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.