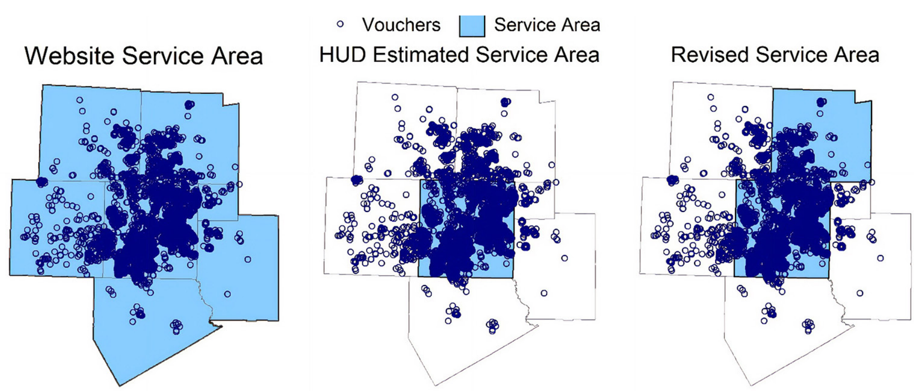
Too Small

The service area of a PHA is considered to be too small if it contains less than 90 percent of the PHA's voucher holders. An example is the Housing Authority of the City of Dallas (PHA code TX009). Exhibit 1 shows the service area described on TX009's website, HUD's estimated service area, and the new, revised service area. The website for the housing authority states that it serves Collin, Dallas, Denton, Ellis, Kaufman, Rockwall, and Tarrant Counties;[4] however, the HUD estimated service area includes only Dallas County. The location of voucher holders that received vouchers from TX009 clearly shows that the PHA serves areas outside Dallas County. Indeed, about 14 percent of the PHA's voucher holders reside outside Dallas County.

A PHA's service area is made larger by adding counties or places that contain at least 5 percent of its voucher holders.[5] The revised service area for TX009 adds Collin County, increasing the share of vouchers covered by the service area from 86 to 92 percent. The other counties listed on the PHA's website are not included because each contains a very small share of the PHA's total vouchers.

Exhibit 1

Housing Authority for the City of Dallas



Sources: Voucher data are from the 2017 administrative PIH Information Center longitudinal dataset; county outlines come from 2019 TIGER/Line Shapefiles via IPUMS NHGIS; HUD's Estimated Service Areas are from the e-GIS Storefront; revised service areas are from the authors' estimations

[3] The authors use different criteria for PHAs that administer less than 50 vouchers. Please see the description of our methodology available on the HUD-eGIS Storefront for more details.

[4] https://dhantx.com/applicants-residents/housing-choice-voucher-program/.

[5] Please see the description of our methodology available on the HUD-eGIS Storefront for more details on the different circumstances under which the authors add counties and places.

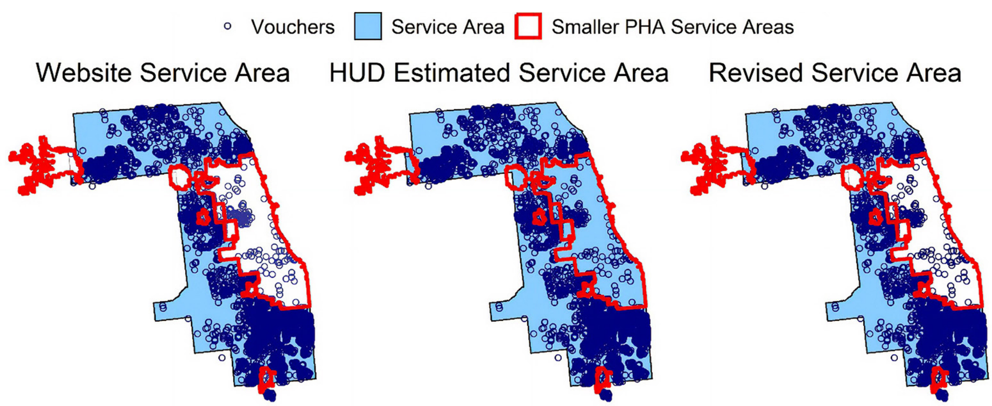
Too Large

A PHA's service area is deemed to be too large if it overlaps with the service area of a second PHA and nearly all voucher holders in that area are from the second PHA.[6] An example of this situation is the HA of Cook County, Illinois (IL025). Its website states that it serves unincorporated or suburban parts of the counties *that are not served by other housing authorities*.[7] However, HUD's estimated service area includes the entirety of Cook County.

Exhibit 2 shows the different service areas for IL025 and the locations of its voucher holders. Six smaller PHAs serve the incorporated parts of Cook County: Chicago, the Village of Oak Park, Maywood, Cicero, Park Forest, and Elgin. The HA of Cook County has just under 1.2 percent of its voucher holders in those areas or outside the county, and vouchers from the HA of Cook County account for less than 5 percent of vouchers in each of these smaller PHAs; thus, the first five jurisdictions from IL025's service area are omitted.[8] The revised service area still covers more than 98 percent of the PHA's voucher holders.

Exhibit 2

Housing Authority of Cook County



PHA = public housing authority/agency.

Sources: Voucher data are from the 2017 administrative PIH Information Center longitudinal dataset; county and CDP outlines come from 2019 TIGER/Line Shapefiles via IPUMS NHGIS; HUD's Estimated Service Areas are from the e-GIS Storefront; revised service areas are from the authors' estimations

[6] Less than 5 percent of the larger PHA's voucher holders and less than 5 percent of the total voucher holders in the second PHA service area come from the larger PHA.

[7] See the HA of Cook County (HACC) website at https://thehacc.org/about/.

[8] The authors do not omit Elgin because HUD's estimated service area for Elgin is Kent County and thus does not overlap with Cook County. The HA of Elgin is therefore not identified as overlapping with the HA of Cook County under the current methodology.

Results

In total, the authors classify the estimated HUD service areas of 100 PHAs as too small, 81 as too large, and 32 as both too small and too large.[9] (HUD's service areas are left as is for 229 PHAs in this study's sample.)[10] Exhibit 3 shows the number of PHAs with HUD estimated service areas and revised service areas by different voucher-coverage bins. Row one shows that the proposed revision process increases the number of PHAs with service areas covering at least 90 percent of vouchers from 272 to 354, or from 72 percent of the sample to 94 percent.

Exhibit 3

Results Summary

Share of Primary PHA Vouchers Covered	HUD Estimated Service Areas	Revised Service Areas
90+	272	354
80–89	43	17
70–79	31	4
< 70	31	2

PHA = public housing authority/agency.

Sources: Voucher data is from the 2017 longitudinal administrative PIH Information Center dataset; HUD's Estimated Service Areas are from the e-GIS storefront; Revised service areas from the authors' estimations

Conclusion

Having good estimates of PHA service areas is critical for policy evaluation, development, and research. The authors believe that their adjustments to HUD's methodology yield more accurate service areas and hope that researchers and program administrators who need more accurate (although not necessarily perfect) information on any PHA's service area find these materials useful.

Notes

Additional details of the methodology and associated data files are available on HUD's geospatial database, the HUD-eGIS Storefront.

Acknowledgments

The authors thank the U.S. Department of Housing and Urban Development (HUD) for access to administrative data and both HUD and the Wells Fargo Foundation for financial support.

[9] A PHA's service area is both too small and too large if it contains smaller jurisdictions that it does not serve and are served by other PHAs and if more than 10 percent of its voucher holders are outside the HUD estimated service area.

[10] One of the PHAs classified as too small, ID016, does not have any geographies that can be added, so its service area remains unadjusted.

Authors

Kristen Tauber is a doctoral student at the New York University (NYU) Wagner School and doctoral fellow at the Furman Center. Ingrid Gould Ellen is a professor of urban policy and planning at the NYU Wagner School and a faculty director at the Furman Center. Katherine O'Regan is a professor of public policy and planning at the NYU Wagner School and a faculty director at the Furman Center.

References

Dastrup, Samuel, Meryl Finkel, Kimberly Burnett, and Tanya De Sousa. 2018. *Small Area Fair Market Rent Demonstration Evaluation: Final Report*. Report prepared for the U.S. Department of Housing and Urban Development, Office of Policy Development and Research. Washington, DC: Government Publishing Office. http://dx.doi.org/10.2139/ssrn.3615783.

Ellen, Ingrid Gould. 2020. "What Do We Know About Housing Choice Vouchers?" *Regional Science and Urban Economics* 80: 103380. https://doi.org/10.1016/j.regsciurbeco.2018.07.003.

Galvez, Martha M. 2010. *What Do We Know About Housing Choice Voucher Program Location Outcomes? A Review of Recent Literature*. What Works Collaborative. Washington, DC: Urban Institute. https://www.urban.org/research/publication/what-do-we-know-about-housing-choice-voucher-program-location-outcomes.

Garboden, Philip M.E. 2021. "You Can't Get There from Here: Mobility Networks and the Housing Choice Voucher Program," *Journal of Planning Education and Research*. https://doi.org/10.1177/0739456X211051774.

McClure, Kirk, Alex F. Schwartz, and Lydia B. Taghavi. 2015. "Housing Choice Voucher Location Patterns a Decade Later," *Housing Policy Debate* 25 (2): 215–33. https://doi.org/10.1080/10511482.2014.921223.

U.S. Department of Housing and Urban Development (HUD). 2018. *Estimated Housing Authority Service Areas*. Washington, DC: HUD Office of Policy Development and Research. https://www.arcgis.com/home/item.html?id=651cfdd0047b463f9aee56d354ad0515.

Wood, Michelle, Jennifer Turnham, and Gregory Mills. 2008. "Housing Affordability and Family Well-Being: Results From the Housing Voucher Evaluation," *Housing Policy Debate* 19 (2): 367–412. https://doi.org/10.1080/10511482.2008.9521639.

Visualizing Veteran and Nonveteran Homelessness Rates in Virginia

Brent D. Mast
Tricia Ruiz
U.S. Department of Housing and Urban Development

The views expressed in this article are those of the authors and do not represent the official positions or policies of the Office of Policy Development and Research, the U.S. Department of Housing and Urban Development, or the U.S. Government.

Introduction

The U.S. Department of Veterans Affairs (VA) publishes veteran population estimates at the national, state, and county levels. The U.S. Department of Housing and Urban Development (HUD) publishes homeless count estimates at the national, state, and Continuum of Care (COC) levels. COCs are service areas within state regions that receive HUD grants to address homelessness.

Studies that focus on improving measures of veteran homelessness rates help to advance work in public policy, social services, and veteran care. For example, having timely data for the veteran population at the state and local levels may help providers of veteran care to target resources to veteran populations at the highest risk for homelessness. Tsai and Rosenheck (2015) found that veterans experiencing mental illness, substance use disorders, social isolation, and past incarceration are at the highest risk for homelessness.

Mast (2023) combined VA state veteran population estimates with HUD state Point-In-Time (PIT)[1] homeless count estimates and U.S. Census Bureau state population estimates to approximate veteran and nonveteran homelessness rates for the 50 states and the District of Columbia for 2008, 2014, and 2020. Mast found that at the state level, on average, veteran homeless rates are higher compared with nonveteran homeless rates, although the mean rates have decreased each year for both groups.

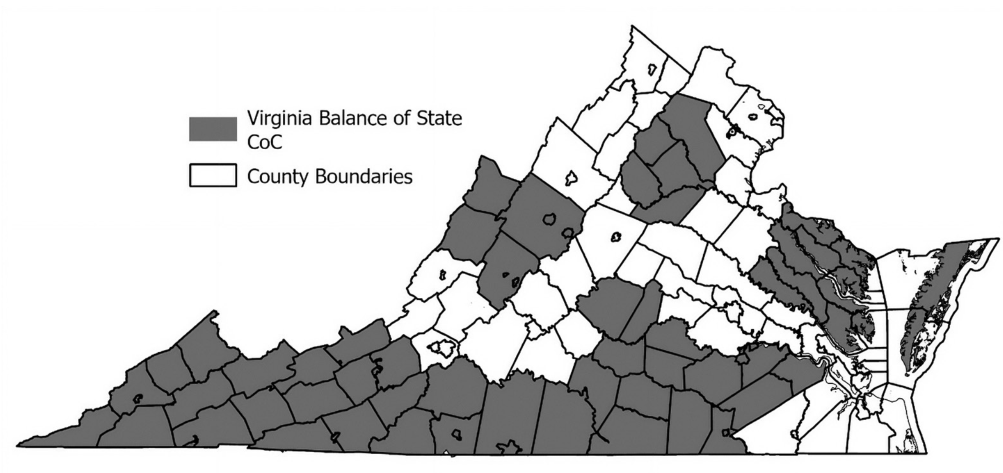
For this study, the authors focus on the same data but at a more local level: within the state of Virginia. Virginia's 16 COCs are county based, and each COC consists of one or more counties or

[1] HUD COC grantees collect PIT counts, which include sheltered and unsheltered homeless persons on a single night in the last 10 days of January for the year reported. More information is available here: https://www.hudexchange.info/programs/COC/COC-homeless-populations-and-subpopulations-reports/.

county equivalents (exhibit 1). For example, the Alexandria COC consists of one county equivalent (Alexandria City), whereas the Virginia Balance of State COC, or areas that are outside individual COCs, consists of 71 counties or county equivalents. In this article, the authors estimate COC-level veteran and nonveteran homelessness rates for Virginia in 2020.

Exhibit 1

Map of Counties and Virginia Balance of State Continuums of Care, 2020



CoC = Continuum of Care.
Source: HUD Office of Policy Development and Research

To the authors' knowledge, few studies focus on veteran and nonveteran homelessness rates at the COC level. One study used VA, HUD, and Census data to calculate risk ratios for veterans at risk of homelessness, using a convenience sample of data for 11 urban COCs, including New York, San Jose, Denver, and Phoenix (Fargo et al., 2012). The authors found that veterans were overrepresented in the homeless population and that risk was related to demographic characteristics such as age, sex, and race.

Data

The authors aggregated 2020 VA veteran population estimates for Virginia counties and county population estimates from the Census Bureau's Population Estimates Program to the COC level. The authors then estimated (1) COC nonveteran populations by subtracting VA veteran population estimates from Census Bureau total population estimates and (2) COC nonveteran homeless populations by subtracting HUD PIT veteran homeless counts from HUD PIT total homeless counts.

Homelessness rates are defined as the rate of homeless persons per 10,000 population. For example, in the Arlington County COC in 2020, the estimated veteran population was 16,843, and the estimated nonveteran population was 236,633. The estimated veteran homeless population was 3, and the estimated nonveteran homeless population was 196. The veteran homelessness rate equaled 10,000*3/16,843, which equals 1.781. The nonveteran homelessness rate equaled 10,000*196/ 236,633, which equals 8.283.

Exhibit 2 reports summary statistics for nonveteran and veteran homelessness rates and the percentage differences between the two rates.

Exhibit 2

Summary Statistics for Nonveteran and Veteran Homelessness Rates in Virginia, 2020

Variable	Mean	StdDev	Min	Median	Max
Nonveteran homelessness rate	7.987	3.527	2.921	7.722	16.140
Veteran homelessness rate	5.919	3.695	1.076	4.627	14.660
Percentage difference	20.122	58.296	– 155.222	41.724	78.496

Max = maximum. Min = minimum. StdDev = standard deviation.

Sources: Census Bureau Population Estimates Program, 2020; HUD Point-in-Time estimates, 2020; U.S. Department of Veterans Affairs, 2020

The percentage difference equals 100 multiplied by the difference between the nonveteran and veteran homelessness rates divided by the nonveteran homelessness rate. A positive percentage difference indicates that nonveteran homelessness rates are greater than veteran homelessness rates; a negative value indicates that veteran homelessness rates are higher than nonveteran homelessness rates.

For example, in the Virginia Beach COC, the estimated nonveteran homelessness rate equals 7.390, the estimated veteran homelessness rate equals 4.116, and the percentage difference equals 100*(7.390 - 4.116)/7.390, which equals 44.307 percent. So, for Virginia Beach, the nonveteran homelessness rate is 44.3 percent higher than the veteran homelessness rate.

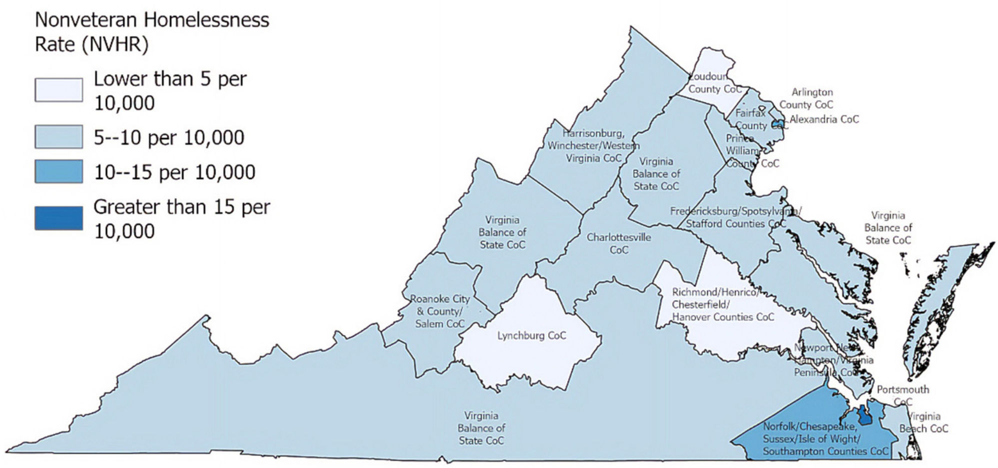
Nonveteran homelessness rates in Virginia range from 2.921 in the Lynchburg COC to 16.140 in the Portsmouth COC, with a mean of 7.987. Veteran homelessness rates range from 1.076 in the Lynchburg COC to 14.660 in the Roanoke City & County/Salem COC, with a mean of 5.919. Percentage differences between nonveteran and veteran rates range from -155.222 percent in the Richmond/Henrico, Chesterfield, Hanover Counties COC to 78.496 percent in the Arlington County COC, with a mean of 20.122 percent.

Data Visualizations

Exhibit 3 is a choropleth map of the nonveteran homelessness rate in Virginia. The majority of COCs experience rates between 5 and 10 per 10,000, illustrated by the medium-shaded COCs on the map. Three COCs—Loudon, Lynchburg, and the multicounty COC including Richmond and Henrico counties—had lower-than-average nonveteran homelessness rates, at lower than 5 per 10,000. Three COCs had the highest rates, greater than 10 per 10,000: Norfolk, Alexandria, and Portsmouth.

Exhibit 3

Choropleth Map of Nonveteran Homelessness Rate in Virginia, 2020



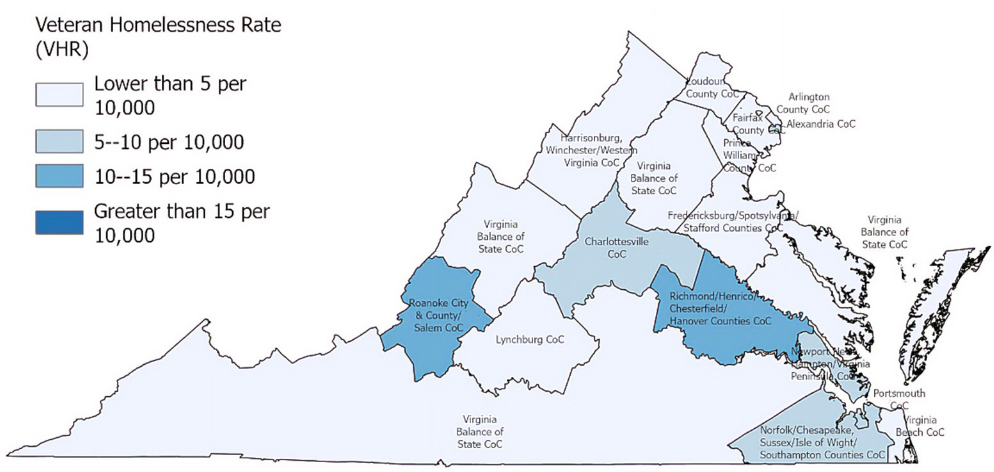
CoC = Continuum of Care. NVHR = nonveteran homelessness rate.

Sources: Census Bureau Population Estimates Program, 2020; HUD Point-in-Time estimates, 2020; U.S. Department of Veterans Affairs, 2020

Using the same choropleth scale and class breaks, exhibit 4 is a choropleth map for the veteran homelessness rate for COCs in Virginia. The maps show that most COCs, represented by the lighter-shaded areas, experience lower rates of homelessness for veterans, lower than 5 per 10,000 persons. Only two COCs, Roanoke and the multicounty COC that includes Richmond and Henrico, have rates higher than 10 per 10,000.

Exhibit 4

Choropleth Map of Veteran Homelessness Rate in Virginia, 2020



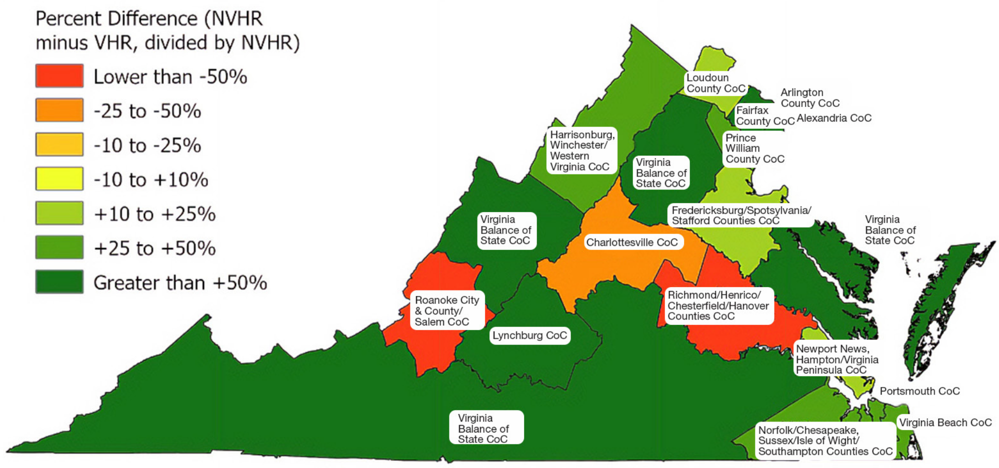
CoC = Continuum of Care. VHR = veteran homelessness rate.

Sources: Census Bureau Population Estimates Program, 2020; HUD Point-in-Time estimates, 2020; U.S. Department of Veterans Affairs, 2020

Exhibit 5 is a choropleth map depicting the percent difference between the two rates. As previously mentioned, a negative value indicates that the veteran homelessness rate is higher than the nonveteran homelessness rate. Exhibit 5 illustrates that three COCs have negative values: Roanoke, Charlottesville, and the multicounty COC that includes Richmond and Henrico; this is illustrated with the orange-shaded COCs. Maps such as these can help service providers and policymakers direct their attention and resources to veterans in those areas, which include Richmond, Virginia's capital, and Roanoke, the largest regional city in southwest Virginia.

Exhibit 5

Choropleth Map of Percent Difference between Nonveteran and Veteran Homelessness Rates in Virginia



CoC = Continuum of Care. NVHR = nonveteran homelessness rate. VHR = veteran homelessness rate.
Sources: Census Bureau Population Estimates Program, 2020; HUD Point-in-Time estimates, 2020; U.S. Department of Veterans Affairs, 2020

Within-state percentage differences for Virginia are in sharp contrast to the state-level results reported by Mast (2023), which indicated that the veteran homelessness rate in 2020 exceeded the nonveteran rate in 35 of 51 cases (the 50 states and the District of Columbia). This awareness can be important for veteran service providers to know in reviewing where resources are most needed at the local level.

Conclusion

The use of geospatial tools and mapping to address homelessness can help determine what areas and populations are most at risk of homelessness and can support efforts to prioritize housing assistance and related social services.[2] In this study, the authors focus on veteran homelessness,

[2] See for example, the Urban Institute's recent work on developing an Emergency Rental Assistance Priority Index: https://www.urban.org/data-tools/mapping-neighborhoods-highest-risk-housing-instability-and-homelessness?&utm_source=urban_newsletters&utm_campaign=DAU.

which often has a more specific set of causes, needs, solutions, and programs targeted at veterans, such as those at the VA and HUD.[3]

This study extends earlier work by Mast (2023), which used veteran population estimates. Mast demonstrated how calculating homelessness rates for both nonveteran and veteran populations can be used to compare and analyze differences across states. In this article, the authors use the same data but at a more local level, using HUD COC service areas in Virginia. Findings indicate that, on average, across Virginia, the veteran homelessness rate is lower compared with the nonveteran homelessness rate, with the exceptions of Roanoke, Charlottesville, and the multicounty COC that includes Richmond and Henrico Counties. Data visualizations such as this provide insight into local experiences of homeless veterans and nonveterans and contribute to the larger body of research methods that can improve local policies to address the issues of veteran homelessness.

Authors

Brent D. Mast is a social science analyst at HUD's Office of Policy Development and Research. Tricia Ruiz is a geographer at HUD's Office of Fair Housing and Equal Opportunity.

References

Fargo, Jamison, Stephen Metraux, Thomas Byrne, Ellen Munley, Ann Elizabeth Montgomery, Harlan Jones, George Sheldon, Vincent Kane, and Dennis Culhane. 2012. "Prevalence and Risk of Homelessness Among US Veterans," *Preventing Chronic Disease* 9: E45. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3337850/.

Mast, Brent D. 2023. "Veteran and Nonveteran Homelessness Rates: New Estimates," *Cityscape* 23 (2): 379–85. https://www.huduser.gov/portal/periodicals/cityscape/vol25num2/article15.html.

Tsai, Jack, and Robert A. Rosenheck. 2015. "Risk Factors for Homelessness Among US Veterans," *Epidemiologic Reviews* 37: 177–95. https://pubmed.ncbi.nlm.nih.gov/25595171/.

[3] Discover more about the unique characteristics of homeless veterans and resources to address their needs at the VA—https://www.va.gov/homeless/—and HUD—https://www.hud.gov/program_offices/comm_planning/veteranhomelessness.

Fewer Public Housing Units and a Greater Spatial Concentration of Housing Choice Voucher Households in the Tampa Metropolitan Statistical Area

Alexander Din
U.S. Department of Housing and Urban Development, Office of Policy Development and Research

Abstract

Between 2012 and 2022, most of the 50 most populous metropolitan statistical areas (MSAs) experienced increases in the concentration of households participating in the Housing Choice Voucher (HCV) Program locating into fewer census tracts. Examining a map for changes in HCV households for the Tampa, Florida, MSA, which saw the largest increase in concentration during that period, revealed that the decline in HCV program households by census tract was widespread, and increases occurred in only about one-third of census tracts, particularly near where public housing used to be.

Concentration of Housing Choice Voucher Program Households into Few Census Tracts

The potential for neighborhood choice and the spatial deconcentration of poverty has been a goal of the Housing Choice Voucher (HCV) Program since its inception, although the program has struggled to achieve either objective (Seicshnaydre, 2016). The program is not without its successes, however: households receive housing and are protected from homelessness (Ellen, 2020). Results are mixed regarding whether HCV households have better life outcomes if they participate in the spatial deconcentration of poverty and move to new neighborhoods (de Souza Briggs, Popkin, and Goering, 2010), but interest remains in HCV households accessing new neighborhoods—particularly households with children (Chetty, Hendren, and Katz, 2015).

Great effort has gone into measuring the spatial locations of HCV households. Analysts typically measure neighborhood access by describing the number of census tracts—neighborhood-level geographic units defined by the Census Bureau—although not all spatial patterns of clustering and dispersion may be captured through this simple method. Previous work (Devine et al., 2003; McClure, Schwartz, and Taghavi, 2015) has found, however, that HCV households are more concentrated than similarly low-income, non-assisted households (Metzger, 2015).

Analysis

Inequality of HCV Concentration by Census Tract by Metropolitan Statistical Area, 2012–22

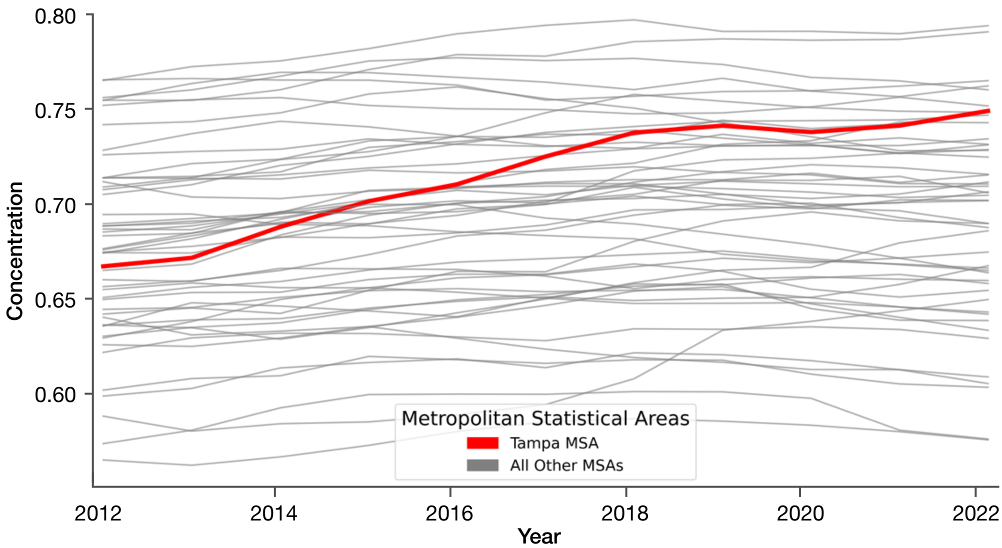
In this article, the author analyzes the Tampa–St. Petersburg–Clearwater Metropolitan Statistical Area (hereafter, Tampa MSA) because it was the metropolitan statistical area (MSA) with the greatest change in concentration of HCV households by census tract from 2012 to 2022 among the 50 most populous MSAs. The concentration and inequality of share of HCV households by census tract was measured for each MSA using the Gini coefficient. The Gini coefficient typically measures wealth inequality—how much wealth is concentrated among how many people—whereas, in this analysis, it measures how many HCV households are concentrated into how many census tracts. Metzger (2015) used a similar inequality index and found that, among similarly large MSAs, HCV households concentrate into fewer census tracts than do similarly low-income non-assisted renter households.

HCV household data between 2012 and 2022 in the 50 most populous MSAs in 2020 were geocoded to 2020 census tracts. The Gini coefficient produces values between 0.0, which would indicate equal dispersion of HCV households in all census tracts, and 1.0, which would indicate that one census tract contains all of the HCV households in an MSA. For each year and each MSA, the Gini coefficient was calculated describing the concentration of HCV households by census tract. Exhibit 1 displays Gini coefficients for each MSA by year as a line graph. Most (40 of 50) MSAs had increased Gini coefficients between 2012 and 2022, indicating that HCV households were in fewer census tracts. Using the Gini coefficient to measure HCV household inequality by census tract can help identify and compare areas where HCV households concentrate or disperse. Further analysis could employ more sophisticated spatial methods and statistics that are not directly comparable across different geographies due to a reliance on distances and spatial weights to inform those metrics.

In 2012, the Tampa MSA had an HCV Gini coefficient of 0.67 and ranked 30th among other large MSAs. However, by 2022, the Tampa MSA experienced the largest HCV Gini coefficient increase (0.08)—to 0.75—and ranked 8th among those same MSAs.

Exhibit 1

Demographic and Neighborhood Characteristics for Housing Choice Voucher Program Households in the Tampa MSA, 2012–22



MSA = metropolitan statistical area
Sources: HUD administrative tenant data; analysis by the author

Tampa, Florida Metropolitan Statistical Area

Households participating in the HCV Program are known to be clustered within the Tampa MSA (Walter and Wang, 2017). Critics of the HCV Program have noted that assisted households often move to similarly low-opportunity neighborhoods, where public housing was located (Greenbaum, Rodriguez, and Ward, 2008). Hillsborough County, which contains about one-half the Tampa MSA's population, has a high share of HCV households locating in Low-Income Housing Tax Credit properties, which may influence the concentration of HCV households into few census tracts (Williamson, Smith, and Strambi-Kramer, 2009).

The overall number of HCV households in the Tampa MSA grew by 18.1 percent between 2012 and 2022 (exhibit 2). The demographics of who was served by the HCV Program in the Tampa MSA also changed: fewer households have a female head-of-household, have children, or have a disabled head-of-household. By the end of the study period, the program had more project-based voucher (PBV) households, more specialty voucher types, and more elderly heads-of-household. Most of the PHAs in the Tampa MSA participate in the Rental Assistance Demonstration,[1] which may explain the increase in PBV households. Nearly 2,300 more specialty vouchers were in use, primarily Tax Credit Union (TCU) and Veterans Affairs Supportive Housing (VASH) vouchers. Few HCV households in the Tampa MSA lived in low-poverty census tracts before the beginning of the study period (McClure, 2013), and the average neighborhood poverty rate for an HCV household increased by 6.5 percentage points.

[1] See PHA Data at RAD Resource Desk, then view data using Extract Data to Excel: https://www.radresource.net/pha_data.cfm.

Exhibit 2

Change in Demographic and Neighborhood Characteristics for Housing Choice Voucher Program Households in the Tampa MSA, 2012–22

Variable	2012	2022
HCV Households	17,336	20,655
Project-Based Vouchers	1.3%	19.4%
Female Head-of-Household	84.6%	79.9%
Household with Children	44.3%	43.0%
Elderly Head-of-Household	17.1%	32.1%
Disabled Head-of-Household	26.4%	21.5%
Specialty Vouchers	3,172	5,458
Neighborhood Poverty Rate	16.4%	22.9%

HCV = housing choice voucher.

Sources: HUD administrative tenant data; HUD Picture of Subsidized Households; American Community Survey data; analysis by the author

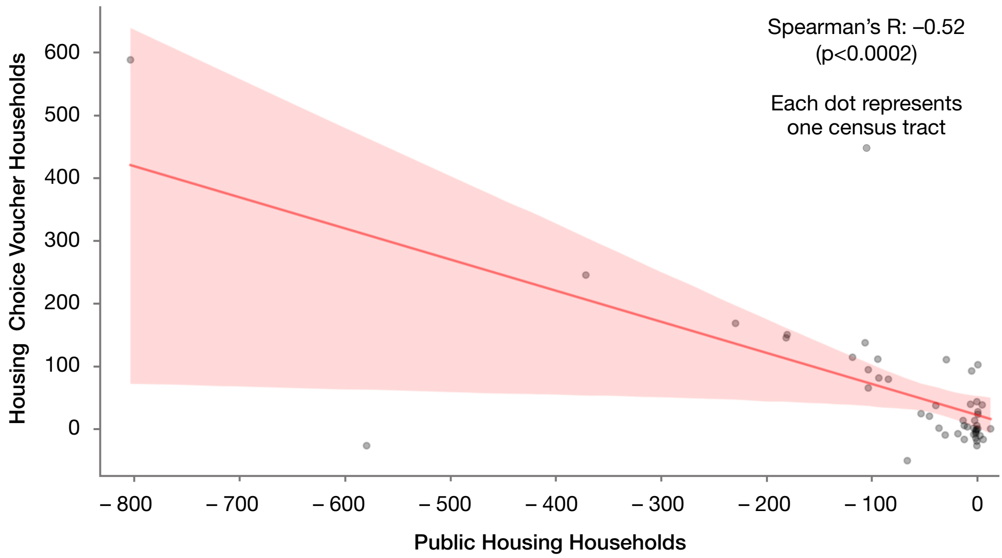
During that period, the number of HUD-assisted households participating in the public housing (PH) program declined from more than 4,500 to approximately 1,100—similar to the increase in the number of HCV households in the Tampa MSA during the same period. The analysis in this article did not track whether specific households moved from PH to the HCV Program, but the Tampa Housing Authority—the largest PHA in the Tampa MSA—participated in the Rental Assistance Demonstration (RAD) program,[2] and a feature of that program was the conversion of PH households to HCV households, including by way of PBVs.

In 2012, more than 4,500 households were served by the public housing program across 48 census tracts; by 2022, that number declined to approximately 1,100 households in only 24 census tracts. The census tracts that had at least one PH household in 2012 gained more than 2,800 HCV households, and, as shown in exhibit 3, there was a strong relationship between them losing PH households and gaining HCV households.

[2] For information on the RAD program in the Tampa MSA, please see: https://www.tampaha.org/rad-program.

Exhibit 3

Change in Public Housing and Housing Choice Voucher Households by Census Tract in the Tampa MSA, 2012–22



MSA = metropolitan statistical area

Note: Only census tracts with at least one public housing household are shown in the visualization (48 census tracts).

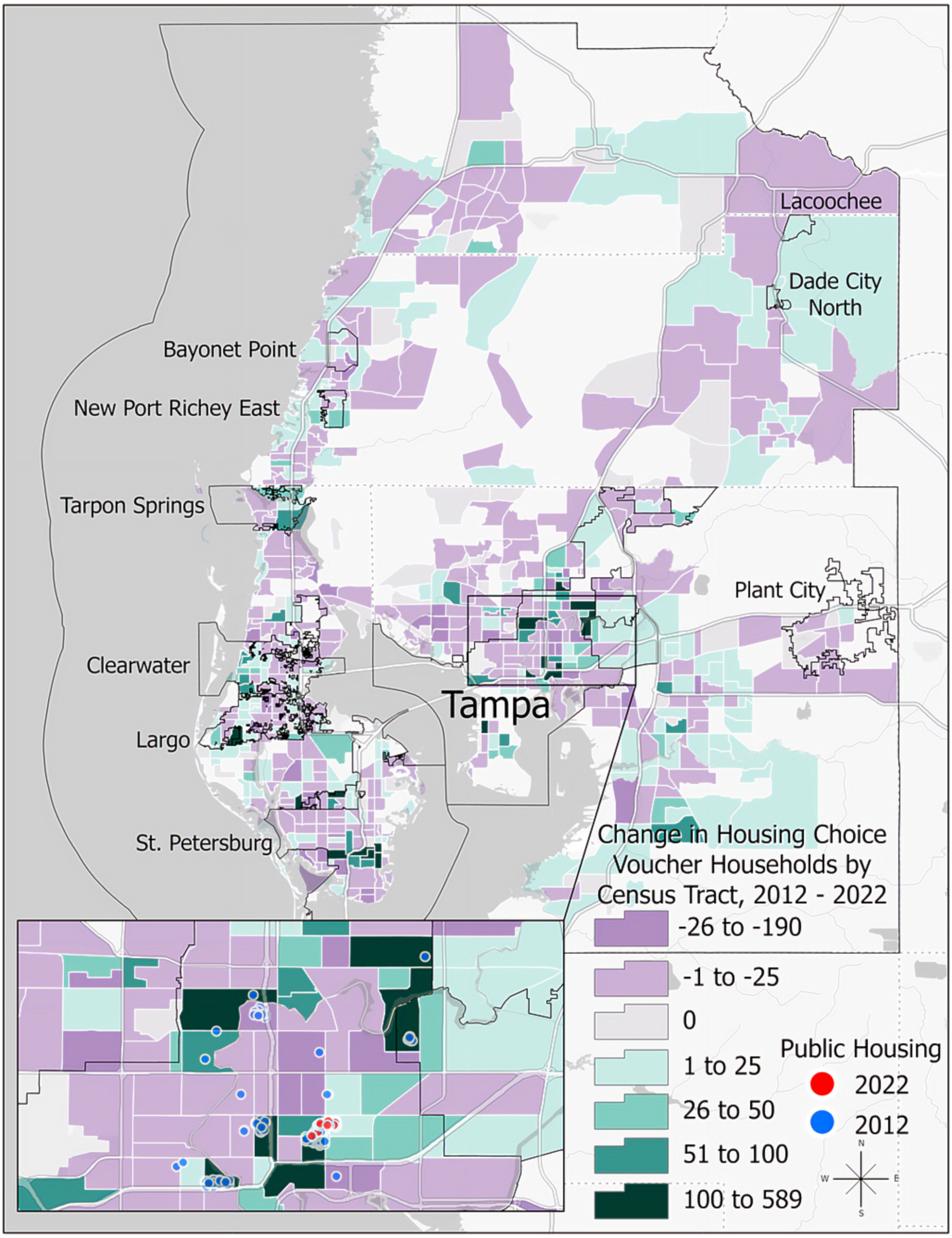
Sources: HUD administrative tenant data; analysis by the author

The map in exhibit 4 shows census tract-level changes in the number of HCV households in the Tampa MSA and includes an inset map around downtown Tampa that also shows public housing households in 2012 and 2022. Throughout the Tampa MSA are census tracts that lost at least one HCV household. Declines in the number of HCV households by census tract were widespread. One-half of the census tracts in the Tampa MSA that had at least one HCV household in 2012 had fewer in 2022, representing an overall loss of about 4,300 fewer HCV households. Increases in the number of HCV households were concentrated into about one-third of census tracts and totaled more than 7,600 more HCV households.

The inset map shows downtown Tampa, which had many census tracts with declining numbers of HCV households but also multiple census tracts with large increases. Overlaid are PH households in 2012 in 2022. Census tracts with high increasing numbers of HCP households in downtown Tampa tend to be located near where many PH households were in 2012 that were no longer present by 2022. Many PH developments existed throughout the downtown Tampa neighborhoods in 2012, but by 2022, only a cluster of PH households remained at Belmont Heights Estates. The map shows that census tracts that gained HCV households were typically located near where PH households used to live.

Exhibit 4

Census Tract-Level Change in the Housing Choice Voucher Program in the Tampa MSA, 2012–22



MSA = metropolitan statistical area

Note: Only Census-defined places that had Public Housing sites in 2012 are displayed on the map.

Sources: HUD administrative tenant data; analysis by the author

Is Further Spatial Concentration of Housing Choice Voucher Households a Fact of Program Changes?

The spatial deconcentration of poverty has long been a goal of HUD, and much analysis has been performed measuring the spatial concentration of HCV households. Examining the inequality of HCV households by census tracts revealed that most MSAs are experiencing further concentration of HCV households. The map of the Tampa MSA illustrates that most census tracts throughout the MSA were declining in the number of HCV households, and the census tracts gaining in HCV households tended to be near where PH households used to live. Despite changes in program type, when considering all HUD-assisted households regardless of program type, little change may have occurred in the concentration of low-income assisted rental households.

Further research should consider the effect of programmatic changes to HUD assistance programs and the impact of those changes on the spatial concentration of HUD-assisted households, particularly HCV households. For example, public housing authorities (PHAs) have so far expressed great interest in converting PH units to PBVs (Mast and Hardiman, 2017). Although changes to how low-income renter households are served may be necessary—whether due to insufficient funding for capital needs to the PH program (Hanlon, 2017) or tight rental markets (Galvez et al., 2020)—programmatic changes must be considered when measuring the spatial concentration of HCV households. Policymakers and stakeholders should also evaluate how changes to rental assistance programs may affect the goal of spatial deconcentration of poverty and whether the changes interrupt patterns of historic segregation and unequal fair housing practices. Further research also should identify success cases and best practices when program changes resulted in HCV households accessing new neighborhoods, not just neighborhoods near where public housing used to be.

Acknowledgments

The author thanks David Hardiman for his helpful comments on changes in the public housing and Housing Choice Voucher programs during the study period.

Author

Alexander Din is a social science analyst in the Office of Policy Development and Research at the U.S. Department of Housing and Urban Development.

References

Chetty, Raj, Nathaniel Hendren, and Lawrence F. Katz. 2015. The Effects of Exposure to Better Neighborhoods on Children: New Evidence from the Moving to Opportunity Experiment. NBER Working Paper No. 21156. Cambridge, MA: National Bureau of Economic Research.

de Souza Briggs, Xavier, Susan J Popkin, and John Goering. 2010. *Moving to Opportunity: The Story of an American Experiment to Fight Ghetto Poverty*. New York: Oxford University Press.

Devine, Deborah J., Robert W. Gray, Lester Rubin, and Lydia B. Taghavi. 2003. "Housing Choice Voucher Location Patterns: Implications for Participant and Neighborhood Welfare." Washington, DC: U.S. Department of Housing and Urban Development, Office of Policy Development and Research.

Ellen, Ingrid Gould. 2020. "What Do We Know About Housing Choice Vouchers?" *Regional Science and Urban Economics* 80:103380.

Galvez, Martha M., Daniel Teles, Alyse D. Oneto, and Matthew Gerken. 2020. "Moving to Work Agencies' Use of Project-Based Voucher Assistance," *Cityscape* 22 (3): 55–84.

Greenbaum, Susan, Cheryl Rodriguez, and Beverly G. Ward. 2008. "Displacement and Deconcentration in Tampa," *Anthropology News*, December 2008, 10–14.

Hanlon, James. 2017. "The Origins of the Rental Assistance Demonstration Program and the End of Public Housing," *Housing Policy Debate* 27 (4): 611–39.

Mast, Brent D., and David Hardiman. 2017. "Project-Based Vouchers," *Cityscape* 19 (2): 301–22.

McClure, Kirk. 2013. "Which Metropolitan Areas Work Best for Poverty Deconcentration with Housing Choice Vouchers?" *Cityscape* 15 (3): 209–36.

McClure, Kirk, Alex F. Schwartz, and Lydia B. Taghavi. 2015. "Housing Choice Voucher Location Patterns a Decade Later," *Housing Policy Debate* 25 (2): 215–33.

Metzger, Molly W. 2015. "The Reconcentration of Poverty: Patterns of Housing Voucher Use, 2000 to 2008," *Housing Policy Debate* 24 (3): 544–67.

Seicshnaydre, Stacy. 2016. "Missed Opportunity: Furthering Fair Housing in the Housing Choice Voucher Program," *Law and Contemporary Problems* 79: 173–97.

Walter, Rebecca J., and Ruoniu Wang. 2017. "A Research Note: The Housing Choice Voucher Program and Access to Opportunity in Florida's Nonmetropolitan Areas," *Cityscape* 19 (2): 239–54.

Williamson, Anne R., Marc T. Smith, and Marta Strambi-Kramer. 2009. "Housing Choice Vouchers, the Low-Income Housing Tax Credit, and the Federal Poverty Deconcentration Goal," *Urban Affairs Review* 45 (1): 119–32.

Industrial Revolution

Every home that is built is a representation of compromises made between different and often competing goals: comfort, convenience, durability, energy consumption, maintenance, construction costs, appearance, strength, community acceptance, and resale value. Consumers and developers tend to make tradeoffs among these goals with incomplete information which increases risks and slows the process of innovation in the housing industry. The slowing of innovation, in turn, negatively affects productivity, quality, performance, and value. This department features a few promising improvements to the U.S. housing stock, illustrating how advancements in housing technologies can play a vital role in transforming the industry in important ways.

A Study of Innovative Assistive Devices for Aging in Place

John B. Peavey
Pranav Phatak
Home Innovation Research Labs

Ed Steinfeld
University of Buffalo

Danise Levine
Center for Inclusive Design and Environmental Access

The U.S. Government does not endorse products or manufacturers. Trade or manufacturers' names appear herein solely because they are considered essential to the object of this article.

Abstract

Home Innovation Research Labs and the Center for Inclusive Design and Environmental Access (IDEA) partnered to study innovative, cost-effective assistive devices that improve accessibility in townhomes and row houses without requiring major renovations. Study participants include seniors, persons with disabilities, caregivers, and design professionals specializing in accessibility.

The U.S. Department of Housing and Urban Development (HUD) identified townhomes and row houses as needing innovative solutions to accommodate the growing population experiencing physical, mental, or sensory challenges. Several programs exist that provide retrofits to existing homes, but in many

Abstract (continued)

cases, the cost of renovating can be prohibitive due to narrow hallways, multilevel stairs, and the lack of bedrooms and full bathrooms on the first floor.

This study addresses three key issues: (1) identifying home improvements needed to make these homes more accessible for individuals with mobility-related limitations; (2) evaluating innovative assistive devices; and (3) determining the availability and cost of such devices.

Introduction

The United States has a large housing stock of attached and semi-attached residential buildings (townhomes and row houses). Narrow floor plans, functional areas spread among levels, and elevated entrances are particularly challenging for seniors and people with disabilities.

The U.S. Department of Housing and Urban Development (HUD) supports affordable housing through many programs, including *Section 202* Supportive Housing for the Elderly and *Section 811* Supportive Housing for Persons with Disabilities. HUD's Office of Lead Hazard Control and Healthy Homes also provides grants focused on low-cost, high-impact modifications through the Older Adult Home Modification Program (OAHMP). According to HUD (n.d.), examples include—

> [I]nstallation of grab bars, railings, and lever-handled doorknobs and faucets . . . installation of adaptive equipment, such as temporary ramp, tub/shower transfer bench, handheld shower head, raised toilet seat, risers for chairs and sofas, and non-slip strips for tub/shower or stairs. The OAHMP model primarily relies on the expertise of a licensed Occupational Therapist (OT) to ensure that the home modification addresses the client's specific goals and needs and promotes their full participation in daily life activities.[1]

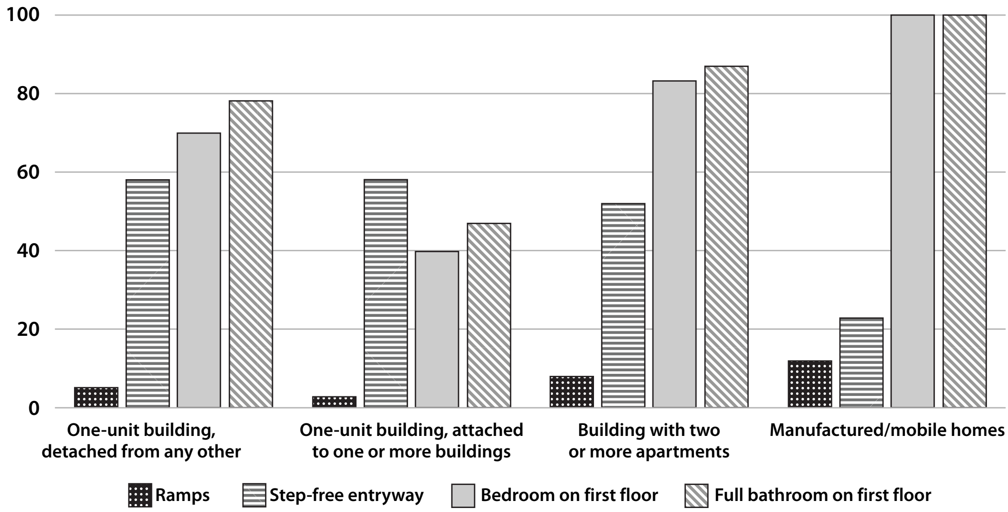
In fiscal year 2022, OAHMP received $30 million in funding. To qualify for the program and receive benefits, applicants must be low-income homeowners who are at least 62 years old with a privately owned primary residence that needs repair or rehabilitation (beyond normal maintenance).

The U.S. Census Bureau and HUD (2021) published the 2019 American Housing Survey (AHS), which includes data concerning home accessibility. The AHS estimated that 48 million households (39 percent of the total households in the United States) have at least one person older than age 65, with a disability, or both. Of people living alone, approximately 35 percent have a disability, and 37 percent are aged 65 or older. The AHS estimated occupied housing units with accessibility features as a percentage of each structure type, demonstrated in exhibit 1.

[1] An overview of OAHMP is available at https://www.hud.gov/program_offices/spm/gmomgmt/grantsinfo/fundingopps/oahmp.

Exhibit 1

Occupied Housing Units with Accessibility Features as a Percentage of Each Building Structure Type



Source: U.S. Census Bureau

Exhibit 1 considers attached and semi-attached houses (townhomes and row houses) to be "One-unit building, attached to one or more buildings." The research team could not identify a single "typical" design for these homes. Instead, the team identified several design types and developed lists of modifications needed to make the units accessible.

Existing Home Modifications Needed for Accessibility

The existing housing stock of attached and semi-attached houses has a wide variety of floor plans based on regional design differences and the age of the homes. The team identified three design features—elevated entryways, garages occupying most of the ground level, and stairs to access multiple floors—that make accessibility challenging.

Raised entryways may present a challenge to accessibility, especially when set-back space in the front of the property is limited. The space constraints in exhibits 2 and 3 make installing a conventional ramp impossible. In such cases, installing a stair lift may be necessary, or the resident may need to relocate if no retrofit solutions are viable. Exhibit 4 illustrates a common design feature of modern townhomes—a garage that occupies most of the ground floor space that requires livable space features, such as a full bedroom and bathroom, to be on the second and third floors of the residence.

Exhibit 2

Raised Entryway: Baltimore Rowhouse

Photo credit: Pinterest.

Exhibit 3

Raised Entryway: New York City Brownstone

Photo credit: Shutterstock.

Exhibit 4

Garage First Level: Virginia Townhouse

Photo credit: Pinterest.

As exhibit 1 shows, the Census Bureau and HUD identified and tracked four key accessibility features, which include ramps, step-free entryways, bedrooms on the first floor, and full bathrooms on the first floor. Exhibit 5 illustrates a typical floor design with the garage on the first floor.

If a resident has a mobility-related disability, three options must be considered. First, the homeowner may do a major renovation to create a bedroom and full bathroom on the first floor. Second, they may install a stair assistive device from the garage to the first floor and another from the first floor to the second floor. Third, the homeowner—or resident with a disability—may move to a different house or assisted-living facility. This study focuses on identifying and evaluating devices that make staying in the home both affordable and accessible.

Exhibit 5

Existing Garage on First Floor (No Bedroom or Full Bathroom)

Source: Home Innovation

Evaluating Innovative Assistive Devices

A large percentage of existing townhomes and rowhouses have bedrooms and bathrooms on the second or third level, making vertical navigation a daily challenge. With vertical mobility in mind, the Home Innovation research team evaluated innovative products that make stairways safer and easier to climb, including easier-to-grip railings and devices such as StairSteady and AssiStep. Initially, evaluating a new lift or elevator technology was expected to be outside the scope due to budget, time, and engineering constraints. However, the team identified a cost-effective option, "FlexStep," a lift product manufactured in Europe. The team also secured an easy-to-use door called "All-Push Door," specifically designed for individuals with difficulty pulling open doors, a common difficulty of those in wheelchairs.

Using Focus Groups

Focus group discussions were conducted among three primary stakeholders—users, caregivers, and professionals. Each focus group discussion lasted approximately 2 hours, and each group had 8 to 10 participants, with 27 total participants.[2] Although extrapolation from this sample size may not be appropriate statistically, the findings from the focus groups corroborate the key findings from the literature search; for example, (1) the appearance of the assistive device is important and must fit with the home's character; (2) assistive devices require a standard method of evaluation to increase trust of the device's safety; and (3) affordability is a primary consideration.

Caregivers in the focus groups were more likely to report challenges than the users or people with disabilities. Caregivers tended to categorize an activity as challenging if they perceived any delay or difficulty, whereas users of the device resisted classifying an activity as a challenge as long as they could ultimately achieve the goal despite delay or difficulty. In every category, the caregivers consistently rated the activity more difficult than the users. Despite this difference, caregivers and users prioritized challenges the same.

Areas of the home such as bathrooms, stairs, laundry rooms, and entryways or entrances were commonly cited as presenting the greatest challenges. Among caregivers and users, the fear of falling was a primary safety concern associated with the bathroom and stairs. Users thought that these areas needed to be modified for them to remain in their homes long term, and caregivers stressed the importance of having good retrofit solutions for these spaces. Caregivers prioritized stairway solutions because they improved mobility inside the house, prevented users from feeling isolated or confined to one part of the house, and allowed mobility-facilitated social interaction.

Occupational therapists in the focus groups stated that individuals who stayed in their homes saw greater benefits in terms of quality of life and added value to their homes if a major renovation expanded the function of the first floor. Most residents prefer to stay in their homes (Goyer, 2021), and professionals have seen an increased demand for aging-in-place solutions.[3]

Remodelers, builders, and architects participating in the research project claimed to encourage their clients to proactively plan for future needs when planning general remodels or designing new homes.[4] Those professionals have found that their clients were more accepting of solutions that highlight empowerment and improve quality of life.

Professionals generally ranked exterior access solutions as the most challenging to implement. In some situations, installing a ramp was not possible due to space limitations. An occupational therapist in the group confirmed this challenge, sharing an example of a client who needed a ramp

[2] The focus group participants were paid an honorarium of $150 in the form of a gift card.

[3] *Aging in place* refers to people remaining in their homes without losing their quality of life as they age.

[4] The remodelers, builders, and architects who participated in this study were accessibility specialists, hence they often recommended Universal Design and accessibility features to their clients.

or lift to access their home, but the entryway did not have space for a ramp, and the cost of a lift was prohibitive (from $15,000 to $20,000). The issue ultimately forced the individual to move.

Professionals agreed safety was a fundamental consideration for all home modifications. Fall prevention was the primary concern, and bathrooms, stairs, door thresholds, and floor material transitions were identified as the areas with the highest probability of falls.

Professionals believed that aesthetics were very important to encourage proactive planning with their clients. Although everyone valued safety and ease of use, their clients also wanted their homes to look and feel like "a home," not a hospital or assisted-living facility. Architects, designers, and remodelers were keenly aware that if all devices were equal in performance, clients preferred the best-looking device.

Across the focus groups, affordability was perceived as the number one barrier to retrofits and renovations. Despite available grant money, rebates, and other funding sources, residents often struggled to get financial assistance. Income or asset thresholds, lack of homeownership—generally a requirement for many of the programs—or a complex application process can prevent residents from qualifying for funding. Even when funding is secured, it may not be enough to cover the total cost. For townhomes and row houses, various space limitations complicate retrofits and renovations. To address those constraints, professionals have proposed complex design solutions that include more expensive devices such as elevators or major renovations.

Selecting the Assistive Devices

Based on the focus group discussions, Home Innovation and Inclusive Design and Environmental Access (IDEA) conducted several brainstorming sessions to identify innovative product solutions that could address the following accessibility challenges typical of townhomes and row houses:

- Limited space.
- Small rooms.
- Narrow hallways.
- Stairways, accessing functional areas on multiple levels of the house.
- Elevated exterior door entry.
- Shared walls between the houses (limits renovation options).
- Small lots with limited space between the house and property line.

From the brainstorming sessions, the team considered four assistive devices (StairSteady, AssiStep, FlexStep, and All-Push Door). StairSteady was selected because it offers a simple, relatively inexpensive assistive device for climbing stairs. It is compact and easy to install but requires some upper body strength to use the manual device. The device is affordable, low maintenance, and compact enough to be concealed when not in use. Because the product can be painted any color, it can easily match the surrounding decor of virtually any home. It is manufactured in Canada and must be special ordered because no U.S. distributors carry it.

AssiStep (similar to the StairSteady device) was chosen as an alternative manual stair-assistive device. It has an adjustable handle and can be installed on stairways with limited space. AssiStep is mounted on a track that requires very little upper body strength to move. The product is affordable, low maintenance, and can be folded up when not in use. It is manufactured in Europe and must be special ordered because no U.S. distributors carry it.

FlexStep was selected because it converts from a lift to a staircase, making it a potential solution for limited space at the main entrance. The product is also well suited for interior doors that enter the house from the garage. It received a high ranking due to its potential to address stairs, entryways, and limited spaces. However, a preliminary review of its cost revealed that it may not be affordable. The product is manufactured in Europe but is readily available for purchase in the United States through distributors.

The All-Push Door was selected because it can be adapted to an exterior or interior door opening. The "push-only" concept was designed for people who may have issues pulling a door open due to cognitive impairment or physical limitations in their hands. The product also eliminates the need for someone in a wheelchair to turn around when entering a door from the pull direction. The product is manufactured in Taiwan and has limited availability in the United States, but the manufacturer plans to have products available to meet anticipated future demand.

In-Person Evaluations of Assistive Devices

Home Innovation recruited 51 participants—users, caregivers, and professionals—to evaluate the assistive devices. The participants included users with mobility-related disabilities or limitations, caregivers of older adults or individuals with mobility-related disabilities or limitations, and professionals with expertise in accessible design, including physical and occupational therapists, architects, designers, and contractors specializing in accessibility renovations.

Participants were given a short orientation that included a review of the evaluation process and a demonstration of how to use each device (exhibits 6 and 7). Participants were instructed to evaluate each device and complete a survey. Next, participants engaged in open-ended discussions with Home Innovation staff, which led to further insight into participants' experiences.

Exhibit 6

StairSteady, AssiStep, and FlexStep Product Test Platform

Photo credit: Home Innovation.

Exhibit 7

All-Push Door Product Test Doorway

Photo credit: Home Innovation.

None of the equipment evaluated are considered do-it-yourself devices. Manufacturers recommend using a remodeler or contractor to ensure correct installation. The StairSteady, AssiStep, and FlexStep devices were installed on the demonstration platform in exhibit 6. A separate doorway mock-up structure was built to install the *All-Push Door* in exhibit 7.

Home Innovation did not conduct performance testing on the accessibility devices and does not endorse or debunk any manufacturer's claims. Instead, Home Innovation captured the participants' perceptions and comments through observational research.

Observational Research Findings for Assistive Devices

The observational research conducted in Home Innovation's laboratory allows large-scale field studies of people interacting with tools, materials, products, or other people while maintaining control and predictability only possible in a structured setting. The product test platforms created for this study are a "true-to-life mock" scenario of a home interior in which participants are led through activities and discussions under the discrete observation of market research professionals. This set-up provides the authenticity and deep insight of ethnographic research with an efficient use of time and resources.

StairSteady Evaluation Results

Participants considered StairSteady easy to install, clean, maintain, and match with any home decor—provided the device was painted. All found the instructions easy to understand, but most thought the device handle required some practice. During the evaluation, the handle did not glide smoothly and required maneuvering to shift from one position to the next. As a result, most participants "had problems using the device" and believed that others would have the same problems.

The professionals and the users and caregivers disagreed about the physical effort needed to use StairSteady. Nearly all professionals thought that the upper body strength needed would make the device difficult to use for most frail senior users or any client recovering from a stroke. When answering this question, the physical and occupational therapists considered a wider range of potential users, not only those individuals with mobility-related disabilities. The users and caregivers were divided about the physical effort necessary. Some believed it would be easier with practice, and others thought the device was not positioned correctly for their body size and stature.

The 51 participants for the in-person evaluation varied in size and stature. StairSteady was installed according to the manufacturer's instructions but could not be adjusted for each participant and is not designed to be adjusted after installation. The in-person evaluations proved that the device must be installed based on each user's physical measurements. Users and caregivers who required little effort to use the device were simply fortunate enough to fit the installation of StairSteady. For taller users, the handle did not line up at 90 degrees with the person's elbow (exhibits 8 and 9), causing the user to push the handle at an angle that kept it from moving smoothly. Furthermore, the person's center of gravity was typically over the handle coming down the stairs (exhibit 10), making it less safe.

Exhibit 8

Height of Handle: Starting Position

Photo credit: Home Innovation.

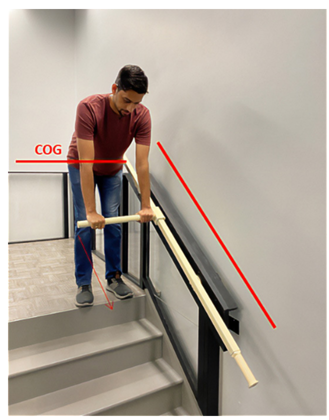
Exhibit 9

Adjust to 90 Degrees: Ascending Stairs

Photo credit: Home Innovation.

Exhibit 10

Handle Too Low: Descending Stairs



COG = center of gravity.
Photo credit: Home Innovation.

The device was determined to be too low for many participants who tried to use it—an unavoidable outcome given the random nature of recruiting participants—and may partially explain some of the poor ratings StairSteady received.

Although StairSteady is straightforward to install, it must be mounted at the correct height for the user. One physical therapist regarded StairSteady as a quasi-medical device that a trained professional should install to ensure that the consumer can use it correctly. One design professional thought that the device's simplicity leads do-it-yourself installers to believe that they can install the device—only to mount it incorrectly.

The professionals were less likely than the users and caregivers to believe that the device was needed. In some cases, users thought they did not need it because they could still navigate the stairs using a cane. Nonetheless, some professionals thought the device needed improvements before they would recommend it.

AssiStep Evaluation Results

Participants' thoughts on AssiStep were similar to those provided for StairSteady. They considered AssiStep easy to install and clean but not easy to match with any home decor. Most participants believed that AssiStep required little maintenance, even though the rail has a sophisticated metal track with special gears inside. All found AssiStep instructions easy to understand, but most thought the device required some practice to use correctly. During the evaluation, most participants found going up the stairs very easy because of the special track on which the handle glides. However, an equal number of participants struggled with disengaging the handle from the track when going down the stairs. As a result, participants were divided about whether they "had problems using the device."

Nearly all professionals thought that others would have problems with AssiStep, even when they did not have difficulty using the device. However, the users and caregivers typically supposed, "If I can do it, anyone can," or "If I had a problem, someone else will have a problem too."

AssiStep was installed according to the manufacturer's instructions. The device's handle can be adjusted vertically after installation on the metal track. No handle adjustments were made for participants, and the handle position was not considered an issue. Instead, most participants were concerned with (1) how to disengage the handle when going downstairs and (2) how to fold the device away when it is not being used.

Participants could hold one of two locations (the upper bar or lower bar) of the handle while allowing the standing user's elbow to be at a 90-degree angle (exhibits 11 and 12).

Exhibit 11

Upper and Lower Bars: Select Bar Position

Photo credit: Home Innovation.

Exhibit 12

Correct Position of Handle: Ready to Ascend Stairs

Photo credit: Home Innovation.

AssiStep's handle can be adjusted up and down, as exhibit 13 shows. When walking down the stairs, the device's handle must be lifted to disengage the metal track, which some professionals thought would be difficult for senior users. Users and caregivers were less concerned, believing that, with practice, a person could safely disengage the handle from the track. As exhibit 14 shows, the person's center of gravity is not over the handle.

Exhibit 13

Adjusting the Handle and Disengaging the Track: Highly Adjustable Design

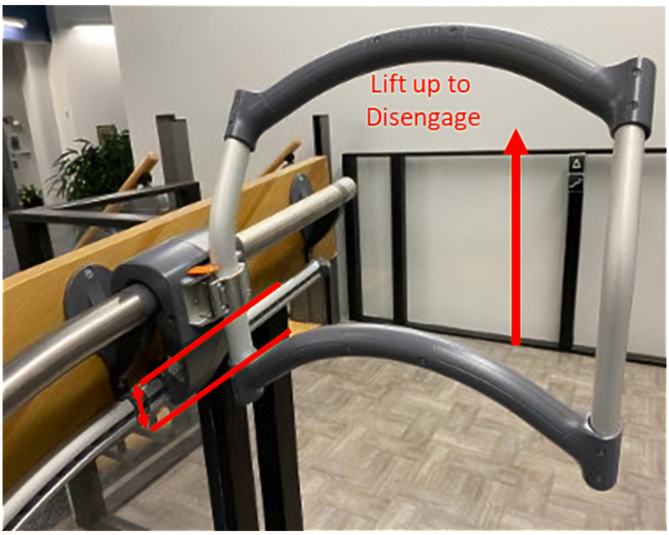


Photo credit: Home Innovation.

Exhibit 14

Descending the Stairs: Can Adjust Without Remounting

COG = center of gravity.
Photo credit: Home Innovation.

The professionals were less likely to believe that the device was needed than the users and caregivers. In some cases, users thought it was not necessary because they could still navigate the stairs using canes. Some professionals said that the device needed an additional safety mechanism to lock in an emergency—in case the handle was disengaged from the track—before they would recommend it.

FlexStep Evaluation Results

Participants considered FlexStep easy to clean, nice-looking, and easy to match with any home decor. All participants found the instructions easy to understand but also thought the control buttons on the device required some practice to use correctly. During the evaluation, most participants had to try the control buttons several times to understand their operation. As a result, most "needed assistance to use FlexStep." They believed that others would need formal training to remember what buttons to push and what to do in a power outage. FlexStep comes with a backup battery, but many participants worried that power would go out and they would be trapped on an upper floor in the house if they depended solely on FlexStep.

All participants agreed that installing FlexStep would not be easy. They assumed that the user would need to hire a professional to complete the installation and provide training to use the device correctly. In addition, all participants were concerned about maintenance and repair options for the FlexStep. Users and caregivers thought that finding service for general maintenance would be difficult. Some respondents also expressed concern about maintenance and repair costs. Many thought that if the purchase of FlexStep was expensive, then maintenance and repair would likely be equally expensive. The manufacturer stated that authorized dealers throughout the United States offer maintenance plans and that FlexStep would not need annual maintenance. The company recommends maintenance every 5 years.

Some participants complained about the noisy operation and audible safety warnings from FlexStep. When in operation, the device made multiple "beeping" noises and flashed lights. The manufacturer stated that many warning cues can be disabled, and the volume can be adjusted. The automatic sensors and safety features on the FlexStep were well received—the device is designed to prevent pets and objects from getting trapped under the stairs as it converts to a lift (exhibits 15 through 17).

Exhibit 15

Stair Mode: Starting Position

Photo credit: Home Innovation.

Exhibit 16

Lifting Mode: Intermediate Position

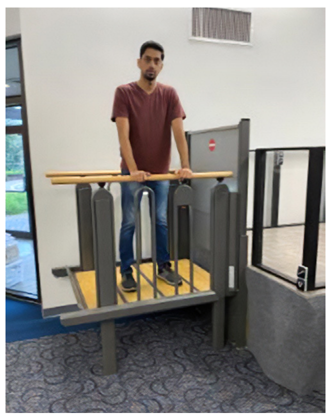


Photo credit: Home Innovation.

Exhibit 17

Final Landing Position: Finished Position

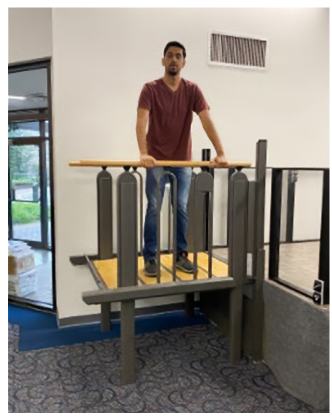


Photo credit: Home Innovation.

Most users and caregivers who did not want FlexStep thought it would not fit in their house due to the layout of their exterior or interior stairs, and they were concerned about the cost of a major renovation to use the device. Many participants thought that FlexStep would be impractical for interior floor spaces but that it could work at entryways where short flights of steps were common.

FlexStep can accommodate a wide range of individuals, including users with moderate mobility limitations and those who use wheelchairs. Participants liked that people of different sizes and statures could use FlexStep without special configurations.

All-Push Door Evaluation Results

Participants considered the All-Push Door easy to install, clean, and match to any home decor. Most participants believed that it required little maintenance. All found the instructions easy to understand but also thought the locking mechanism on the door should be easier to operate. Some participants thought the multiple locks made it difficult to know whether the door was locked. Most participants had to try the locking sequence several times to understand how to operate it. Therefore, most professionals "needed assistance to use the All Push Door." In addition, many respondents "made mistakes that required them to redo some steps." Participants believed that others would need some practice using the double set of locks.

Participants disagreed about how much assistance was needed to operate the All-Push Door. Users and caregivers appeared to make fewer mistakes operating the door than did the professionals. In fact, most users and caregivers did not need much assistance to use the All Push Door.

Many professionals thought that the small locking mechanisms of the device could present a problem to those individuals with arthritis or other hand grasping issues. Many professionals also thought that some senior clients or patients could accidentally leave one of the multiple locks open.

The All-Push Door has a double-hinge mechanism, allowing the door to be pushed from either direction. In exhibit 18, the door is being pushed from the inside. In exhibit 19, the door is being pushed from the outside.

Exhibit 18

Outer Door Frame Operation: Exiting the House

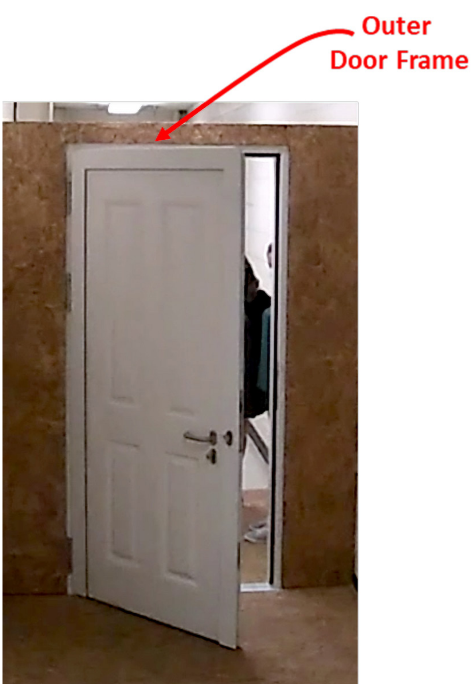


Photo credit: Home Innovation.

Exhibit 19

Inner Door Frame Operation: Entering the House

Photo credit: Home Innovation.

Participants believed that the All-Push Door addresses the problem of having to back up in a wheelchair when approaching a door that opens toward the person. Whether the individual is in a wheelchair, using a walker or cane, or simply having a mobility issue, being able to push the door open from either side is believed to be a "great benefit."

Although operating the door was not completely intuitive due to the double set of locks (best seen in exhibit 18 near the door handle), participants saw it as a minor issue that could be overcome with some instruction and practice.

The manufacturer claims that the door can be made available in many styles to fit almost any interior or exterior decor. Other features such as peepholes, glass panes, and keypad entries instead of a lock and key can be easily added to the All-Push Door. The door is not widely available in the United States, but the manufacturer plans to distribute the product in the United States soon.

Availability and Costs of Innovative Assistive Devices

All four assistive devices in this study were manufactured outside the United States, meaning their availability was limited, and additional costs may be associated with shipping. European countries and Canada have developed assistive devices that install on stairs and doors at a much greater frequency than the United States. Most products in this study must be purchased directly through the manufacturers' websites, with turnaround times of several weeks. Purchasers of StairSteady, AssiStep, and All-Push Door must hire a remodeler to install the products. FlexStep distributors are in the United States, so the product can be readily found, and the distributor can provide installation, maintenance, and warranty services. The All-Push Door manufacturer is interested in making the product in the United States, which may lead to lower costs and wider availability in the long term.

StairSteady costs less than a conventional stair lift product, and AssiStep is similar in cost to a conventional stair lift product. FlexStep is more expensive than a similarly sized wheelchair lift device but has the added benefit of converting into a flight of stairs and superior aesthetics, which users of accessibility devices prefer. The All-Push Door is comparable in cost to an automatic door opener, but it can easily exceed that cost if installation requires a major door frame modification.

Conclusion

Study participants expressed a strong willingness to adapt their homes to meet their needs. An inherent value is associated with staying in one's home. It provides a sense of independence and well-being. The accessibility devices evaluated in this study offer some new options, but in terms of performance, the manual devices are not "game changers." Nonetheless, StairSteady is the most affordable option, but only individuals with some remaining upper body strength can use it. If the person suffers from a degenerative disease or osteoarthritis, using a manual stair-assistive device may cause the person's joints to deteriorate faster. AssiStep is also a manual device for stairs, and it is somewhat easier to use, requiring less upper body strength. FlexStep is innovative because it can transform from a lift to a flight of stairs. The price is about 30 percent more than a traditional chair lift, but the innovation may be worth that premium because it eliminates the need for a lift *and* a separate set of stairs. The All Push Door, a manual door that can be "pushed open" in both directions, is comparable in price to an automatic door opening device.

In closing, the four innovative products in this study show that many opportunities exist for further innovation. They also demonstrate the economic value of government support for home modification services. Such support may spur further innovation, improve product quality, and reduce the cost of retrofits and adaptations.

Acknowledgments

The authors gratefully acknowledge the financial support HUD provided for this research under grant number H21673CA. Further, the authors thank Jagruti Rekhi for managing and coordinating this work on HUD's behalf. Finally, the authors thank the editors and two referees for their helpful comments.

Authors

John B. Peavey, PE, is the director of building science at Home Innovation Research Labs. Pranav Phatak, EIT, is a research engineer at Home Innovation Research Labs. Ed Steinfeld, ArchD, AIA, is the affiliation SUNY distinguished professor at the University of Buffalo. Danise Levine, M.Arch., AIA, R.A. CAPS, is the affiliation assistant director at the Center for Inclusive Design and Environmental Access.

References

Goyer, Amy. 2021. "Resources, Gadgets and Tools to Care for an Older Adult at Home." Washington, DC: AARP. https://www.aarp.org/caregiving/home-care/info-2021/tips-for-aging-independently.html.

U.S. Census Bureau and U.S. Department of Housing and Urban Development (HUD). 2021. "2019 Home Accessibility." 2019 American Housing Survey. https://www.census.gov/library/visualizations/2021/demo/2019-home-accessibility.html.

U.S. Department of Housing and Urban Development (HUD). n.d. "Older Adult Homes Modification Program." https://www.hud.gov/program_offices/spm/gmomgmt/grantsinfo/fundingopps/oahmp.

Policy Briefs

The Policy Briefs department summarizes new, emerging, or overlooked topics in housing and community development from a neutral point of view. If you have a submission or proposal, please contact the editor at David.L.Hardiman@hud.gov.

Point Access Block Building Design: Options for Building More Single-Stair Apartment Buildings in North America

Stephen Smith
Center for Building in North America

Eduardo Mendoza
Livable Communities Initiative

Abstract

Because of historical circumstances, building codes in the United States and Canada have stricter provisions around the number of stairways required in multifamily buildings above certain building heights than codes in other countries. Most building codes in the United States, with some exceptions, require multiple stairways and exits in multifamily buildings with four stories or more. These requirements pose a problem for smaller scale developments, particularly on small infill lots and when trying to build family-sized apartments. Within the United States, the cities of Seattle, Honolulu, and New York already have more flexible exit rules for mid-rise buildings, allowing "point access block" (sometimes referred to as "vertical shared access") buildings of up to six stories with a single stairwell, with additional fire and life safety restrictions not otherwise required for buildings of such size. Legislation has also been introduced or passed in several states to study the issues involved.

Introduction

Residents of multifamily buildings typically think of their homes in terms of what lies within their apartments, but to an architect, circulation—the ways residents move in, around, and out of a building—comes first. Circulation can take many forms, but one element that every multistory building must have is stairs. The number and arrangement of stairs and the hallways that lead to and from them are surprisingly important in defining the rest of the building—where windows can be placed, how rooms can be used, how big apartments can be, how much space everything consumes, how much the building costs to build, how high the rent or sales price is, and, ultimately, where people live.

In the United States and Canada, longstanding rules around means of egress—or, put simply, stairs—drive many of these considerations. Unlike building codes in most of the rest of the developed world, and even the National Fire Protection Association's (NFPA) model building code, most adopted codes in the United States dictate that even small apartment buildings above three stories (or two stories in Canada) must have two stairs, with a minimum distance between their doorways, as exhibit 1 shows. Combined with a suite of other rules, the result is that apartment buildings are typically laid out as what architects refer to as double-loaded corridors—what people in other countries think of as hotels—with small apartments (often as many studios and one-bedrooms as the zoning code will allow) flanking a central hallway. Efficient and affordable family-sized apartments found in the rest of the world—and in shorter and older apartment buildings in the United States—have become harder to build.

Exhibit 1

Plan of a Typical Floor From a Small Building in Jersey City, New Jersey

Notes: Small buildings in New Jersey typically require two stairs per building. In this example, the second stairway (in yellow) consumes 7 percent of the space. An additional 2 percent of the space (in green) is rendered unrentable, because it must sit in the common corridor rather than inside the apartment on the right. Eliminating the second stairway would cut construction costs by an amount roughly proportional to the floor area eliminated and also offer better light to the two rooms facing the interior courtyard and a private entryway to the two-bedroom apartment on the right.

Source: Plan redrawn and analyzed by Alfred Twu

History

In the early hours of Sunday, September 2, 1666, a fire broke out in a bakery on Pudding Lane. The fire burned for 4 days, and by the time it was finally extinguished, it had burned through four-fifths of London (LFB Museum, n.d.). The dominant use of wood as a building material contributed to the spread of the fire, with English diarist John Evelyn (1661) calling London a "wooden, northern, and inartificial congestion of Houses." London was quick to respond to the fire, and the next year, a parliamentary act was passed that ordered that the "outsides of all Buildings in and about the said City be henceforth made of Brick or Stone," among other building regulations to protect London's buildings against fire (Great Britain Record Commission, 1819). Although wood has never entirely fallen out of use in Europe, laws, customs, and economics moved European cities toward less combustible building materials both sooner and more completely than in North America (Fantin, 2012), largely solving the problem of great urban fires by the 19th century (Hensler, 2011).

In the United States and Canada, however, wood continues to be used extensively in construction. The seemingly endless forests of North America made it an economically appealing building material, and sprawling, abundant land reduced the risk of fires spreading between buildings (Schulz, 1993)—at least until the late 18th and early 19th centuries when Americans started packing together in dense cities, which began to be consumed by massive urban conflagrations. Manhattan experienced multiple "great fires" starting in 1776, with large fires in Chicago and San Francisco following later into the 19th and early 20th centuries. The policy response in some cities was to adopt a more European attitude toward wood, but the material was so common that even in cities like New York, which went furthest to restrict its use, wood remained a common material in new construction into the 20th century (Friedman, 2010; Howe, 2013).

Therefore, a distinct approach to fire safety was developed, which continues to shape North American architecture today—the second exit. New York was, as is often the case in building technologies and regulations, a leader in requiring a second means of escaping a building, with rules requiring fire escapes beginning in 1860. The fire escape—a series of lightweight balconies connected by ladders, with access through a window above the ground floor and retractable ladder rungs at street level—spread across North America. Its advantages were that it was cheap to build and could be added to a design without altering the interior floor plan (Wermiel, 2003).

The fire escape had flaws, because window egress and ladders made it difficult for anybody without full physical faculties to use. Therefore, the narrow, flimsy, open-air balconies and ladders of the 19th century eventually evolved into America's general requirement in the 20th century that apartment buildings with at least four stories have two staircases (known as "exits" in code parlance), often fully walled off from the rest of the floor plan and on opposite ends of the building (Wermiel, 2003).[1]

The Double-Loaded Corridor Floor Plan

The requirement for a second exit led to a distinctive design in North America—the double-loaded corridor floor plan. Multifamily buildings in the United States and Canada generally center around a long, straight hallway bisecting the building along the narrower dimension, with many units arrayed on either side. In much of the world, this design is most characteristic of hotels or student dormitories (Eliason, 2023). In most of North America, it has become the only realistic way to build apartments.

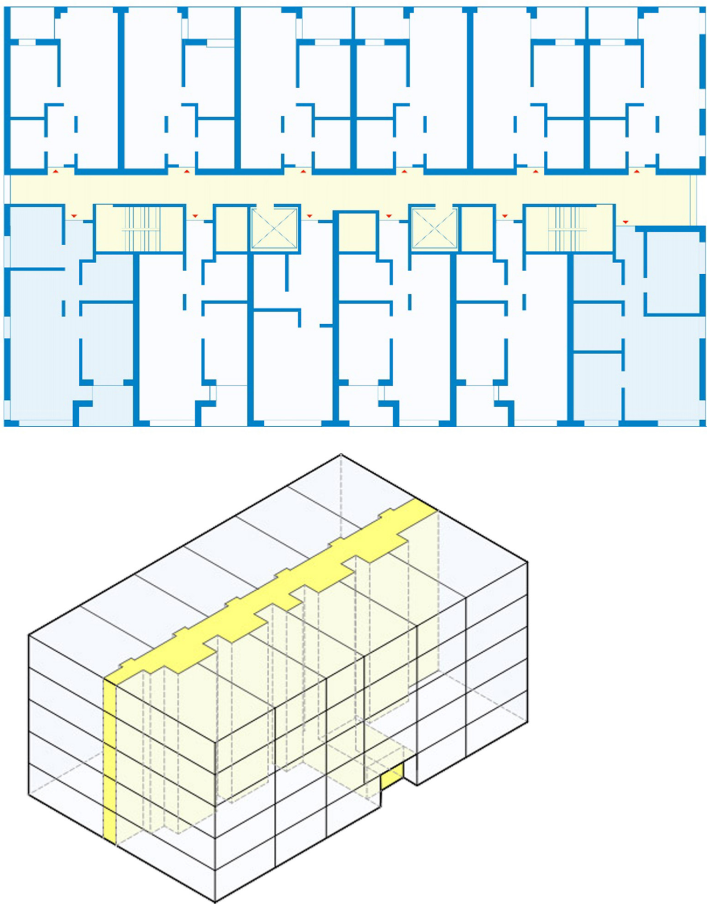
The design flows from the North American building code requirement that each unit has access to two exits—that is, staircases—along with a few related vertical circulation elements, such as elevators and trash storage. Housing developers all around the world, whether government or private, generally seek to include as much floor area as possible within apartments and minimize the space that circulation (both vertical circulation and corridors) consumes. Therefore, to ensure a

[1] 2021 International Building Code, Chapter 10, Means of Egress, https://codes.iccsafe.org/content/IBC2021P1/chapter-10-means-of-egress.

high enough percentage of the floor area sits within apartments, more apartments must be built on each floor, pushing architects and developers toward a double-loaded corridor design, as exhibit 2 shows. Because double-loaded corridors mean that each apartment generally has only a window on one side, designing efficient layouts for family-sized apartments becomes more difficult, contributing to developers building studio and one-bedroom apartments (Smith, 2023).

Exhibit 2

Double-Loaded Corridor Building, Viewed in Plan and Axonometrically



Source: Design by Michael Eliason, Larch Lab

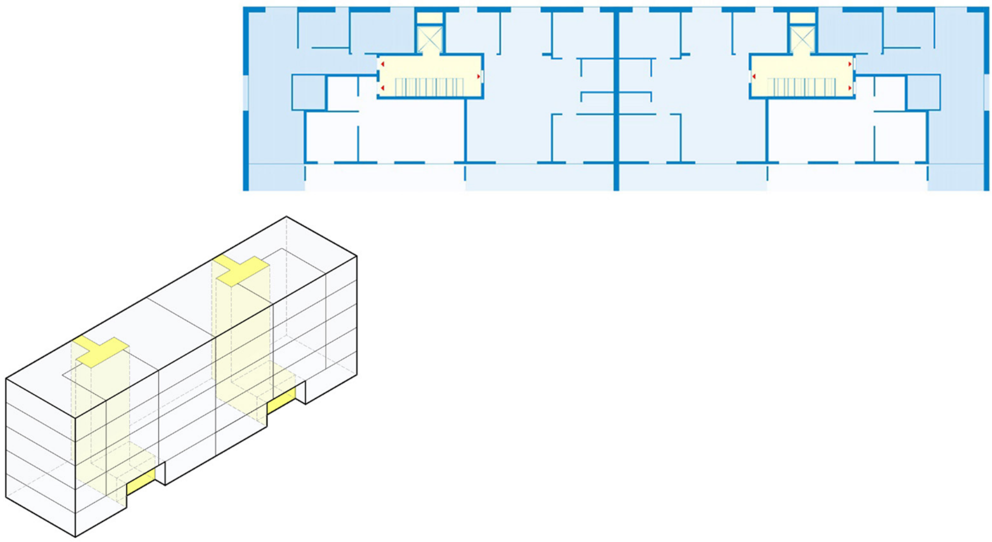
An Alternative Model: Point Access Blocks

A double-loaded corridor floor plan stands in contrast to the typical way of designing apartment buildings in the rest of the world, characterized by what architects in North America refer to as *point access blocks* (Eliason, 2021), a design with a single point of access from which every apartment is easily accessible without the need for long corridors. Point access blocks, as exhibit 3 shows, can have a variant of different designs, including lowrise typologies like the mid-century

American "garden apartment" or West Coast "dingbat," mid-rise housing blocks found in Europe, and highrise single-stair towers more common in East Asia. In North America, because rules typically limit these types of buildings to three stories, proponents usually focus on incremental mid-rise building types rather than the more aggressive highrise versions.

Exhibit 3

Point Access Block (Two Single-Stair Conditions, Attached Together), Viewed in Plan and Axonometrically



Source: Design by Michael Eliason, Larch Lab

Point access blocks involve only a handful of apartments on each floor compared with often more than a dozen per floor in a double-loaded corridor. Requiring apartments to be served by less vertical and horizontal circulation space makes this design efficient, thereby maintaining a reasonable ratio of unrentable stair and hallway area relative to a smaller number of apartments. Without a long corridor bisecting the building, architects also have more flexibility in laying out units. Especially at mid-rise heights, this flexibility can allow for apartments with windows on multiple sides to stretch from the front of the building to the back, allowing more light and air into each room and making it easier to position the bedrooms and living area on opposite sides of the building for privacy and sound attenuation. The greater number of windows allows for more efficient family-sized layouts because windows are what make space usable for bedrooms, in lieu of shifting space to interior areas without direct sunlight, through the use of more and larger bathrooms or closets, or even the growing American trend of windowless bedrooms (Wagner, 2023).

Outside the United States and Canada, building multifamily dwellings with a single staircase is much more common. Most such buildings tend to be mid-rises, such as the freestanding *palazzine* of mid-century Italy, or the affordable, wooden walkup *apaato* or newer and more luxurious concrete *manshon* buildings with elevators in Japan. Low- and mid-rise point access blocks are the most common type globally, with building codes and regulators considering taller buildings to be more of a risk and many requiring a second stair at a certain height, thereby eliminating some of the efficiencies of the model. That said, highrise, single-stair point access blocks can be found in many countries, particularly in Asia,[2] Latin America,[3] Eastern Europe (Teoalida, n.d.), and increasingly Western Europe (Burkard Meyer Architects, 2023),[4] as planners become more comfortable with height due to rising housing costs, growing demand to live in cities, and declining fire death rates with new technologies.

Americans might be most familiar with a humbler style of point access block, one of the most recent low-cost housing typologies in North America—the garden apartment. These two- and three-story complexes, typically built on the fringes of American cities (at least at the time of their construction), feature a series of connected point access blocks with a single staircase each, giving access to typically one or two apartments on either side per floor. Because the requirement for two stairs in building codes in most of America kicks in only at the fourth floor, such complexes can be built with minimal space "wasted" from a financial perspective on a second staircase and long, unrentable corridors connecting the vertical circulation to the apartments. This type of housing, which often offers two- and three-bedroom apartments alongside smaller units, used to be common in newer parts of American cities and inner suburbs in the decades after World War II (Hess, 2005). However, developers are building fewer new garden apartment complexes because, among other reasons, greenfield land within a reasonable commute of downtown is becoming less available, and tenant and condo-buyer demand shifts toward more central locations with different land economics and zoning are pushing developers to build taller and more costly housing types (Anderson, 2019).

The photo of Colonial Village in Arlington, Virginia, in exhibit 4 illustrates a typical garden-style apartment design familiar to most readers. Because the buildings in this complex are two to three stories high, they are already below the current U.S. single-stair height limit. The same basic garden-style layout can also be used for buildings above three stories, with apartments centered around single access points and interior stairwells, and the entire development that includes multiple sets of apartment clusters and entrances.

[2] For example, https://www.archdaily.com/619665/the-minton-dp-architects?ad_source=search&ad_medium=projects_tab.

[3] For example, https://www.archdaily.com/999364/oscar-by-you-residential-building-perkins-and-will?ad_source=search&ad_medium=projects_tab.

[4] For example, https://www.b22.it/en/projects/cascina-merlata-housing-sociale/.

Exhibit 4

Garden Apartments of Colonial Village in Arlington, Virginia

A photo of Colonial Village, a garden apartment complex in Arlington, Virginia. Colonial Village in Arlington, Virginia, was the first Federal Housing Administration-insured large-scale rental housing project. The 50-acre garden apartment complex was built between 1935 and 1940, had a 10,000-person waiting list for 276 units, and still provides affordable housing today. Some claimed that at the time of construction, it was "the most widely copied piece of real estate in the country" (Boodman, 1977).

Photo credit: Winerman (2011).

Policy Adoption of Point Access Blocks

A growing interest in the United States and Canada seeks to combine the advantages of the garden apartment with the density of mid-rise buildings by reforming building codes that currently restrict point access blocks with a single stairwell for each group of apartments to three stories in the United States and two stories in Canada. In 2023, Washington Governor Jay Inslee, Oregon Governor Tina Kotek, and California Governor Gavin Newsom all signed into law bills ordering studies to determine whether and how to allow taller single-stair buildings in their respective states.[5] Bills have also been introduced in New York City and at the state level to either study or enact similar changes (California YIMBY, 2023; SPUR Digital Discourse, 2023),[6] and the Canadian

[5] WA SB5491, https://legiscan.com/WA/text/SB5491/2023; HB 3395, https://olis.oregonlegislature.gov/liz/2023R1/Measures/Overview/HB3395; CA AB835, https://legiscan.com/CA/text/AB835/2023; Senate Bill S6573, https://www.nysenate.gov/legislation/bills/2023/S6573.

[6] The New York City Council, Legislation Int 0794-2022, https://legistar.council.nyc.gov/LegislationDetail.aspx?GUID=37FE9CD9-BF36-4B6F-A20A-70FA1A52E99C&ID=5898998.

Board for Harmonized Construction Codes has decided to develop a code change to allow taller single-exit buildings as well.[7]

Within the United States, the cities of Honolulu, New York, and Seattle already have more flexible exit rules for mid-rise buildings, allowing point access blocks of up to six stories with a single stairwell, with additional fire and life safety restrictions not otherwise required for buildings of such size. Seattle's code since 1977 has allowed for new apartment buildings up to six stories if built with 1-hour fire-resistive construction,[8] with at most four apartments per floor, and a travel distance of no more than 125 feet from the farthest point in each apartment to the stairwell.[9] The consolidated city and county of Honolulu (which encompasses all the island of Oahu, home to two-thirds of the state of Hawaii's population) recently adopted similar language to Seattle in its building code.[10] New York City allows new buildings of up to six stories under a somewhat different set of conditions, requiring steel or concrete construction (as it generally does throughout the whole city, regardless of the number of stairs) and limiting floor area to 2,000 square feet per story (with legislation introduced in 2022 to double that size, bringing it more in line with Seattle's allowance).[11]

Beyond local building codes in Seattle and New York City, NFPA also recommends somewhat looser standards for single-stair buildings than most adopted U.S. building codes. NFPA is a nonprofit organization that sets standards for fire protection in the United States that are often adopted into code, and one of its first model codes, published in 1927, was called the Building Exits Code (Bukowski and Kuligowski, n.d.). Today, the NFPA 101 Life Safety Code (the modern successor to the Building Exits Code) and the NFPA 5000 Building Construction and Safety Code recommend allowing up to four stories of apartments around a single stairwell, subject to a few conditions that are similar to those found in Seattle's code—for example, four units per floor at most, sprinklers throughout the building, and a 1-hour fire resistance rating for walls around the exit stair and corridors leading to it.[12]

Outside the United States and Canada, codes in other countries often go much further in allowing taller single-stair configurations in point access blocks (exhibit 5). Germany allows buildings of up to 60 meters (197 feet), or around 20 stories, to have a single stairway. Switzerland has no height limit, and single-stair towers of more than 30 stories are now being built (Burkard Meyer Architects, 2023). In Asia and Oceania, taller single-stair buildings are also common. South Korea allows them of unlimited height, mainland China allows them up to 18 stories, and Australia allows them up to 25 meters (82 feet, or roughly eight stories).[13]

---

[7] CCRs 1815 and 1816, https://codes.nrc-cnrc.gc.ca/en/certifications-evaluations-standards/codes-canada/codes-development-process/code_change_requests.html.

[8] The Second Egress: Building a Code Change, Seattle, https://secondegress.ca/Seattle.

[9] UpCodes, 2018 Seattle Building Code 1006.3.3, Single Exits. https://up.codes/viewer/seattle/ibc-2018/chapter/10/means-of-egress#1006.3.3.

[10] Hawaii State Building Code, https://codelibrary.amlegal.com/codes/honolulu/latest/honolulu/0-0-0-14009?fbclid=IwAR24kHU1ZsmJN_mXmTt1eNYWd1y3Y7fGDLJ2Jm5G5sR2ALXygJIGMMB5pXc.

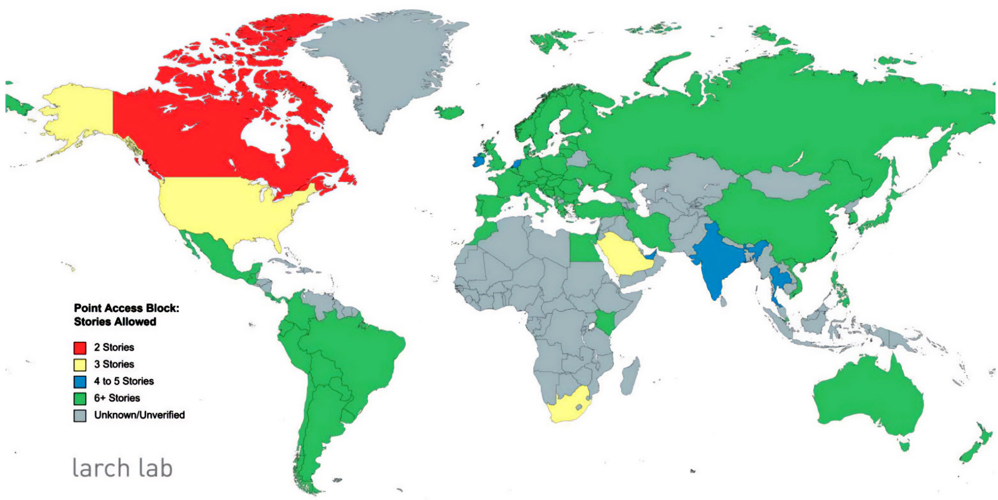
[11] NYC BC, 1006.3.2.

[12] NFPA 5000, 2021, 25.2.4.5; NFPA 101, 2021, 30.2.4.6.

[13] The Second Egress: Building a Code Change, Jurisdictions, https://secondegress.ca/Jurisdictions.

Exhibit 5

Legality of Single-Stair Buildings by Number of Stories



Note: Excludes city-specific modifications in the United States (for example, Honolulu, New York City, and Seattle allowing up to six stories).
Source: Eliason (2021)

Although single-stair buildings are commonplace and the codes allowing them have not attracted controversy in most of the developed world outside of the United States and Canada, they have recently come under scrutiny in the United Kingdom following the loss of 72 lives after a fire in 2017 enveloped the 24-story Grenfell Tower apartment block in London. The building's single staircase was cited as one of many factors leading to the deaths, and tighter limits are now being considered for new single-stair buildings going forward, with London imposing a 30-meter (98-foot) limit for apartment buildings and some discussion of an 18-meter (59-foot) limit in the future for the United Kingdom outside of Scotland (NFCC, 2022; Spocchia, 2023).

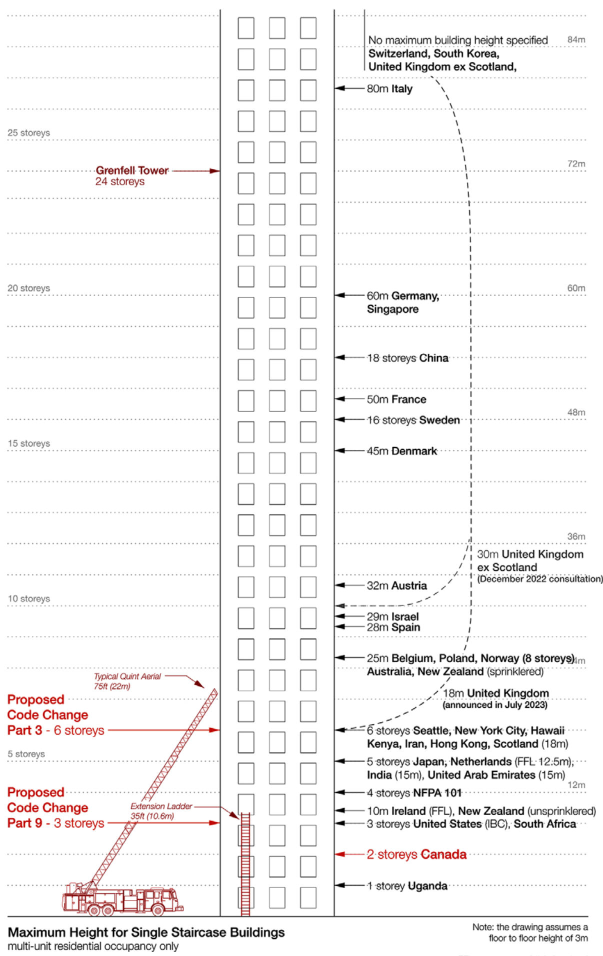
Following the fire, several of Grenfell's features were rethought, leading to code changes that were implemented before the current reevaluation of single-stair rules. Flammable cladding, which allowed a small fire to rapidly spread across the façade of the building, was seen as the most egregious contributor to the fire and led to the swiftest reaction—not only for buildings going forward but also for existing buildings with the same cladding, which is being stripped at an enormous cost (Longley, 2020). The lack of fire sprinklers meant that no active suppression was available until the fire service arrived, and a sprinkler requirement was imposed a few years after the fire (FPA, 2020). Beyond those issues, a "stay-in-place" policy remained in effect for more than an hour after the fire began before a general evacuation order was finally ordered (Agerholm, 2018). The fire led to other changes in building regulation in the United Kingdom, including a controversial ban on mass timber construction in buildings above 18 meters (59 feet), even though the building that burned used none of the material (Rudgard, 2023).

American and Canadian proponents of raising the single-stair height limit have focused on allowing mid-rise buildings, of at most a few more stories than the current two- and three-story height limits, although the debate in the United Kingdom over single-stair skyscrapers sometimes obscures the debate in North America. As noted previously, three U.S. cities already allow multifamily buildings up to six stories with a single stairwell and, in the case of Seattle, have done so for decades without major incident or controversy.

Conrad Speckert, a Toronto-based architect who requested that the national code change to raise Canada's height limit for single-exit buildings, created the graphic in exhibit 6. It is meant to generally represent the maximum allowed height, with further details such as occupancy limits per floor and construction types omitted for simplicity's sake. It illustrates the various building height requirements in different countries where multiple stairwells are required but does not necessarily demonstrate the range of building types that can rely on a single staircase, which are more varied than a slim, freestanding highrise building.

Exhibit 6

Legality of Single-Stair Housing by Height



Source: The Second Egress: Building a Code Change, Jurisdictions, https://secondegress.ca/Jurisdictions

Challenges in Building Code Development

Determining the appropriate limits of building codes is a difficult challenge because of the serious life safety consequences of getting it wrong combined with the informational challenges involved in evaluating fire safety. Historically, building codes have been developed through trial and error, reacting—sometimes underreacting and sometimes overreacting—to major fires. Data collection is incomplete, complicating large-scale empirical approaches, and fires have become relatively rare, making small-scale data collection less useful. Engineered approaches using advanced computational modeling can offer a way out of traditional challenges, but numbers must also contend with historical traditions involved in code development.

The Triangle Shirtwaist Factory fire in Manhattan in 1911 exemplifies the challenges of legislating exits when considering the uncertainty and politics involved in major fires. A 1915 law passed in reaction to the fire that required four exits, rather than the previous three, in a commercial building like the one that burned (which was, by modern standards, not terribly tall at just 10 stories of under 10,000 square feet each), only to be bumped back down to the current two in a 1938 building code that aimed for quality over quantity in commercial building exits (Spivack, n.d.).

A lack of uniform data and information on outcomes complicates data-driven approaches to determining the safety of single-stair requirements, taking advantage of variations in different countries' rules. Although the United States does have an unusually high rate of fire death among its high-income peers around the world (Brushlinsky et al., 2022), a fact proponents often use to argue that our relatively low single-stair height limit is not keeping us safe from fires (Eliason, 2021), these top-line numbers cannot speak to the safety of specific building types in countries with many different types of housing. Finer analyses may be possible in some countries but have yet to be performed on available data to provide specific insights about the safety of taller buildings.

Our divergent tradition of building with wood further complicates the use of codes from safer countries abroad to inform codes in North America. Although modern materials and devices like gypsum board and automatic sprinklers and engineering approaches like fire-resistance ratings provide a high degree of safety to light wood frame structures, the inherent flammability of the light wood frame used in much of the United States for low- and mid-rise structures does give some pause about adopting global norms for building exits, because the dominant building materials abroad are noncombustible concrete and masonry. The National Fire Protection Association published an article in 2022 critical of the effort to adopt more European approaches to multifamily building exits, citing Europe's greater use of noncombustible materials like brick and concrete and stricter standards for compartmentalization within buildings. In the article, NFPA recommended its own four-story single-stair height limit (taller than the two- and three-story limits enacted in most of the United States and Canada) as a possible path forward (Ziavras, et al. 2022).

Urban Infill Brings New Code Challenges

Across North America, governments and the private sector are increasingly embracing infill development, often due to the shrinking supply of available land, growing demand for urban living, and new worries about the environmental consequences of sprawl. States and cities are reconsidering strict single-family zoning in some cases (Fox, 2022; Healey and Ballinger, 2021;

Jeffords, 2023; Wamsley, 2019), and multifamily permitting in the United States is at highs not seen since the 1980s.[14] New growth within urban cores since then has largely occurred on commercial, industrial, or vacant land. However, with those sites becoming scarcer, cities across the continent are turning toward upzoning built-out single-family neighborhoods for small apartment buildings. Cities are now looking to bolster the so-called "missing middle"—that is, housing typologies denser than a detached single-family house but smaller than a large mid- or highrise apartment building requiring large lots (Parolek, 2020). Therefore, planners are turning to the most abundant type of land in cities that is also one of the more controversial for development—single-family lots.

Unlike large parking lots or industrial areas that have served as sites for new housing during the past few decades, single-family lots tend to be small; anywhere from 2,500 to 7,500 square feet is typical in urban settings. These small lots struggle to accommodate apartment buildings under today's building codes. Developers typically target an efficiency ratio—that is, the ratio of rentable floor area inside of apartments to the total area of a typical upper floor—of somewhere around 85 percent. To keep this ratio relatively stable, every extra square foot of circulation area—whether a second staircase, a corridor to increase the fire separation between the staircase and dwelling units, the required distance between the two stairs, or a larger elevator—must come with an additional 5 to 10 square feet of extra space within apartments. On a small urban lot, growing the size of each floor to accommodate the second stairwell and other features of North American buildings while still maintaining a reasonable efficiency ratio under current building code constraints means eating into prized open space.

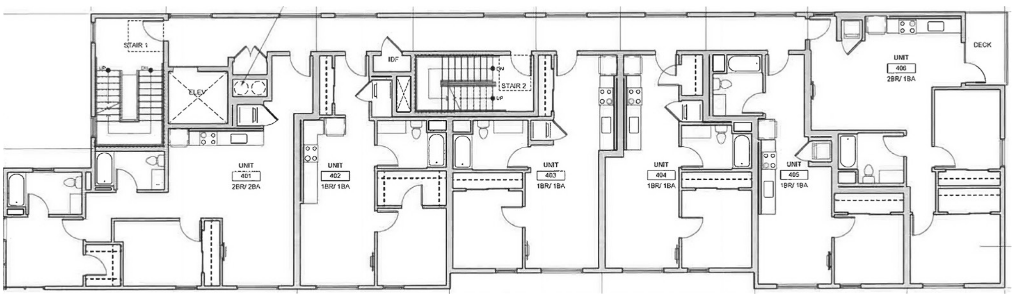
In Los Angeles, for example, new mid-rise multifamily buildings on small lots end up coming within 10 feet of their rear property line, and even with this fairly high lot coverage, they often have efficiency ratios short of 80 percent (exhibit 7). Shoehorning apartments into single-family neighborhoods will inevitably require compromising on these high multifamily lot coverages (current allowed lot coverage in single-family neighborhoods in Los Angeles tops out around 50 percent),[15] but shrinking buildings while maintaining the same building code requirements for two stairs in protected shafts, in addition to elevators needed for mid-rise buildings, would drive efficiency ratios low enough to call into question the feasibility of development. This problem can be mitigated to some extent if developers buy multiple small lots next to each other to enlarge their development sites, but timing purchases is tricky, and so-called land assembly can drive up the price of land in Los Angeles by up to 40 percent (Brooks and Lutz, 2016). Winning political support for multifamily zoning in single-family neighborhoods is difficult enough for mayors, city council members, and urban planners in North America, and building code requirements for an additional staircase and some other related features—reducing the total amount of residential units that can be built on the available lot size—can make the task more difficult.

[14] Building Permit Survey, https://socds.huduser.gov/permits/.

[15] Los Angelese Department of City Planning, https://planning.lacity.org/odocument/eadcb225-a16b-4ce6-bc94-c915408c2b04/Zoning_Code_Summary.pdf.

Exhibit 7

Typical Floor Plan of a Los Angeles Multifamily Building on a Single 50’x100’ Lot



Source: DFH Architects

Future Directions

Reforms and modifications to building codes in the United States can be considered in multiple ways. Building codes are written variously at the national, state, county, and local levels, and change can come from any and all of these levels. The most widespread change tends to come through the national model building codes, with states, counties, and cities often proving grounds for new rules of building that do—or do not—eventually make it into the model code, which then get promulgated back downward through cyclical adoptions by the various levels of government. The allowance of mass timber construction in the United States, for example—which is more fire-resistant than typical American light wood frame but with strength that allows it to be used for building types for which, previously, only concrete or wood would do—first came at the state level in the Pacific Northwest (Pacheco, 2018; Washington Forest Protection Association and Forterra, 2018). It was later adopted into the International Building Code (AWC, n.d.), the latest version of which is finally being adopted in jurisdictions across the United States.[16]

Proponents of raising the limit for single-stair buildings are, so far, focusing state legislative efforts on studying the issue. How such studies might occur is an open question because existing code provisions are quite old and were arrived at largely through trial and error and in response to major fires. Data beyond anecdotes—domestically in Honolulu, New York City, and Seattle and also abroad—are spotty, because fire services do not keep records of fire fatalities by number of exits per building. Newer, more advanced computational models of smoke spread and time to exit buildings or holistic methods like decision tree analysis could fill some gaps in data, although engineered approaches will still have to contend with more than one-and-a-half centuries of precedent for two means of egress in the United States.[17]

Finally, reform has a wide range of possibilities. Rules in the rest of the world run the gamut from the ultra-permissive (for example, no height limits for single-stair buildings in Switzerland or South Korea) to the very strict (Uganda requires a second staircase connecting any story above or

[16] Building Code Allowances for Tall Mass Timber, https://www.woodworks.org/resources/status-of-building-code-allowances-for-tall-mass-timber-in-the-ibc/.

[17] NFPA 550, https://link.nfpa.org/free-access/publications/550/2022; NFPA 92, https://link.nfpa.org/free-access/publications/92/2021.

below the ground floor),[18] but advocates in North America have focused on liberalizing rules for mid-rise buildings of up to six stories (Gordon, 2022). Two NFPA model codes—NFPA 101 and NFPA 5000—offer a more incremental path forward, allowing for single-stair apartment buildings of up to four stories.[19] More permissive rules around stairs could also be paired with stricter rules for construction type (for example, requiring more fire-resistant materials like concrete or mass timber) and compartmentalization, in addition to the current requirements for fire sprinklers (which are not typically found abroad) to mitigate the risk and bring overall standards in line with those abroad.[20]

Conclusion

Building codes are a new area of interest for housing advocacy in recent years, which has, until now, largely focused on land use and transportation. Focusing on the ways in which buildings are constructed, and not just where and to what size, opens new avenues for shaping the building environment but also new perils if mistakes are made. Unlike other urban policy measures, building codes have a much more critical function in life safety, and reforms will have to meet a high burden of proof. A history of fires and resource availability shaped the North American way of building, and the onus will be on advocates for change to prove that their suggestions are necessary and safe.

Authors

Stephen Smith is the director of the Center for Building in North America. Eduardo Mendoza is the policy director for the Livable Communities Initiative and a professional city planner based in Los Angeles.

References

Agerholm, Harriet. 2018. "Grenfell Tower Inquiry Hears Trapped Residents Were Never Told to Evacuate, Even After Order was Given," *The Independent*, September 13. https://www.independent.co.uk/news/uk/home-news/grenfell-tower-inquiry-residents-999-evacuate-stay-put-advice-fsg-a8536511.html.

American Wood Council (AWC). n.d. "Understanding the Tall Mass Timber Code Changes: A Toolkit for Fire Officials." https://awc.org/wp-content/uploads/2022/01/tmt_toolkit.pdf.

[18] NFPA 550, https://link.nfpa.org/free-access/publications/550/2022; NFPA 92, https://link.nfpa.org/free-access/publications/92/2021; The Second Egress: Building a Code Change, Jurisdictions, https://secondegress.ca/Jurisdictions.

[19] NFPA 101, https://www.nfpa.org/codes-and-standards/all-codes-and-standards/list-of-codes-and-standards/detail?code=101; NFPA 5000, https://www.nfpa.org/codes-and-standards/all-codes-and-standards/list-of-codes-and-standards/detail?code=5000.

[20] Legislative Incentives for Fire Sprinklers in New Buildings, https://www.eurosprinkler.org/wp-content/uploads/2023/02/Summary-Legislation.pdf.

Anderson, Bendix. 2019. "Garden-Style Apartment Projects Allow Developers to Expand in the Suburbs," *WealthManagement.com*, July 30. https://www.wealthmanagement.com/multifamily/garden-style-apartment-projects-allow-developers-expand-suburbs.

Boodman, Sara G. 1977. "First FHA Garden Apartments, Colonial Village, May Be Sold," *The Washington Post*, October 6. https://www.washingtonpost.com/archive/local/1977/10/06/first-fha-garden-apartments-colonial-village-may-be-sold/17776da1-e7ac-4212-9d24-93bd8ec92bf3/.

Brooks, Leah, and Byron Lutz. 2016. "From Today's City to Tomorrow's City: An Empirical Investigation of Urban Land Assembly," *American Economic Journal: Economic Policy* 8 (3): 69–105. https://www.aeaweb.org/articles?id=10.1257/pol.20130399.

Brushlinsky, Nikolai, Sergei Sokolov, Peter Wagner, and Birgitte Messerschmidt. 2022. *World Fire Statistics* 27. CTIF International Association of Fire and Rescue Services. https://www.ctif.org/sites/default/files/2022-08/CTIF_Report27_ESG_0.pdf.

Bukowski, Richard W., and Erica Kuligowski. n.d. *The Basis for Egress Provisions in U.S. Building Codes*. Gaithersburg, MD: NIST Building and Fire Research Laboratory. https://tsapps.nist.gov/publication/get_pdf.cfm?pub_id=861281.

Burkard Meyer Architects. 2023. "Growing Together Over Height: High-Rise Residential Building in Ostermundingen," *BauNetz*, January 5. https://www.baunetz.de/mobil/meldung.html?cid=8119437#bildergalerie.

California YIMBY. 2023. "The Single Stairway to Heaven," *California YIMBY blog*, February 1. https://cayimby.org/blog/the-single-stairway-to-heaven/.

Eliason, Michael. 2023. "Why Does American Multifamily Architecture Look So Banal? Here's One Reason," *The Architect's Newspaper*, March 29. https://www.archpaper.com/2023/03/why-does-american-multifamily-architecture-look-so-banal-heres-one-reason/.

———. 2021. *Unlocking Livable, Resilient, Decarbonized Housing with Point Access Blocks*. Report prepared by Larch Lab for the City of Vancouver. https://www.larchlab.com/city-of-vancouver-report-on-point-access-blocks/.

Evelyn, John. 1661. *Fumifugium, or, the Inconveniencie of the Aer and Smoak of London Dissipated: Together With Some Remedies Humbly Proposed.* London, UK. https://collections.library.yale.edu/catalog/17256624.

Fantin, Mathias. 2012. "Metal Floors with Double T Joists," August 5. http://combiencaporte.blogspot.com/2012/08/les-planchers-metalliques-poutrelles-en.html.

Fire Protection Association (FPA). 2020. "Sprinklers to Be Required in 11M Residential Buildings." https://www.thefpa.co.uk/news/sprinklers-to-be-required-in-11m-residential-buildings.

Fox, Justin. 2022. "What Happened When Minneapolis Ended Single-Family Zoning," *Bloomberg*, August 20. https://www.bloomberg.com/opinion/articles/2022-08-20/what-happened-when-minneapolis-ended-single-family-zoning.

Friedman, Donald. 2010. *Historical Building Construction: Design, Materials, and Technology*, 2nd ed. New York: W.W. Norton & Company.

Gordon, Wyatt. 2022. "How Allowing Single-Staircase Buildings Could Change Virginia's Housing Market," *Virginia Mercury*, May 5. https://www.virginiamercury.com/2022/05/05/how-allowing-single-staircase-buildings-could-change-virginias-housing-market/.

Great Britain Record Commission. 1819. "Charles II, 1666: An Act for Rebuilding the City of London," *Statutes of the Realm* 5: 1628–80. British History Online. https://www.british-history.ac.uk/statutes-realm/vol5/pp603-612.

Healey, Jon, and Matthew Ballinger. 2021. "What Just Happened with Single-Family Zoning in California?" *Los Angeles Times*, September 17. https://www.latimes.com/homeless-housing/story/2021-09-17/what-just-happened-with-single-family-zoning-in-california.

Hensler, Bruce. 2011. *Crucible of Fire: Nineteenth-Century Urban Fires and the Making of the Modern Fire Service*. Sterling, VA: Potomac Books.

Hess, Paul Mitchell. 2005. "Rediscovering the Logic of Garden Apartments," *Places* 17 (2): 30–35. https://escholarship.org/content/qt1fk783z2/qt1fk783z2.pdf.

Howe, Richard. 2013. "Notes on Invisible Wood," *The Gotham Center for New York History blog*, January 30. https://www.gothamcenter.org/blog/notes-on-invisible-wood.

Jeffords, Shawn. 2023. "Toronto City Council Approves Multiplexes to Address Growing Housing Crisis," *CBC*, May 10. https://www.cbc.ca/news/canada/toronto/toronto-approves-multiplex-vote-1.6839296.

London Fire Brigade (LFB) Museum. n.d. "The Great Fire of London, 1666." https://www.london-fire.gov.uk/media/5079/72157_museum-fs_great-fire-of-london-lores.pdf.

Longley, Cathy. 2020. "Three Years After Grenfell, UK Announces Major Funding to Strip High-Rises of Dangerous Cladding," *National Fire Protection Association blog*. https://web.archive.org/web/20220705055020/https://www.nfpa.org/News-and-Research/Publications-and-media/Blogs-Landing-Page/NFPA-Today/Blog-Posts/2020/05/29/three-years-after-grenfell-uk-announces-major-funding-to-strip-high-rises-of-dangerous-cladding.

National Fire Chiefs Council (NFCC). 2022. "NFCC Calls for New High Rise Residential Buildings to Have More Than One Fire Escape Staircase." https://www.nationalfirechiefs.org.uk/News/nfcc-calls-for-new-high-rise-residential-buildings-to-have-more-than-one-fire-escape-staircase.

Pacheco, Antonio. 2018. "Washington State is Embracing Mass Timber Construction," *The Architect's Newspaper*, March 19. https://www.archpaper.com/2018/03/washington-state-embracing-mass-timber-construction/.

Parolek, Daniel G. 2020. *Missing Middle Housing: Thinking Big and Building Small to Respond to Today's Housing*. Washington, DC: Island Press.

Rudgard, Olivia. 2023. “Fire Fears After Grenfell Disaster Set Back Wood Building in UK,” *Bloomberg*, April 19. https://www.bloomberg.com/news/features/2023-04-19/low-carbon-wood-building-designs-face-fire-safety-doubts-in-uk#xj4y7vzkg.

Schulz, Horst. 1993. “The Development of Wood Utilization in the 19th, 20th and 21st Centuries,” *The Forestry Chronicle* 69 (4): 413–418. https://pubs.cif-ifc.org/doi/pdf/10.5558/tfc69413-4.

Smith, Stephen. 2023. “Why We Can’t Build Family-Sized Apartments in North America,” *Center for Building in North America blog*, May 4. https://www.centerforbuilding.org/blog/we-we-cant-build-family-sized-apartments-in-north-america.

Spivack, Dolores. n.d. *Amending the Building Code of the City of New York: Exploring Forces that Influenced Change*. New Jersey Institute of Technology. http://archives.njit.edu/vol01/etd/2010s/2016/njit-etd2016-040/njit-etd2016-040.pdf.

Spocchia, Gino. 2023. “Second Staircases Required on All London Towers Over 30m, Rules Mayor,” *Architects’ Journal*, February 14. https://www.architectsjournal.co.uk/news/second-staircases-required-on-all-london-towers-over-30m-rules-mayor.

SPUR Digital Discourse. 2023. “Stairway to Affordability: How We Can Diversify Multifamily Housing,” *SPUR.org*, June 14. https://www.spur.org/events/2023-06-14/stairway-affordability-how-we-can-diversify-multifamily-housing.

Teoalida, Mihai. n.d. “The Evolution of Communist Blocks and Apartments.” https://www.hartablocuri.ro/planuri/.

Wagner, Kate. 2023. “We Cannot Countenance Windowless Bedrooms,” *The Nation*, March 21. https://www.thenation.com/article/society/windowless-bedrooms-housing-crisis/.

Wamsley, Laurel. 2019. “Oregon Legislature Votes to Essentially Ban Single-Family Zoning,” *NPR*, July 1. https://www.npr.org/2019/07/01/737798440/oregon-legislature-votes-to-essentially-ban-single-family-zoning.

Washington Forest Protection Association and Forterra. 2018. “Washington State to Allow Mid and High-Rise Mass-Timber Buildings.” https://www.prnewswire.com/news-releases/washington-state-to-allow-mid-and-high-rise-mass-timber-buildings-300760329.html.

Wermiel, Sara E. 2003. “No Exit: The Rise and Demise of the Outside Fire Escape,” *Technology and Culture* 44 (2): 258–284.

Winerman, Lea. 2011. “Where We Live: Colonial Village,” *The Washington Post*, November 14. https://www.washingtonpost.com/realestate/where-we-live-colonial-village/2011/11/14/gIQAGfOWUN_story.html.

Ziavras, Valeria, Shawn Mahoney, Jonathan Hart, and Corey Hannahs. 2022. “Single-Exit Design in Multistory Apartment Buildings,” *NFPA Journal: In Compliance*, August 8. https://web.archive.org/web/20231001000000*/https://www.nfpa.org/News-and-Research/Publications-and-media/NFPA-Journal/2022/Fall-2022/In-Compliance/In-Comp-Package.

U.S. DEPARTMENT OF HOUSING AND URBAN DEVELOPMENT

Office of Policy Development and Research

Cityscape
VOLUME 26 NUMBER 1 • 2024