Evaluation of the Project SOAR Demonstration

Final Report

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

Evaluation of the Project SOAR Demonstration
Final Report

Michael DiDomenico
Rebecca Johnson

Office of Evaluation Sciences
U.S. General Services Administration

July 2020

Acknowledgments

There are many people without whom this program and the accompanying evaluation would not have been possible. Maria-Lana Queen from the Department of Housing and Urban Development (HUD) managed the SOAR program and provided instrumental help in understanding operations and connecting us with grantees. Teresa Souza was a great partner in managing the research contract and generally helped us overcome any problems we encountered. Jagruti Rekhi, Calvin Johnson, and Mark Shroder provided insightful feedback on the research design and execution. Barry Goldstein and Michael Collins from the Department of Education were integral in helping to navigate the interagency data sharing process that allowed us to track educational outcomes. Site visitors included Mindy Ault, Justin Brock, Marissa Dolan, Richard Duckworth, Tiesha Lawrence, Daniel Marcin, Calvin Johnson, Heidi Joseph, Ransford Osafo-Danso, Jagruti Rekhi, Elizabeth Rudd, Blair Russell, Mark Schroder, and Carol Star from HUD and Crystal Hall from the Office of Evaluation Science. We received additional support and feedback from Rekha Balu, Katherine Christie, Jason Houle, Ryan Moore, Mary Clair Turner, Miles Williams, and the rest of the Office of Evaluation Sciences (OES) team. Finally, we are thankful to all of the people who worked tirelessly at the nine grantees to implement the program and were generous in giving us access to learn more about the challenges and successes they faced. And a special thank you to all of the education navigators who were the ones actually working with students to help them overcome obstacles and meet their goals.

Foreword

Project SOAR (Students + Opportunities + Achievements = Results), also known as “ROSS for Education,” was a 2-year demonstration conducted by the U.S. Department of Housing and Urban Development (HUD), implemented between 2017 and 2019, to expand educational services to youth living in public housing. Project SOAR provided grant funding to nine public housing agencies (PHAs) to hire education navigators to help youth between the ages of 15 and 20 to complete the necessary steps to transition to and succeed in postsecondary educational programs. The primary goal of the demonstration was to help students complete the Free Application for Federal Student Aid (FAFSA). Other grant objectives included improving financial literacy and college readiness, completing postsecondary program applications, and completing other tasks necessary for postsecondary education enrollment. As part of the grant, PHAs agreed to participate in an impact evaluation.

This report presents the results from the experimental and quasi-experimental impact evaluations and the findings from an implementation study based on site visits and a navigators’ tracking tool. The implementation study found that it took time for SOAR navigators to engage residents. Navigators encountered reluctance from residents to participate, and they provided in-person assistance to fewer than half of age-eligible residents. Both the experimental and non-experimental impact evaluations found that Project SOAR did not lead to statistically significant improvements in FAFSA completion, college enrollment, or Pell Grant receipt during and immediately after the completion of the grant (between October 2018 and March 2020). On average, 28 percent of eligible residents in the treatment group submitted a FAFSA application, and 29 percent enrolled in college in 2019.

SOAR was an important demonstration aimed at addressing the gap in college attendance by family income and overcoming barriers to college enrollment among HUD-assisted youth. Despite null results, the findings can assist in the design of future interventions to increase educational attainment among HUD-assisted tenants. The data match between HUD and the U.S. Department of Education demonstrated the potential of this approach for future research on housing and educational outcomes at a low cost and relatively quick turnaround.

Todd Richardson
General Deputy Assistant Secretary for Policy Development and Research
U.S. Department of Housing and Urban Development

Contents

List of Exhibits

Abstract

Project SOAR (Students + Opportunities + Achievements = Results), also known as ROSS for Education, was a demonstration program reflecting the U.S. Department of Housing and Urban Development's (HUD) commitment to expand educational services to youth living in HUD-assisted housing. Project SOAR provided grant funding to nine public housing authorities (PHAs) to hire and deploy counselors—or "education navigators"—to help youth between the ages of 15 and 20 living in public housing complete the necessary steps to transition to and succeed in postsecondary educational programs.

The overall goal of this study is to explore whether the SOAR grants improved postsecondary outcomes. The outcomes of interest include completion of the Free Application for Federal Student Aid (FAFSA), college enrollment, and receipt of a Pell Grant. In evaluating SOAR's impact on these outcomes, the study contributes to a large body of literature on how to reduce disparities in educational opportunity (Section 2). Past research shows that for low-income students, interventions that involve less time and resources—for instance, letter or email outreach campaigns—can fail to change long-term outcomes. Project SOAR navigators were trained to not only provide informational materials about steps like completing the FAFSA but also to supplement the information with hands-on assistance and support.

To document program implementation (Section 3), the report draws on two sources of data: qualitative notes from site visits to the grantees and detailed records that navigators kept about their outreach efforts and interactions with residents. The key findings are that:

1. **SOAR navigators needed time to establish themselves in the community and engage residents.** In the site visits, navigators reported the importance of taking the time to get to know residents and establishing themselves in the community. In the quantitative data, navigators' interactions with residents started off slowly but began to meaningfully increase by the summer of 2018. This means that navigators had limited interactions with residents to prepare them for the first focal FAFSA season. In addition, by the time navigators were building more recognition in the community, grantees were already under pressure to spend the funds prior to the grant's termination.

2. **SOAR navigators were only able to provide in-person assistance to a subset of age-eligible residents.** While navigators were able to reach many residents through methods like mailings and e-mail, they were only able to meet one-quarter to one-half of residents at least once in person. The low level of engagement has important implications for the main analysis, which estimates the average effect among all eligible residents.

3. **Among residents who met in person with navigators, there was wide variation in the frequency and number of attendees at those meetings.** Decisions to complete the FAFSA and enroll in college are often family-level rather than student-level; that is,

students weigh the benefits of postsecondary education against challenges such as getting their parents to provide the necessary financial information or pressure to forgo college for immediate earnings. We see wide variation, however, in whether and where families of students were engaged. Some sites often included parents and caregivers as attendees in navigator meetings; other sites rarely did so.

Section 4 discusses the methods used for the impact evaluation. First, in four of the nine PHAs, residents were randomized to treatment and control conditions by clusters of buildings (Asset Management Projects, or AMPs). The primary analysis estimates the Intent-to-Treat (ITT) effect of SOAR. An exploratory analysis also examines the estimated effect of SOAR on youth who met with a navigator in-person at least once. A secondary analysis of the remaining five PHAs that did not randomize residents to assistance uses a non-experimental design to evaluate the impact of SOAR. A synthetic control method matches the non-experimental SOAR PHAs to other PHAs with similar trajectories of FAFSA completion (and other postsecondary outcomes) *prior* to the SOAR grant period and examines whether the five PHAs had higher-than-expected FAFSA completion rates given their trajectories.

Section 5 presents the results. The **key findings** are that:

1. **Based on the experimental analysis, SOAR did not lead to statistically significant improvements in FAFSA completion, college enrollment, or Pell Grant receipt.** Treatment AMPs had slightly lower completion rates than control AMPs, but the differences were not distinguishable from zero.
2. **Based on the non-experimental analysis, SOAR did not lead to statistically significant improvements in FAFSA completion, college enrollment, or Pell Grant receipt.** SOAR may have led to some improvements relative to PHAs that did not receive any SOAR funding, but we cannot rule out that the improvements were due to chance.

Section 6 discusses the findings and their policy implications. SOAR navigators faced a difficult task. In a short time period, they needed to get to know a large number of PHA residents and their families, begin providing services to these residents, and guide them through complicated financial aid and college enrollment decisions. The results show that the SOAR model, when evaluated as the average effect across PHAs, did not lead to meaningful improvements in postsecondary outcomes. The results leave open the possibility that the SOAR model, implemented differently, could produce improvements. For instance, a model that requires youth to express interest in a navigator, or that partners with a local school district to identify students for help, could help navigators focus services on the students for whom their assistance makes a larger difference.

Overall, while past research shows that the SOAR model of in-person assistance can be effective for low-income youth, the program model needs to be more defined, sufficient resources must be in place to support navigators, and more effort needs to be made to focus on the students who are both interested in college and in need of additional assistance.

1 Introduction

Project SOAR (Students + Opportunities + Achievements = Results), also known as ROSS for Education, was a demonstration program reflecting the U.S. Department of Housing and Urban Development's (HUD) commitment to expand educational services to youth living in HUD-assisted housing.[1] Project SOAR provided grant funding to nine public housing authorities (PHAs) to hire and deploy counselors—or "education navigators"—to help youth between the ages of 15 and 20 living in public housing complete the necessary steps to transition to and succeed in postsecondary educational programs. The most emphasized step was to complete the Free Application for Federal Student Aid (FAFSA).[2,3] Other targeted activities included increasing financial literacy and college readiness, completing program applications, and completing other administrative steps necessary for enrollment.

The Office of Evaluation Sciences (OES) at the U.S. General Services Administration worked with HUD to design an evaluation of Project SOAR to understand program implementation and ultimately learn the extent to which Project SOAR improved postsecondary outcomes for eligible students. This report summarizes the findings of both the implementation study and the impact evaluation. This research adds to the body of evidence exploring the effectiveness of college access interventions.

The remainder of this report is structured as follows. Section 2 discusses the disparities in college access among students from low-income families compared with their more-affluent peers and summarizes prior research of interventions designed to increase low-income students' postsecondary access and success. Section 3 discusses how the SOAR intervention was designed and implemented. Section 4 discusses both the planned experimental and non-experimental methods used to estimate the impact of Project SOAR on key outcomes of interest. Section 5 describes the estimated effects of Project SOAR. Previewing the results, the analysis suggests that SOAR had no statistically significant impact on either FAFSA completion or college enrollment. There was also, however, wide variation in how the intervention was delivered by each grantee, and navigators only engaged a limited subset of eligible residents. Section 6 offers some interpretation of the results and suggestions for future research and policy efforts.

---

[1] In this report college is used interchangeably with postsecondary educational programs more generally and includes not only 4- and 2-year college programs but also any other Title IV eligible institution of higher education at which U.S. federal student aid can be used. We also use the terms student and (eligible) youth interchangeably in the context of who was targeted by and participated in services made possible by Project SOAR grants, even though in some cases the person engaged with an education navigator may not have been a student enrolled in high school or college.

[2] The report uses PHA and grantee interchangeably when referring to PHAs that hired education navigators.

[3] The decision to focus on residents aged 15–20 was based on (1) the desire to use some eligibility criteria to make the navigators' task of assisting residents more feasible, and (2) the assumption that a higher proportion of 15–20-year-olds were interested in college than, for instance, 25–30-year-olds.

2 Disparities in educational opportunity

Project SOAR was designed to help reduce persistent gaps in college-going between students from low-income families and their more affluent peers. Students from families with a lower socioeconomic status (SES) both attend and graduate from college at lower rates than students from wealthier families, even when taking into account measures of academic preparation.[4] For example, among students with math scores in the top two quintiles, only 63 percent of students in lower SES families are enrolled in college within 3 years of high school graduation, whereas 85 percent of students from high SES families are (Ma, Pender, and Welch, 2019). Even when looking at some of the best-prepared high school students—those with a GPA between 3.5 and 4.0—the 6-year completion rate among those who enroll at a public, 4-year college is 61 percent for students from lower SES families compared to a rate of 89 percent for students from higher SES families (Ma, Pender, and Welch, 2019).

2.1 Barriers to college

There are multiple barriers for students on the path to a postsecondary education that can help explain these disparities. Of particular relevance to low-income families are barriers related to financial constraints, including a lack of information about the financial aid process.

Information about financial aid: Understanding the financial aid process is crucial for students from low-income families. While the sticker price of college has increased rapidly over the past 2 decades, the out-of-pocket costs for low-income students have remained fairly low because of the availability of financial aid. For example, while the average published price of tuition and fees for a full-time student at a public 4-year university increased from $5,170 in the 1999–2000 school year to $10,140 in the 2015–2016 school year (both in 2019 dollars), the net tuition and fees after grant aid for a dependent student whose parents earned less than $35,000 per year were only $2,340 for the 2015–2016 school year (Ma et al., 2019). Unfortunately, many low-income families may have inaccurate information about the out-of-pocket costs of college. In surveys, low-income families have overestimated costs by between 80 and 150 percent (Scott-Clayton, 2012).

One common source of information for students is family or friends who have had to navigate the college-enrollment process before. Low-income students are more likely to lack these resources because they are more likely to be the first in their family to go to college, and the community resources to which they have access may be less robust. Staff at five SOAR grantee sites reported that low-performing and low-resourced high schools limited youths' academic preparation and access to peers or counselors who could help them navigate the college enrollment process.[5] In particular, respondents reported schools had no counselors or had very

---

[4] In this paragraph, high SES is defined as the top two quintiles or approximately above the 60th percentile. Low SES is defined as the bottom two quintiles, or roughly at or below the 40th percentile. SES is not only a measure of household income but also considers parental education and occupations (Ma, Pender, and Welch, 2019).

[5] Grantee observations were collected from a set of site visits described in further detail in Section 3.2.

high counselor-to-student ratios in some cities, and school choice systems were viewed as fragmented, limiting students' ability to attend high-quality and well-resourced schools. This lack of information can discourage some students from going to college because they may incorrectly assume it is not affordable or otherwise within reach.

Financial constraints: Youth in public housing can face a considerable amount of pressure to prioritize helping with family obligations over postsecondary education. Grantee staff suggested that residents tend to prioritize the near term over potential future payoffs when making college enrollment decisions. For example, parents in many cases wanted youth to help with the household and care for siblings in the immediate term rather than go to college to earn money that could support the family several years later. Many youth wanted to work right away and often had not considered Career Technical Education or certificate programs that could be completed quickly.[6] Given these pressures, part of the job of navigators was to better inform SOAR participants about the available options to help them make more-informed decisions about their futures.

Low-income families also are more likely to face setbacks from small financial barriers and opportunity costs related to the college-enrollment process. For example, grantees reported residents had difficulty paying small fees, and that in some cases there were limited fee waivers for the ACT and SAT tests. One grantee reported that even though a statewide fee waiver existed for traditional public school students, they were not available to youth who attended charter schools. These fees could prevent otherwise interested students from taking the necessary steps for applying to college. One role of navigators was to help students navigate the procedures for obtaining fee waivers for tests and applications to reduce this barrier.

Although barriers related to financial information and other financial constraints are among the most important faced by students from low-income families, grantee staff cited many additional barriers, including worries about childcare for youth with their own children, general problems with procrastination, and reluctance among the student's parents to share financial information. While SOAR was designed with a focus on reducing financial barriers by helping students understand the financial aid process, most navigators saw their role as identifying any number of barriers and helping where they could.

2.2 Prior interventions

There have been many attempts at reducing barriers to college access among low-income students and increasing the success of those who enroll. College access and college success interventions range from those which are extremely light touch (for example, one-way text messages, not personalized to the student) to those which are very intensive (such as significant

[6] Interestingly, some of the push to earn money quickly could be considered rational in an economic sense. Carrell and Sacerdote (2017) find in a survey that male high school students predict their wages with only a high school degree to be 52 percent higher than females expect, and ACS data suggest that in New Hampshire, where the study took place, high school men do, in fact, earn about as much as men with 1 to 3 years of college.

in-person assistance combined with additional financial and academic supports). Exhibit 1 summarizes prior research (in rough order of increasing intervention intensity). Light-touch interventions have gained significant popularity recently, not the least because they can be easy to implement and fund at a large scale, meaning even small gains can be cost-efficient. The emerging consensus, though, seems to be that while light-touch interventions have been effective in some cases, the positive effects of light-touch interventions are difficult to replicate, especially when they are taken to scale.

Over the last decade, there has been an explosion in the number of evaluations of light-touch, "nudge" interventions aimed at increasing college enrollment and persistence. These efforts have occasionally shown promise, but their effectiveness has been uneven. Early efforts to increase college enrollment (Castleman and Page, 2015) and FAFSA renewal (Castleman and Page, 2016) with encouraging text messages were effective at increasing 2-year college enrollment and persistence, respectively, but were not successful at changing 4-year college outcomes. Efforts at scaling these early interventions did not show effects (Bird et al., 2017). Page, Castleman, and Meyer (2020) finds that emails to students were able to increase FAFSA completion in the short run, but the effects deteriorated over time and were not significant within a few months. An intervention undertaken by the Department of Education's Office of Federal Student Aid was effective at increasing FAFSA completion rates with a simple email, although the effects were only measured over a 20-day period, and the persistence of the effect is unknown (OES, 2017). There have been some successes with light-touch, communications-based interventions, but the mechanism for why they work in some cases is poorly understood, meaning it is not likely that designing just any communication intervention is likely to change behavior.

Since 2016, HUD has attempted three interventions geared towards increasing FAFSA completion among residents. All three focused on informative communications—sent via postal mail, email, and robocall—broadly targeted to residents either between 17 to 20 years of age or between 17 to 24 years of age (OES, 2016, 2019a, 2019b). None of these efforts significantly improved FAFSA completion. Taken together, the results of the HUD experiments and attempts by other researchers suggest that information alone generally is not enough to elicit action.

Two studies that provide a more direct comparison of interventions with different levels of intensity seem to support the general conclusion that information alone is not enough to change behavior. Bettinger et al. (2012) tests one intervention that provides low-income students interested in learning more about college financing with information about likely financial aid packages and encouragement to complete the FAFSA on their own. That intervention was ineffective. However, when the same information was paired with in-person help by the tax preparer to complete and submit the FAFSA, students both completed the FAFSA at significantly higher rates and attended college at higher rates. In a similar vein, Carrell and Sacerdote (2017) test both an intervention in which colleges send information to potential students indicating the colleges' interest in the students, which was not effective, and a much more intensive intervention that offered a combination of weekly academic mentoring, financial support, and financial incentives for completing activities related to college applications, which was effective.

Exhibit 1: Overview of prior interventions

Study	Intervention	Population	Effect
OES (2016)	Sent residents one of nine letter variations encouraging FAFSA completion.	Residents ages 17–20 who lived in households (nationwide) using Housing Choice Vouchers.	No significant effect.
OES (2019a)	Sent a combination of four mailings, three emails (for those with email addresses), and one robocall to encourage FAFSA completion.	Residents of the New York City Housing Authority ages 17–24.	No significant effect.
OES (2019b)	Sent two letters encouraging FAFSA completion.	Youth living in public housing at the Seattle Housing Authority and King County Housing Authority ages 17–24.	No significant effect.
Castleman and Page (2015)	Sent students a series of two-way text messages related to tasks students accepted to college need to take prior to starting college in the fall.	High school students in Boston, Dallas, and Philadelphia.	The text messages increased 2-year enrollment by about 3 percentage points, on average, but did not significantly change overall enrollment.
Castleman and Page (2016)	Sent first time freshmen students a series of 12 text messages focusing on the importance of renewing their FAFSAs.	First time college freshmen who participated in Castleman and Page (2015).	Increased enrollment in the sophomore year for community college students by around 12–14 percentage points, but no effect on 4-year students.
Page, Castleman, and Meyer (2020)	Two-way text messages to students updating them on where they were in the FAFSA process (such as not started, started but not submitted, submitted and selected for verification).	High school students in Texas.	There was an increase in FAFSA completion over the short run, but the effect decreased and was not significant by the end of summer.
OES (2017)	Federal Student Aid sent emails to students reminding them to refile their FAFSAs and alerting them to changes in the FAFSA process for the current cycle.	Approximately 14 million students who filed a FAFSA for the previous (2016–2017) academic year.	Increased FAFSA renewal rates by 3.4 percentage points over a control group the first 20 days after the email.
Bird et al. (2017)	Sent a series of text messages encouraging FAFSA completion.	Students enrolled in the Common Application (nationwide) who were low-income or first-generation college students.	The intervention did not change the rate of FAFSA completion but did lead to a small increase in college enrollment, mainly driven by an increase in people going to 2-year programs who otherwise would not have enrolled in a program.

Study	Intervention	Population	Effect
Phillips and Reber (2019)	One condition sent email and text messages reminding students of key activities and deadlines. Another condition added the assistance of an advisor who communicated with students via email, text, and phone calls.	Low SES, college-interested, high school juniors in California who applied for the program.	The advising intervention increased applications to 4-year colleges, but neither intervention had a significant effect on college enrollment.
Bettinger et al. (2012)	Provided estimates of financial aid based on tax information. One condition provided only the information and encouragement for people to fill out the FAFSA on their own. A separate condition automatically filled in most of the FAFSA with tax information and offered personal assistance from the tax preparer to complete the rest of the FAFSA.	Low-income students between 15–30 years old getting tax preparation at H&R Block who say they are interested in learning more about college finances.	Increased FAFSA completion by 16 percentage points and college enrollment by 8 percentage points for dependent students in the personal assistance condition, but no effect for information only.
Scrivener et al. (2015)	Provides intensive counseling, tutoring, and career advising, last dollar funding, transit cards for NYC, free use of textbooks. Requires full time enrollment and emphasizes developmental courses and graduation within 3 years.	Low-income students at three New York City community colleges in need of developmental coursework who were willing to attend college full time.	Increased 3-year graduation from 22 to 40 percent. Also increased credits earned by 9 credits (39 to 48 credits) and increased enrollment in 4-year colleges 17 to 25 percent.
Miller et al. (2020)	Replication of Scrivener et al. (2015) in three Ohio community colleges.	Low-income students at three Ohio community colleges in need of developmental coursework who were willing to attend college full time.	Increased 3-year graduation from 19 to 35 percent. Increased 4-year college enrollment from 12 to 18 percent. Increased credits earned after 3 years by 8.5 credits.
Bertrand et al. (2019)	Intensive advising, tutoring, professional mentorship, financial support, and financial incentives.	Low-income students in Chicago planning to enroll or already enrolled in community college.	Increase in college enrollment of between 7 and 9 percentage points, increase in full-time enrollment of 13 percentage points, and an increase in persistence to spring of 11 percentage points.

Study	Intervention	Population	Effect
Carrell and Sacerdote (2017)	One intervention includes mentoring by a Dartmouth student, paying for application and test fees, and a financial incentive for completing activities. A second intervention asks colleges to send students a personalized letter and emails saying the college is interested in them.	High school students in New Hampshire who have expressed interest in college but have not completed an application.	The intensive intervention increased college enrollment by 6 percentage points, driven entirely by increases in women going to 4-year schools. The interest letter was not effective.

Note: All studies used experimental designs except for the HUD Seattle and King County study, which used a synthetic control method.

The research suggests that more intensive interventions more consistently produce significant and sizable gains in college enrollment and persistence. The interventions studied by Scrivener et al. (2015), Miller et al. (2020), Bertrand et al. (2019), and Carrell and Sacerdote (2017) all include a combination of intensive academic advising and mentoring and financial support in the form of scholarships, monetary incentives, or in-kind benefits like the use of textbooks or a transit card. Providing this more holistic array of supports has more consistently shown large impacts, even when replicated, as was done with the Accelerated Study in Associate Programs (ASAP) intervention (Scrivener et al., 2015; Miller et al., 2020).

One key feature of the interventions cited (regardless of intervention intensity) is that they are generally targeted at students who have been identified as interested in and academically prepared for college. For instance, Page et al. (2019) studied the impact of a program targeted at “low-income students identified as having the potential to enroll and succeed in college” (p. 6). Carrell and Sacerdote (2017) selected a population where “The high school guidance departments identify students who have expressed interest in college but have taken few or no steps to apply. The intent is to capture students who are right at the margin of applying to college or failing to apply” (p. 126). Phillips and Reber (2019) studied a program that “targeted students who were likely to be eligible for admission to public 4-year colleges in California based on their prior grades and course-taking” and who “had relatively high-grade point averages (about 75 percent reported B averages or above) and very high educational aspirations (nearly 80 percent aspired to a graduate degree)” (p. 8). The specific targeting of these interventions helps limit the study to students who are likely college-interested, but it also could mean that a large part of the sample will go to college regardless of the intervention.

The present intervention targeted public housing residents by age rather than by high school grades or interest in college. This choice was largely due to the lack of any educational data collected directly by HUD, but it makes the design a relatively novel attempt not only to help students already interested in college but also to potentially foster college aspirations. Grantees mentioned in several interviews that they felt SOAR was important, in part because it is able to proactively reach some students who are missed by other programs that rely on students to actively engage in college preparation before being recruited into an intervention. But while SOAR had the benefit of reaching students who wanted and needed help but who might have been missed by other programs, it also likely targeted youth who either were not college-interested and had made other plans or who were planning to go to college and did not require any assistance to do so, which could make the effectiveness of the program more difficult to measure.

3 Project SOAR design and implementation

3.1 Program design

Project SOAR targeted multiple behaviors related to pursuing a postsecondary education. Each grantee, a PHA, hired between one and three navigators to help residents between 15 to 20 years of age. Each navigator was expected to assist approximately 100 to 125 students over the course of a year. While the primary goal of the intervention was to help students complete the FAFSA, there were four total objectives specified in the grant announcement:

1. Help students complete the FAFSA.
2. Improve student financial literacy and college readiness.
3. Help students complete postsecondary program applications.
4. Help students complete tasks necessary for enrollment.[7]

Nine PHAs were selected for the demonstration and were funded to employ education navigators to carry out tasks in support of the main program objectives. Exhibit 2 lists each housing authority and the number of navigators awarded to each. As part of the grant, PHAs agreed to participate in an evaluation. The four grantees that participated in the experimental impact evaluation—where navigators were randomly assigned to help a subset of eligible residents—are listed in bold. The remaining five grantees were part of a non-experimental component of the impact evaluation. The exhibit also includes the number of 15- to 20-year-olds who were living in public housing at each PHA at the time of the grant award and the number of youth each PHA proposed to serve.[8]

Exhibit 3 summarizes the timeline of the grants relative to the two FAFSA cycles that occurred while navigators were providing services. The grants were awarded in May 2017, and grantees were encouraged to hire navigators early in the summer so they could be in place and offering services prior to the start of the 2018–2019 FAFSA season, which began on October 1, 2017.[9]

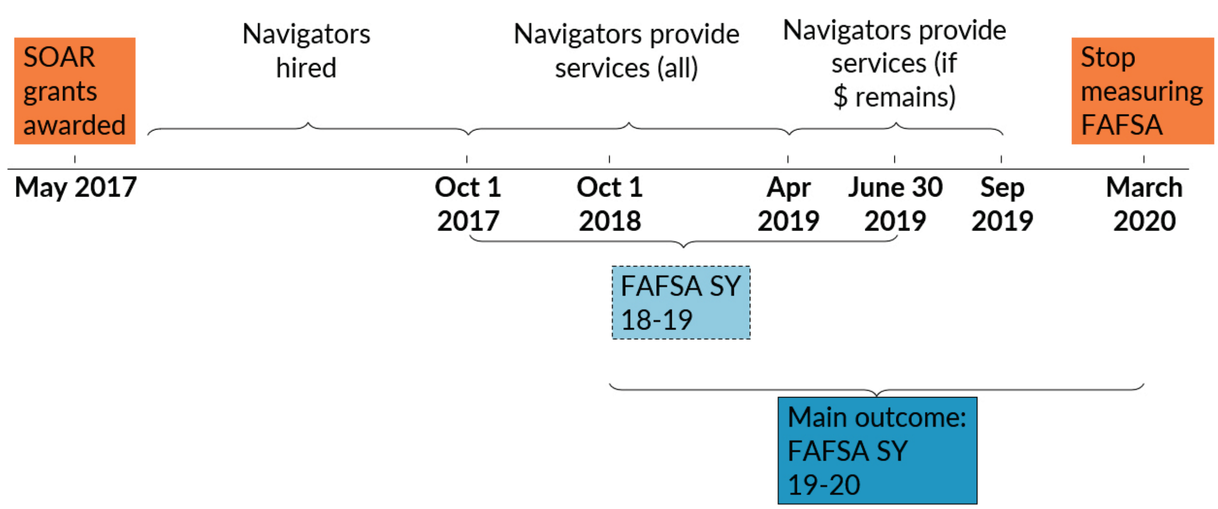
[7] Those tasks included both ones preliminary to the FAFSA and applications, like forming a "College Action Plan" where students outline the timeline for completing key steps, as well as tasks related to avoiding "summer melt" where students who enroll in a college do not show up, such as help registering for courses and figuring out their living arrangements.
[8] The number of navigators was roughly scaled to the size of the eligible student population that PHAs reported to HUD when PHAs applied for the grants. PHAs also differed in whether the navigators were employed full time or part time.
[9] Since the 2016–2017 season, each FAFSA cycle runs for 21 months from October 1 of the year preceding the intended school year to June 30 of the same school year. For example, if a student intended to attend college in the 2019–2020 school year, she could complete the FAFSA at any point between October 1, 2018 and June 30, 2020.

Exhibit 2: SOAR Grantees.

PHA	State	Total Residents 15–20	Residents to Serve	Navigators	Residents per Navigator
Chicago Housing Authority	IL	3,207	750	3	250
Philadelphia Housing Authority	PA	3,189	250	2	125
Housing Authority of the City of Los Angeles	CA	3,103	250	3	83
Seattle Housing Authority	WA	862	427	3	142
City of Phoenix Housing Department	AZ	655	298	3	99
Housing Authority of the City of Milwaukee	WI	536	208	1	208
High Point Housing Authority	NC	312	347	1	347
Prichard Housing Authority	AL	119	101	1	101
Northwest Georgia Housing Authority	GA	89	80	1	80

Note: Grantees in **bold** are included in the experimental evaluation. Other grantees are included in the non-experimental evaluation. The sample sizes represent estimates at the time of randomization, which are distinct from the actual, analytic sample sizes depicted in Exhibit 18.

Exhibit 3: Timeline of grants in relation to FAFSA cycles. Exhibit is not drawn to scale.



Given the differing speed of the hiring process, each grantee effectively began providing services at different points in time between August and November 2017.[10] The grant supported navigators through April 2019, which means that the navigators' services could help students

[10] As a result, the end date of hiring and start date of services in Exhibit 3 is an approximation and varies across PHAs.

with both the 2018–2019 FAFSA season and the 2019–2020 FAFSA season.[11] As is discussed further in Section 3, it took grantees time to ramp up and develop a steady state of program activities. As such, it was more likely for navigators to have an impact in the second year of the grant (for example, from summer 2018), and the primary analysis focuses on FAFSA completion for the 2019–2020 school year.[12]

3.2 Implementation analysis

Project SOAR was designed to utilize features of both lighter-touch and more intensive interventions. Navigators were told to focus on lighter-touch interactions for important one-time actions (such as personalized FAFSA completion assistance) and encourage follow-through on those actions (for example, frequent check-ins) while supplementing lighter-touch efforts with more intensive interactions for those who wanted it (such as one-on-one assistance with college search, scholarship search, or academic preparation).

The grant was designed to give PHAs maximum flexibility to tailor their program models to local context. As such, what navigators were doing at each grantee site looked somewhat different. This local variation has implications for whom navigators focused their outreach efforts on, who they ended up engaging, and the content of engagements. To understand better how SOAR was implemented, the research team completed an implementation analysis based on a set of site visits to grantees and an exploration of data collected on navigator-resident interactions collected by grantees.

3.3 Implementation data sources

Two main data sources inform the implementation analysis: data gathered from site visits to grantees and data collected by grantees on program participants and activities.

Implementation site visits

Site visits were conducted in July and August 2018 by teams of between two and four staff members from OES and HUD.[13] Most site visits took place over two days, although for some of the smaller grantees, site visits were completed over the course of a single day. Each visit included interviews with education navigators, grantee staff responsible for overseeing day-to-day operations of Project SOAR and direct management of education navigators, and grantee leadership. Site visitors were provided with interview guides to create semi-structured conversations that covered the same general topics but allowed for the interviewers to probe for more detail according to their discretion. When possible, site visits also included observations of

[11] Grantees were allowed to continue providing services until September 30, 2019, if they had unobligated funds after April.

[12] As we discuss later, the timing of our analysis means that our observations of the 2019–2020 FAFSA season are truncated to the last application dates we can observe in the Department of Education data before matching, which was March 12, 2020.

[13] There is a more detailed report related to implementation available from HUD upon request.

interactions with potential SOAR participants, observations of group events, and visits to several of the housing developments. Site visitors were asked to record notes of their interactions and observations, but the observations were not structured.

One limitation of the site visits is that young adults and their families were not interviewed directly. The decision was made only to speak with grantee staff, so the study team did not have to take additional steps that would be needed to get a research design approved that involved interviewing young adults and their families. As such, any descriptions of the challenges faced by young adults and their families reflect grantee staff's understanding of the situation gained from their interactions with young adults and their families—in other words, most of the information generated from the site visits is second-hand.

At the completion of each site visit, the site visit team was asked to document their notes and write a summary of the visit in a provided template. The study team decided not to record or transcribe conversations because the purpose of the site visits was to understand general experiences and think about opportunities for program improvement rather than to provide the types of "thick description" seen in other types of qualitative analyses.

For the analysis, the study team completed a preliminary review of site visit summaries and had a meeting to identify key themes and settle on a coding schema. Two reviewers then coded notes from one grantee, and the codes were compared to discuss decisions and improve reliability. The two reviewers then coded the remaining site visit summaries (each site was coded by one reviewer), referring to interaction-specific notes when possible. The coding schema focused on identifying structural and behavioral barriers and facilitators to taking action falling into three general categories: attributes and activities of the Project SOAR model, community context and programs, and descriptions of residents (including their financial circumstances and postsecondary awareness).

SOAR data systems

For Project SOAR, HUD developed a data tracking tool that navigators were instructed to use to document program activities. The tracking tool includes a roster of residents (the participant tracker) and a transaction-level accounting of program activities (the interaction tracker). The interaction tracker contains the date and type of an interaction between a resident and a navigator. Grantees were expected to use the tracking tool consistently, and HUD designed the tracker to utilize pre-specified categories to encourage data alignment (a list of the categories and definitions is provided in Appendix Section 7.2). Despite these efforts, there seemed to be a good deal of variation in how different grantees used the tracker. As such, any interpretation of cross-grantee comparisons should take this potential variation into account. In addition, while site visits were conducted at all nine grantee sites, one of the grantees in the non-experimental evaluation—Prichard Housing Authority in Alabama—did not provide HUD with usable participant or interaction tracker data. Therefore, the quantitative analysis of interactions discussed in Section 3.5 focuses on the remaining eight grantees.

The analysis relies on merging the data in the participant tracker and the interaction tracker to data from the Public and Indian Housing Information Center (PIC) to obtain a limited set of demographic characteristics for each student listed in the tracker, including age, sex, race and ethnicity, citizenship status, household income, and the number of 15- to 20-year-olds in that student's household.

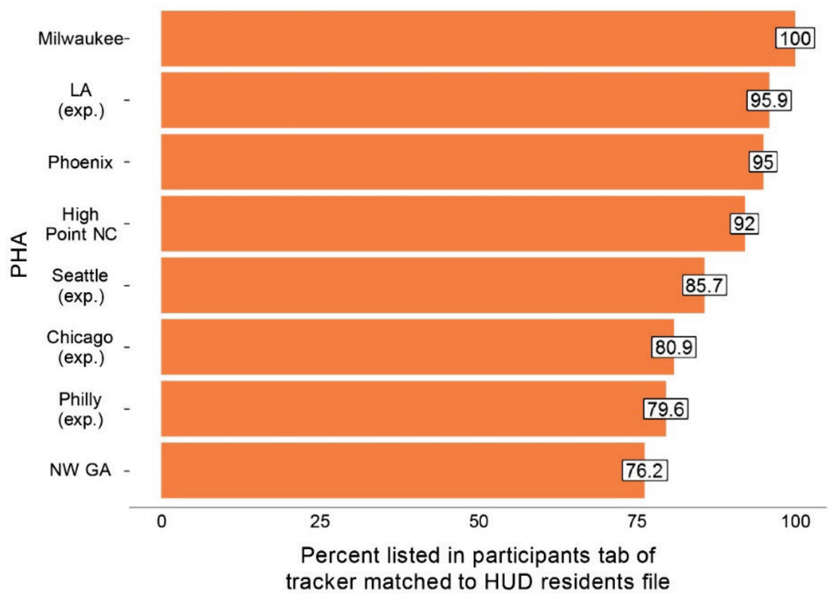
To merge the data, we performed a fuzzy match on first name, last name, and date of birth (DOB).[14] There were three ways that a record from the interaction tracker and a record from PIC could count as a match:

1. Match exactly on first name, match exactly on last name, and match exactly on DOB.
2. Fuzzy match on first and last name and match exactly on DOB.
3. Match exactly on both first name and last name and fuzzy match on DOB.

A perfect match was not expected because navigators may have interacted with youth who did not reside in the PHA or who resided there but were not included in the PIC data because they were outside the eligible age range (which was limited to 15- to 20-year-old residents) or were perhaps not included in the PIC data for other reasons (for example, not reported on the household roster). Exhibit 4 shows that the match rate differs among grantees, with rates above 95 percent for Milwaukee, Los Angeles, and Phoenix to rates of just under 80 percent for Philadelphia and Northwest Georgia.

[14] In practice, matching was performed in `R` using `fastLink` (Enamorado, Fifield, and Imai, 2019) and a Jaro-Winkler string distance with a threshold equal to 0.9.

Exhibit 4: Match rate of individuals in the activity tracker and PIC. The match rate is defined as the percentage of students listed in the participant tracker who could be matched to the PIC residents file for each PHA. For shorthand in this figure and in remaining figures, the names of Los Angeles (LA), Northwest Georgia (NW GA), and Philadelphia (Philly) are abbreviated. PHAs in the experimental impact evaluation are noted with "(exp.)" below the PHA name.



3.4 Program development

Outside of the grant announcement, HUD gave few firm guidelines to grantees as to how to structure their programs. When SOAR launched, some PHAs decided to adopt existing models with minimal adaptation, such as College Depot (Phoenix) and the College Access Plan (Los Angeles). The benefits of relying on existing organizations included well-defined training systems and more fully-developed materials and structures, including dedicated space at the public library in the case of College Depot, placement test boot camps, and other supports. While the use of core program structures held certain advantages, grantees still had to find ways to adapt the programs to the PHA context. For example, the College Access Plan model was based on education navigators being placed on-site in high schools. Los Angeles had to make adjustments to base the program in public housing developments, which included having education navigators learn how to recruit individuals in the community.

When grantees did not build SOAR on an existing model, they often made use of referrals to institutions in the community such as the Boys and Girls Club, Catholic Community Services, YearUp, or other groups offering tutoring and other forms of academic preparation rather than trying to replicate the services in-house. Some education navigators focused on connecting young adults to other existing services offered by the grantee, such as resident services programs like FamilyWorks in Chicago and the chess program in High Point. Some grantees already offered college-related benefits, including tuition scholarships in Chicago and transportation assistance to incentivize college applications and enrollment in Seattle.

Leveraging existing grantee programs allowed navigators both to approach students at a location where they knew students would be and to use existing service referral systems.

In several instances, education navigators took it upon themselves to build partnerships when they did not formally exist, such as by meeting with local school principals and college counselors to ensure SOAR students were receiving attention and support both at school and the PHA. Because high school counselors also were balancing large caseloads (estimated to be as many as 500 students per counselor in Chicago), education navigators recognized that schools may not provide all students with college information and indicated that, in some instances, counselors may even have steered SOAR students with poor grades away from college (Los Angeles).

3.5 Outreach and engagement

Navigators had a broad pool of age-eligible (15 to 20) students they were asked to assist. The characteristics of those whom navigators ended up assisting, which was a product of both navigators' efforts to engage students and students' willingness to meet with navigators, may have had implications for the effectiveness of the intervention. For instance, if navigators were spending the majority of their time meeting with youth who were reluctant to attend college, they may have needed to spend more time with each individual in order to get the desired outcome. On the other hand, navigators who spent the majority of their time meeting with youth who already had well-defined plans for going to college may have been supporting students who needed less intensive help (and may have been successful without any help from the navigators). Education navigators pursued a variety of strategies for interacting with potential participants by trial and error in an attempt to address the specific community context and personal needs of each student.

Variation in the types of interactions

Many interviewees agreed that navigators needed to engage the whole family, not just the student. Grantees interviewed described one of the more common strategies in larger developments as the "knock and talk"—visiting residents (sometimes unscheduled on a weekend) at their homes so that education navigators could make a personal connection with both the student and parent(s). Buildings with open access made this approach easier. Other geographic and physical features made this approach more difficult. Scattered-site housing designs in Phoenix and the scale of developments in Los Angeles made it harder for one navigator to reach enough students efficiently, and education navigators in Chicago had trouble reaching residents in mixed-income housing where they needed to be buzzed in by a resident.

Though the navigators we interviewed thought door-to-door interactions were the most effective way to engage students, grantees supplemented door knocking with other types of outreach, which commonly included newsletters, mass postal mailings, emails, and posters. Grantees (and navigators in particular) did not think these efforts were as effective as door-to-door outreach, but they thought they were helpful for raising the overall visibility of the program.

These impressions are supported by the interaction tracker data. As seen in Exhibit 5, which examines the total number of times each mode of interaction was used, the three most common strategies reported were mailing literature, emailing, and in-person, one-on-one counseling. The same general results hold when looking at the total number of students exposed to each strategy. Exhibit 6 shows the same three activities as the most popular, but more students were exposed to interactions classified as one-on-one counseling than any other strategy.

Exhibit 5: Total number of interactions by type. In these counts, the unit of analysis is the interaction. If the same student met with a navigator for four separate one-on-one counseling sessions, it appears in the count four times. The figure shows that while low-cost outreach methods such as mailing literature and emails comprised the majority of interactions, interactions classified as one-on-one counseling were also frequently reported.

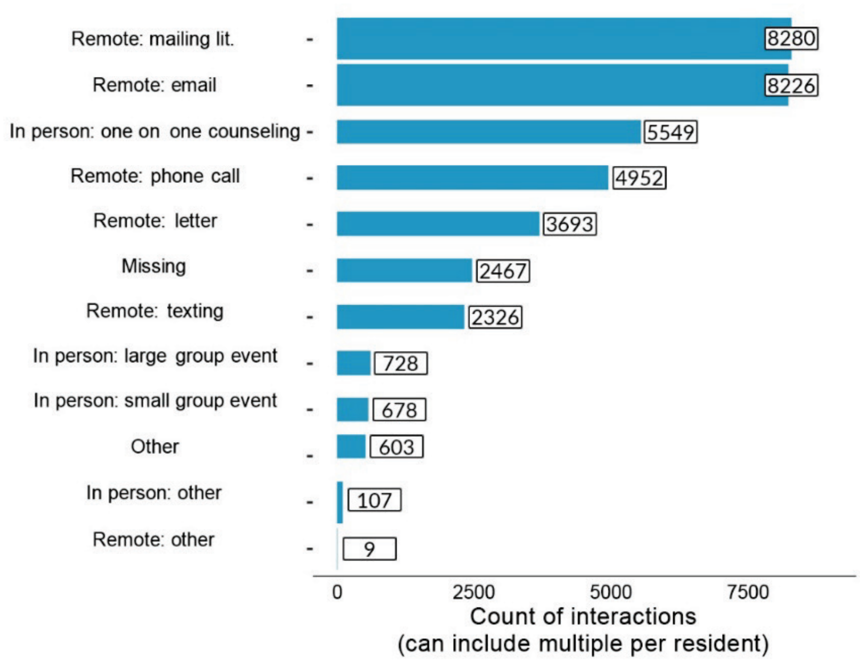
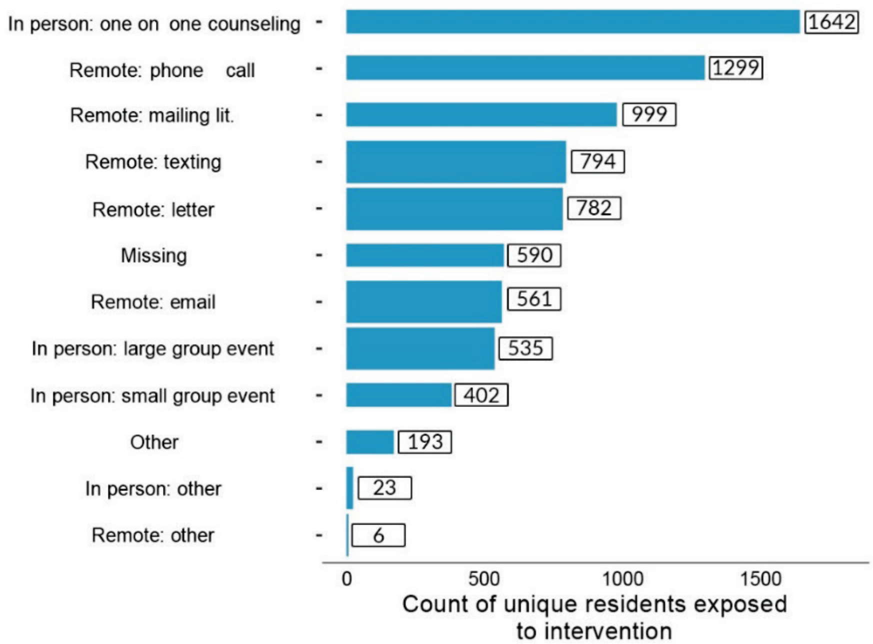
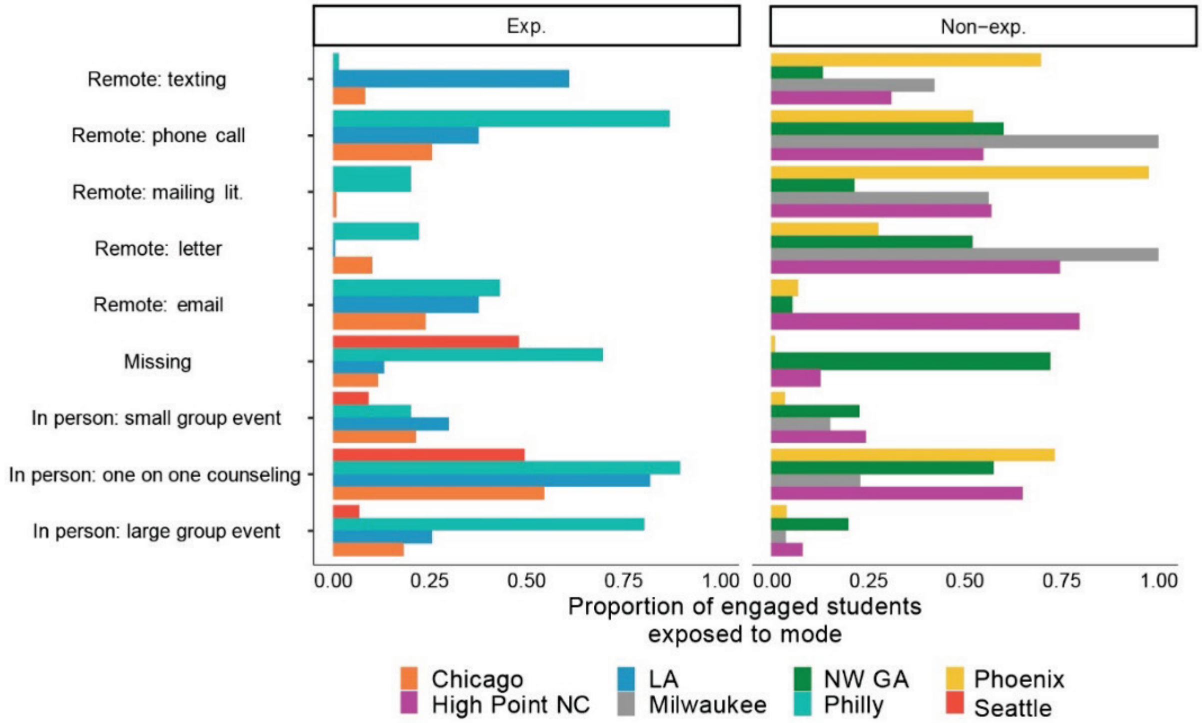


Exhibit 6: Number of residents exposed to each type of interaction. In these counts, if a student receives multiple "doses" of the same type of interaction—for example, four in-person meetings—she only appears in the count once. The counts are summed across all grantees.



Within these general patterns of engagement, there was significant variation among grantees in the methods navigators reported using. Exhibit 7 shows the proportion of engaged students who were exposed to each mode of interaction. In general, non-experimental grantees more frequently used communications like postal mail, email, phone calls, and texting. While it is speculation, it is possible that this is a reflection of grantee capacity. For example, non-experimental grantees—with the exception of Phoenix—had a single education navigator and may have felt these methods of communication were a more effective use of scarce navigator time, especially when the navigator was responsible for the whole of the PHA's geographic area instead of having a smaller grouping of assigned buildings.

Exhibit 7: Interaction mode by grantee. The denominator in each proportion is the number of students within that PHA who had any interaction with the navigator. The numerator in each proportion is the number of students with at least one interaction falling into that category.



The variation in the mode of interaction that Exhibit 7 depicts is also a result of variation in how navigators used the interaction tracker (Appendix Section 7.2 shows the definitions of categories that HUD provided to grantees, such as a one-on-one counseling meeting). For instance, examining navigators' free text notes shows that one-on-one counseling encompassed a wider range of activities than sit-down meetings to focus on specific tasks. Exhibit 8 displays a random sample of nine text descriptions of one-on-one counseling. Approximately half of such interactions seem to describe knock and talks. Analysis of all the navigators' free text notes shows variation across grantees in the most frequent words used for interactions marked as one-on-one counseling.[15] For instance, Exhibit 7 shows that Philadelphia reported that over 75 percent of engaged students had a one-on-one counseling session. But navigators there were much more likely to have the word "door" in their notes on counseling sessions than in Seattle or Chicago. This highlights how between-grantee variation in reported engagement likely reflects a mix of different strategies and different data recording practices.

Exhibit 8: Random selection of one-on-one counseling text descriptions

Sat with the student and developed a work plan for applying for scholarships
Door to door

[15] See Appendix Section 7.2 Exhibits 37 and 38.

Door-to-door visits for project soar to introduce myself and the program services
FAFSA
Door-to-door visits for Project SOAR at Legends South to introduce myself and the services available. Was not home. Spoke to grandmother.
Introduction
Interns provided flyer for upcoming soar field trip
Met with [name] to advise her on how to enroll in the Partners in Education program for the summer session. Gave her the new Chicago Housing Authority scholarship information as well.
Resource fair invites

Timing of engagement

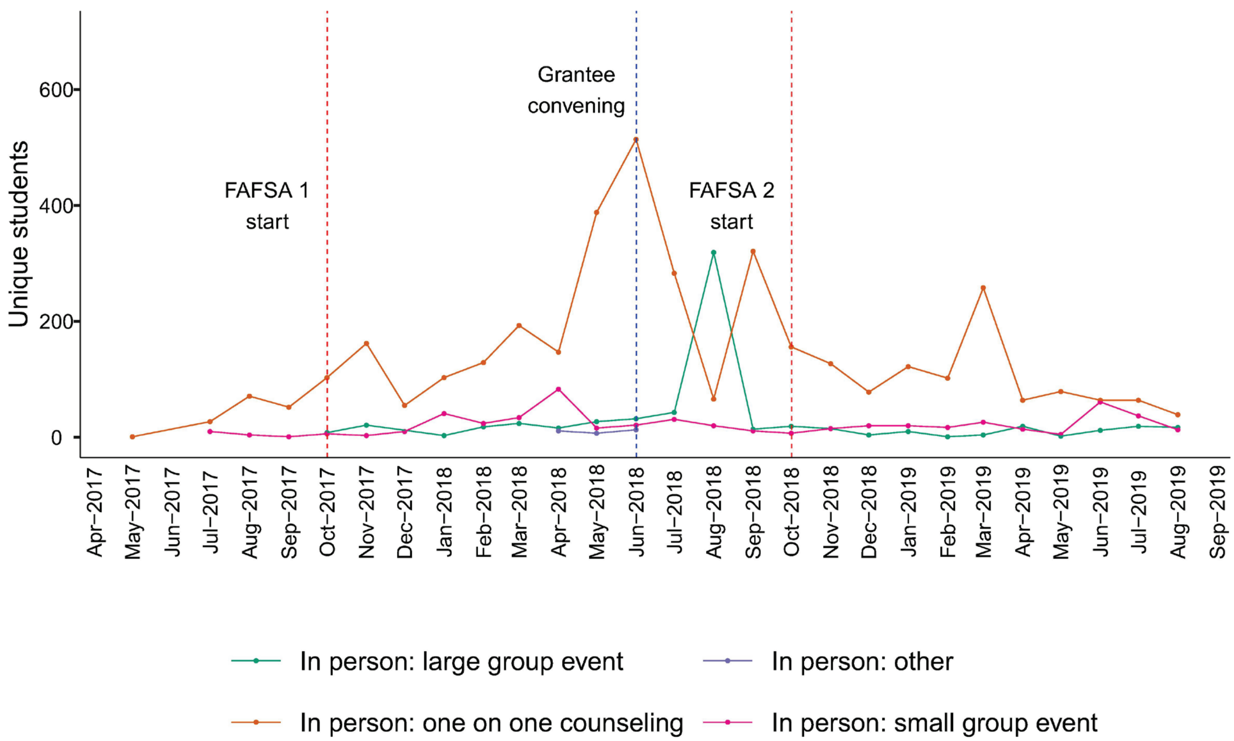
Over time, grantees modified their approaches in an attempt to gain more traction in the community. By the summer of 2018, education navigators in Philadelphia had begun to write personalized messages on the outside of materials they were leaving behind when residents did not answer the door. Communications campaigns also served as a way to communicate success stories or fun events that could encourage interest. Education navigators in Phoenix and Milwaukee found that residents were most likely to read communications when included in rent mailers. Some education navigators began going to other well-attended (non-SOAR) community events as part of their recruitment strategy (as in High Point), while others recognized the importance of working with resident councils to get more buy-in from the community (as in Philadelphia).

Overall, the intensity of interactions seemed to pick up in the summer of 2018. Exhibit 9 shows the number of students engaging with in-person interactions over time. There appears to be a gradual ramp up in activity over the fall of 2017, when navigators were perhaps trying to make initial inroads in the communities. Then there was a sizable increase in activity over the summer of 2018 before activity again faded heading into the fall of 2018.

There are several possible explanations for the increase in activity. First, as several grantees noted, it took time to develop relationships in the community; going into the second year of the SOAR grant, navigators had more ongoing relationships with students and their families. Also, the summer may have given students more time to meet with navigators without the pressure of balancing school, homework, and other family responsibilities. One thing that may temper these explanations is that a grantee convening took place in June, and there was a push by HUD for sites to submit data to inform some of the presentations. It is possible that some of the spike in activity was more related to variation in data quality than a reflection of on-the-ground interactions. The data seem to support the idea that grantees and navigators needed a good amount of time to achieve more robust operations. Even a full year may not have been enough time to achieve steady-state operations. For example, the CEO of one grantee thought that the 2-year timeline for the project was too short to see real change and develop trust with residents.

Based in part on the site visits, a decision was made to concentrate on the second year of the grant for the main specifications of the analysis. Concentrating on outcomes related to the first year of the grant would not be a test of a fully functioning program. Navigators were spending more time trying to engage students than helping them with concrete tasks during a good portion of the 2018–2019 FAFSA season. The peak in activity in the summer of 2018 suggests that navigators were in a much better position to help students with tasks related to the 2019–2020 FAFSA season and that the analysis would be more likely to detect program effects by focusing on outcomes related to the 2019–2020 school year.

Exhibit 9: Frequency of in-person interactions over time



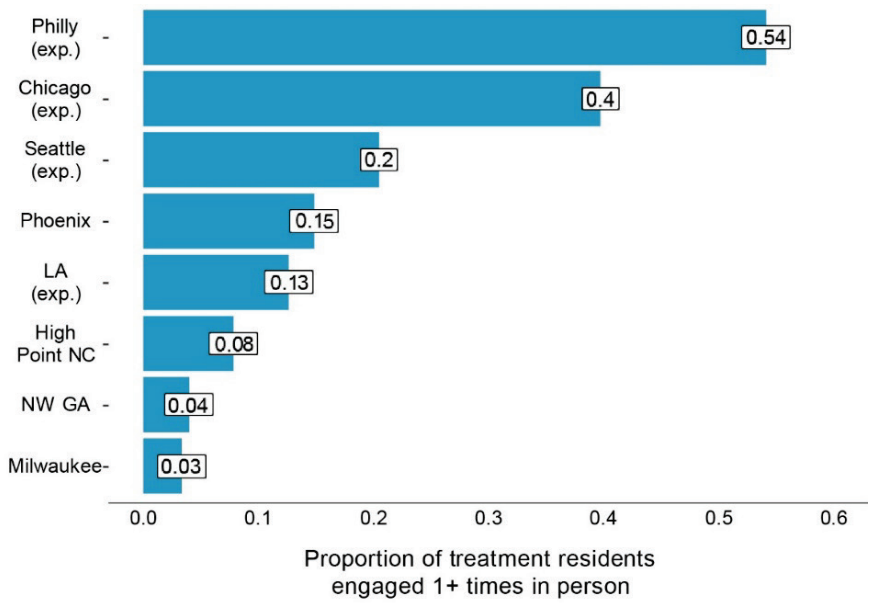
Outreach challenges

Navigators encountered resistance to participation in a number of ways. First, residents were reluctant to engage with new services that they thought could soon disappear, which they had experienced before. Second, nearby schools often had limited resources and poor academic preparation for young adults living in public housing, making the college application process seem even more daunting. Behavioral issues in school, including experiences with trauma, were prevalent among the population. Third, some students were parents themselves, which made it difficult for them to balance school with childcare (both because of cost and balancing schedules), and many did not think postsecondary education was a possibility and were not aware of childcare supports that were potentially available. Finally, in several communities, immigration/citizenship status presented a barrier. Undocumented students could face additional bureaucratic hurdles or have to fill out additional information, which made the process more complicated. Undocumented students in some cases were not eligible for as many programs, scholarships, or other types of financial aid, and they could be reluctant to provide information if they feared that it would be used for an immigration enforcement action. All of these things contributed to reluctance among students to participate in a program focused on postsecondary education.

The struggles in recruitment are apparent in the activity tracker data. Exhibit 10 shows the proportion of eligible residents who had at least one in-person interaction with a navigator. Navigators in only two grantees met with at least 40 percent of eligible students in person, and in most grantees, navigators met with fewer than 20 percent of eligible residents in person. This does not necessarily indicate that grantees were completely unsuccessful. It is difficult to define what a "good" level of interaction is, given that different grantees may have decided to focus on more intensive interactions with a smaller set of individuals, and the numbers may reflect different strategic decisions. Still, given what grantee staff reported, combined with the administrative data, it seems like outreach was one area where grantees tended to struggle.

The level of engagement has implications for the analysis. Whether or not grantees were achieving the right balance for caseloads according to their program model, the primary specification for analysis estimates effect of the intent to treat or the effect of the program on all eligible students, regardless of whether the students heard from or met with a navigator. Relatively low engagement rates could lead to underestimating the program's effect on those who were engaged. As discussed in Section 5.1, a secondary analysis estimates the effect of the program on those who have at least one in-person interaction with a navigator to account for low levels of engagement.

Exhibit 10: Proportion of eligible residents with at least one in-person interaction. The denominator is either: (1) all students aged 15–20 in the non-experimental PHAs, or (2) students aged 15–20 residing in treatment AMPs in the experimental PHAs.



Demographic characteristics of engagement

Descriptive tests reveal that certain demographic characteristics were associated with a higher (or lower) likelihood of a student having any in-person interactions with a navigator. Exhibit 11 shows the results from a model that separately regresses the binary indicator of any in-person meeting on each student demographic characteristic.[16] The results show that, overall, navigators were more likely to meet with female residents, more likely to meet with students at the older end of the 15 to 20 age range, and less likely to meet with students in larger households or who identified as Hispanic and non-White (relative to non-Hispanic White). However, as Exhibit 10 shows, these overall relationships may just be highlighting the between-grantee variation in rates of in-person engagement; for instance, Los Angeles has both a higher concentration of Hispanic residents and a much lower in-person engagement rate.

Exhibit 12 shows the results when analyzing variation within each grantee by including a grantee fixed effect. Although most relationships remain the same, the analysis shows, for instance, that within grantees that have a mix of citizens and non-citizens, non-citizen students are more likely to meet in person with a navigator.

[16] Because the analysis is meant to be descriptive rather than causal, and because attributes such as a student's race/ethnicity are correlated with household income, we ran separate regressions for each type of attribute: citizenship, race/ethnicity, household income, and total household members 15–20.

Exhibit 11: Demographic predictors of engagement: overall

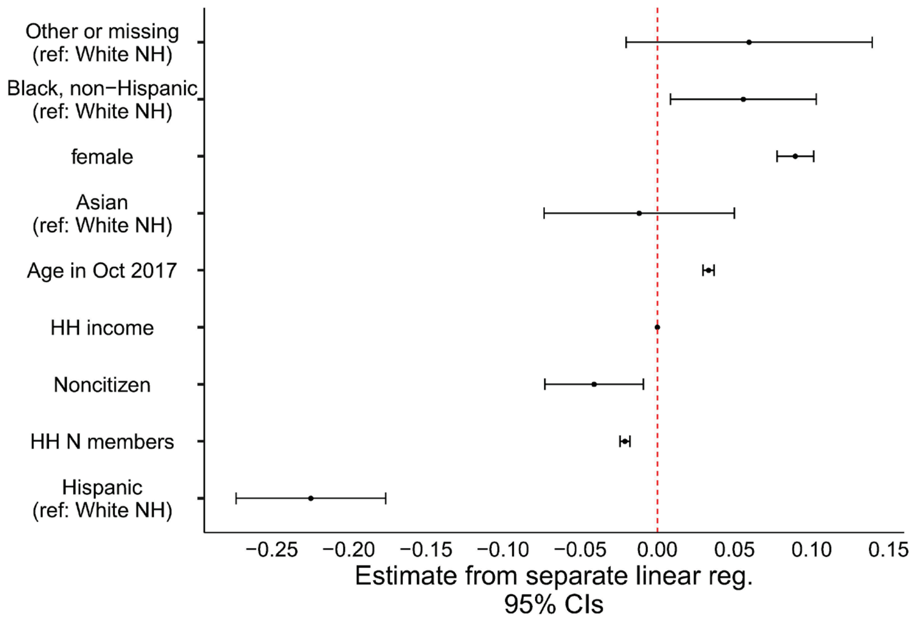
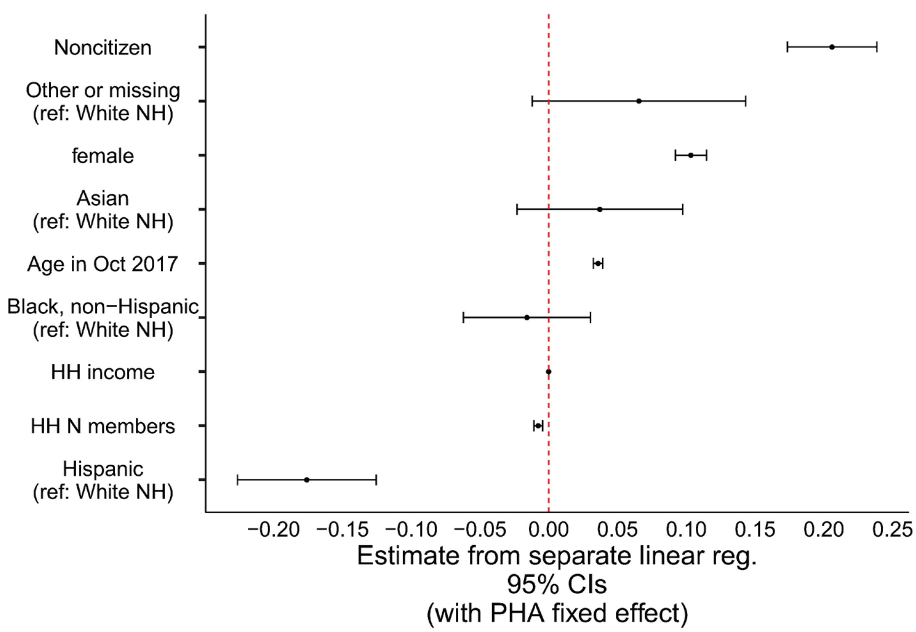


Exhibit 12: Demographic predictors of engagement: variation within each PHA



3.6 Clusters of variation

The variation in engagement shows two interesting patterns. First, navigators were meeting with female students who already have higher odds of attending college (Clark and Shi, 2020). These students may have required less help from navigators. For example, they may already have had a general plan and may have only needed help with smaller steps as part of that larger plan.

Second, navigators were meeting with students who may have faced unique and complex barriers to college, such as non-citizens who had to navigate eligibility rules. For these students, navigators may have needed to provide more-intensive or more-individualized assistance.

Grantees varied not only in terms of engagement when it is defined as a binary measure of whether or not a student met with a navigator at least once in person, but also in terms of engagement as defined as the complete set of interactions navigators and students had. Students with at least one in-person meeting varied along two additional dimensions of engagement: the "dose" of meetings (with some students having long trajectories of repeated meetings with navigators and others having only one meeting) and the presence of the student's parents at that meeting.[17]

Analysis examining these other dimensions clustered students' interaction trajectories—their full history of interactions across the entire period—using sequence analysis.[18] We defined seven "states" that a student could be in during a particular month of the study window:[19]

1. No meeting between student and navigator.
2. Navigator met with parent only.
3. Navigator met with student only.
4. Navigator met with parent and student together.
5. Navigator met with school official only.
6. Navigator met with school official and student and/or parent.
7. Other in-person meeting.

Each student's sequence is structured as an ordered set of interactions, and an optimal matching algorithm and k-means clustering groups together students with similar sequences of navigator interactions.[20] Exhibit 13 shows five distinct clusters of interaction histories identified in the analysis. The clusters show that the binary measure of any meeting or no meetings hides significant variation in the types of and time sequencing of meetings.

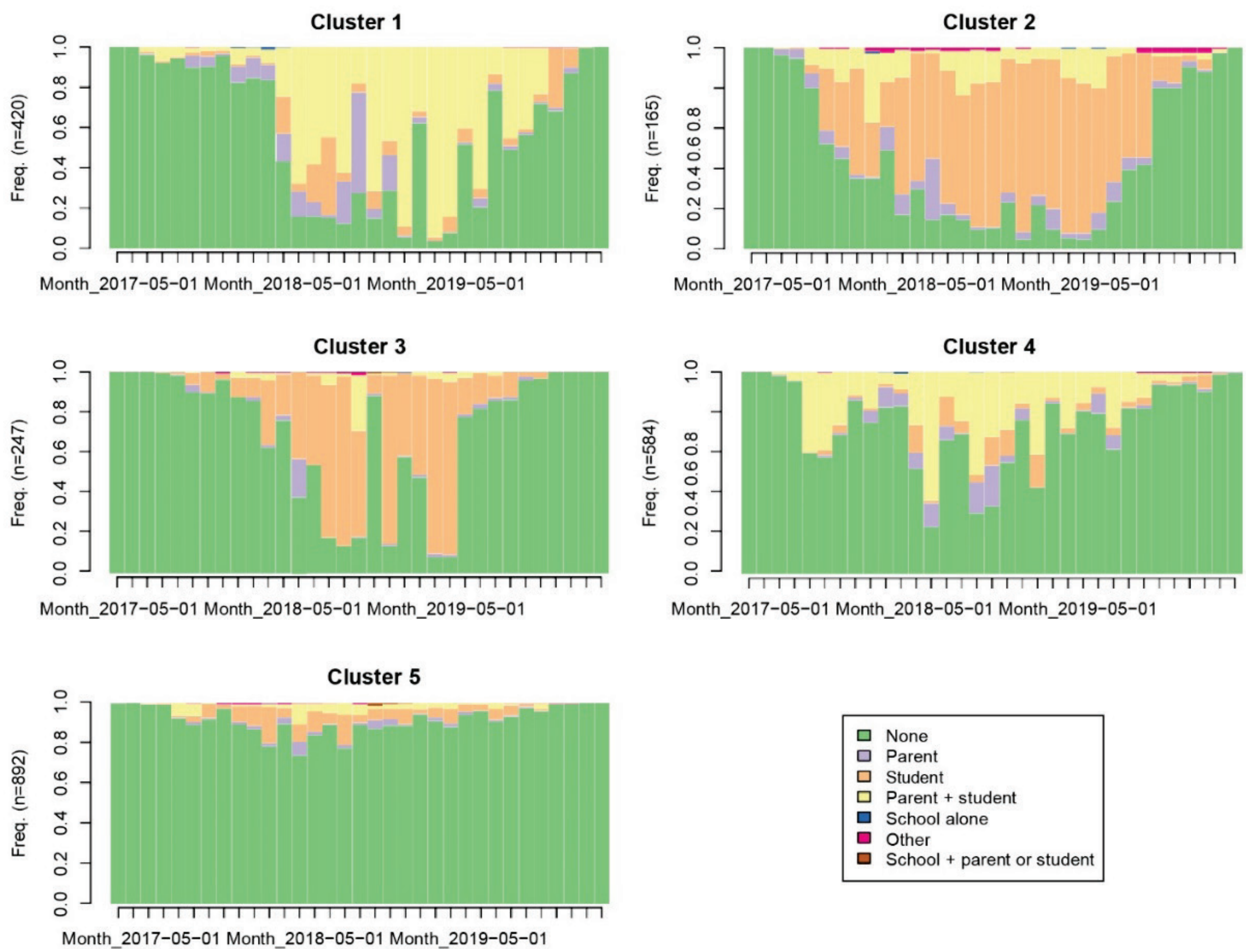
The five clusters can be further grouped into two broad types: students whose parents co-attended the meetings with navigators (Clusters 1 and 4) and students whose parents rarely attended (Clusters 2 and 3). A regression analysis shows that households with higher total household income and fewer teenagers aged 15 to 20 were more likely to fall into the "heavier parent involvement" cluster, with Philadelphia and Seattle also having more parent involvement in meetings than Chicago and LA. Engagement varies considerably across grantees and is accompanied by significant heterogeneity in the type and sequencing of in-person meetings.

---

[17] As shown in Section 7.2, the interaction tracker category was broader than the student's biological or legal parent, defining parent/guardian as "A primary caregiver for FAFSA eligible AMP resident." For shorthand, we use the phrase "parent," but the results should be interpreted as the involvement of the student's primary caregiver.
[18] We use `traminer` in `R` for the analysis.
[19] For students with multiple states in the same month, we use the first meeting of that month; this was rare enough that we do not believe a different decision—such as using the last meeting of the month—would alter findings.
[20] Exhibit 39 in the Appendix shows an example with twenty randomly-chosen students.

Given this heterogeneity, it is possible to define engagement in numerous ways, and estimates of the effect of the program on those who were treated may be sensitive to decisions about how engagement is defined.

Exhibit 13: Clustering interaction sequences among students with at least one in-person meeting. K-means clustering with k = 5 shows five distinct clusters of interaction trajectories. Cluster 1 is characterized by a high prevalence of the navigator jointly meeting with the student and his or her parent. Cluster 4 represents a similar focus on joint meetings but with fewer meetings overall. Clusters 2 and 3 are characterized by a high prevalence of student-only meetings, with Cluster 2 having these more spread out over the study period and Cluster 3 having a peak in summer of 2018. Finally, Cluster 5 is characterized by many months of no in-person meetings.



4 Impact analysis

The impact analysis of the effectiveness of Project SOAR includes two distinct parts: an experimental analysis of four grantees and a non-experimental analysis of the other five grantees.[21] Four of the grantees—Chicago, Los Angeles, Philadelphia, and Seattle—had particularly high youth-to-navigator ratios to the extent that it was impractical to have the navigators attempt to serve everyone. As a feature of the grant, these four PHAs were chosen to have navigators offer assistance only to a group of randomly selected youth (the treatment group). The remaining youth were not eligible for SOAR services (the control group). These four PHAs comprise the experimental component of the impact evaluation. This analysis compares the postsecondary outcomes of the youth selected for the treatment group with outcomes of the youth selected for the control group in those four PHAs.

The other five grantees were not included in the experimental evaluation for various reasons; they were included in the non-experimental impact evaluation. Three of the five PHAs not included in the experimental evaluation had a smaller relative number of eligible youth (High Point, Northwest Georgia, and Prichard). HUD chose not to ration or restrict navigator assistance in these PHAs to a subset of eligible individuals because it was not anticipated that there would be excess demand (for example, caseloads were small enough that it was expected that navigators could attempt to engage with all eligible students[22]), and these PHAs were excluded from the experimental component of the impact evaluation. One PHA (Milwaukee) had similar youth-to-navigator ratios as the experimental PHAs but declined to participate in the randomized evaluation. Another PHA (Phoenix) had a configuration of buildings that made random assignment less feasible.[23] In each of the five PHAs participating in the non-experimental component, navigators attempted to provide assistance to all residents in the targeted age range. The analysis of these five grantees uses quasi-experimental methods to estimate the impact of Project SOAR by comparing the postsecondary outcomes of youth in these five PHAs to youth in (similar) PHAs that were not selected for the grant.

4.1 Data and data structure

There are several sources of data which inform the impact analyses.[24]

The Public and Indian Housing Information Center (PIC): PIC is a HUD system developed to collect and maintain certified tenant and other data for processing from Public Housing Agencies. The PIC data extracts are point-in-time quarterly extracts created by HUD for research, reporting, and monitoring purposes.

---

[21] All of these choices were preregistered in an analysis plan, posted at this link on the OES website on February 12, 2020. The analysis plan upload date can be verified at this link. The analytic choices were made prior to us taking any outcome data into possession. Deviations from the analysis plan are noted in Section 4.6.

[22] Although this assumption may have been incorrect in retrospect.

[23] Phoenix had only four AMPs (described below) eligible for assignment, two of which were very large and two of which were small, making it likely that all would be selected to the treatment group given the random allocation mechanism.

[24] The analysis plan included some exploratory analysis based on records from one local high school district; however, these data ultimately were not available.

- *What these data help us investigate:* PIC contains the identifiers used for matching to outcome data held by the Department of Education (described below). PIC also contains basic demographic information for individuals and households, including age, race, ethnicity, household income, and household size. These data allow us to control for baseline demographic characteristics of the public housing residents that remained imbalanced following randomization.
- *Limitations:* The data are limited to information collected for the purposes of verifying eligibility for housing assistance. As such, it lacks residents' grades, test scores, or other academic outcomes relevant for college going.
- *Analyses:* Experimental.

Picture of Subsidized Households (PSH): The PSH is a publicly-available data source containing demographic characteristics of those residing in subsidized housing. The analysis restricts the data to the Public Housing program.

- *What these data help us investigate:* While PIC data inform (1) the main experimental results focused on five grantees and (2) the analysis of demographic variation in interactions across all grantees, the non-experimental analysis involves comparing the five non-experimental grantees to all other PHAs not receiving a SOAR grant. The scale of a customized extract of these data—over 2,000 PHAs observed over a period of 13 years—led us to use the publicly-available PSH data to collect resident demographics for these donor PHAs.[25]
- *Limitations:* While PIC data are individual-level and restricted to age-eligible residents, the PSH data are aggregated to the PHA and contain demographics of all PHA residents. Depending on the similarity between (1) the demographics of the entire PHA and (2) the demographics of PHA residents aged 15 to 20, the former is only a general approximation for the latter.
- *Analyses:* Non-experimental.

Enterprise Data Warehouse and Analytics (EDWA): EDWA is a data warehouse maintained by the Department of Education that contains information on students' interactions with the federal postsecondary educational system. EDWA contains information on FAFSA completion, postsecondary enrollment, and federal student aid. A memorandum of understanding between HUD and ED allowed for HUD to send person-level files to ED to be matched to EDWA using individuals' Social Security numbers, names, and dates of birth. ED provided the AMP- and PHA-level aggregated outcome data used in the main analysis and conducted the individual-level complier analysis on individual-level data.

- *What these data help us investigate:* These data include the main outcomes of the study. They were used to estimate if there were any changes in FAFSA completion or other postsecondary outcomes for students eligible for SOAR services.

[25] For comparability, we also use the demographics from PSH for the grantees in the synthetic control analyses.

- *Limitations:* Navigators served residents during two FAFSA completion cycles (Exhibit 3). One began shortly after navigators were hired (October 1, 2017), ending June 30, 2019. The other began when navigators were about a year into their tenure, starting October 1, 2018 and ending June 30, 2020. As Exhibit 3 highlights, due to the timing of our analysis, we only observe FAFSA completion through March 2020, which truncates the full 2019–2020 FAFSA cycle.
- *Analyses:* Experimental; Non-experimental.

College type: In addition to affecting whether students enroll in college, the intervention might also impact the type of college at which students enroll. Navigators were instructed to work with students to find a college in line with their preferences and constraints, constraints that might include family obligations (that mean students prioritize commuter schools) or financial obligations (that mean students prioritize shorter degree programs). Following prior work (Chetty et al., 2017; Deming et al., 2015), we grouped colleges into four tiers of selectivity:[26]

1. Highly selective or selective colleges (Tier I in results presentation): these encompass tiers 1 through 6 in the Barron's rating system, or about 1,200 colleges.
2. Non-selective 4-year colleges (Tier II in results presentation): tiers 7 and 8.
3. Non-selective 2-year public and not-for-profit colleges (Tier III in results presentation): tier 9.
4. Non-selective private, for-profit colleges, 2 or 4-year (Tier IV in results presentation): tiers 10 and 11.

- *What these data help us investigate:* If there are general increases in postsecondary enrollment, these data can help determine what types of schools the enrollment increases are concentrated in.
- *Limitations:* College enrollment is a function of (1) which colleges a student applies to, (2) which colleges accept the student's application (if applicable), and (3) which college a student chooses to enroll in. By observing enrollment, we miss possible impacts on outcomes such as the breadth of colleges to which a student applies.
- *Analyses:* Experimental, Quasi-experimental.

4.2 Outcome variables to be analyzed

The primary outcome of interest is FAFSA completion for the 2019–2020 academic year. The FAFSA cycle for the 2019–2020 academic year begins October 1, 2018, and ends June 30, 2020. Exhibit 3 shows the relationship between academic cycles and the timeline for navigator services. The impact analysis restricts the analytic sample to the population of students who are expected to be high school seniors or older, a group for whom FAFSA completion is more relevant.

[26] We begin with the same 12 original tiers in the Barron's system as past studies but group these tiers differently to ensure adequate cell sizes in each tier.

Exhibit 14 illustrates the "effective window" for FAFSA completion in the two focal cycles for the experimental analysis. The effective window is defined based on school registration rules around student age. In particular, most school districts in the study require a student to be 5 years old on or around September 1 of the year in which they enter kindergarten. The effective window uses this rule to define rising high school seniors as those who turn 17 as of September 1 of the relevant outcome year. All students who are between 17 and 20 years old at some point during the effective window are age-eligible for that cycle. If the student meets the age eligibility requirement and lives in an area eligible to receive SOAR assistance at some point during the effective window, he or she is included in the analytic sample for that year.[27]

Exhibit 14: Time windows for each FAFSA cycle

Cycle	Actual window	Effective window
Primary: FAFSA 2019–2020	10/01/2018–06/30/2020	09/01/2019–08/31/2020
Secondary: FAFSA 2018–2019	10/01/2017–06/30/2019	09/01/2018–08/31/2019

Secondary outcomes include rates of postsecondary enrollment, institution type (public, private, or proprietary), program length (2-year or 4-year), program selectivity, and Pell Grant receipt. The analysis only explores these outcomes for the 2019–2020 school year, when it is more likely SOAR has an effect.

4.3 Experimental impact analysis

This section describes the statistical models and hypothesis tests for the experimental impact analysis.

Random assignment process

The experimental design uses an administrative unit called the Asset Management Project (AMP). Generally speaking, AMPs are individual buildings or groups of buildings in close proximity.

The decision to use AMPs as the unit of randomization was based on two main considerations. First, randomly assigning individuals (including within buildings) would be logistically challenging for navigators. Individual random assignment would not only make it more difficult to verify eligibility, it would make it much more time consuming for navigators to reach eligible individuals. For example, a navigator may have to recruit and assist individuals in different developments over a large geographic area. Even with the presence of multiple navigators, the area any one navigator would have to cover would mean a substantial amount of time would be devoted to travel rather than direct assistance. Assigning AMPs guaranteed that navigators would have more limited geographic areas in which they would need to operate. Second, using

[27] Note that this means a resident may only be eligible for SOAR for a short period of time, but this decision reflects the fact that navigators could help anyone eligible for any length of time. For the non-experimental analysis, the logic is similar but extended back to start with the 2007–2008 FAFSA cycle. Appendix 7, Exhibit 40 shows in more detail the eligibility algorithm.

AMPs as the unit of randomization was expected to minimize crossovers, which is defined as individuals assigned to the control group receiving assistance from navigators. It was anticipated that it would be more difficult for navigators to both verify eligibility and to turn individuals in the control group away if individual-level random assignment was used. For example, it would be difficult to build an on-site presence in a development while having to explain that only certain residents were eligible for SOAR. As described below, this was still a concern when using AMPs as the unit of randomization, due to the close physical proximity of different buildings in some cases.

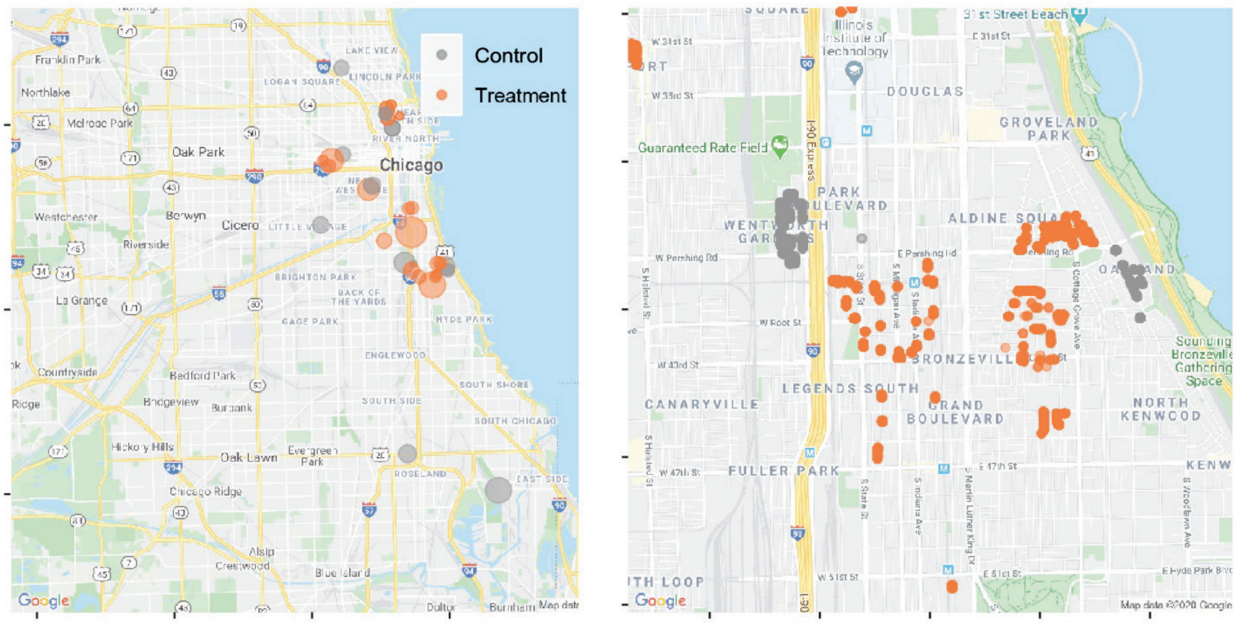
While the choice to use AMPs as the unit of random assignment made program implementation easier, it also limited the statistical power of the experiment. Relative to the number of eligible individuals, there were a small number of AMPs. This limitation is discussed more in Section 6.

Office of Evaluation Sciences relied heavily on local PHA knowledge to select AMPs for the experiment that would have clear geographic, and sometimes social, boundaries. In a majority of cases, individual AMPs were treated as unique randomization units; however, some AMPs were grouped together to avoid confusion and limit possible non-compliance because of unclear boundaries. Additionally, each PHA provided HUD with a list of AMPs they requested be removed from consideration for various reasons—for example, due to geography that would make travel difficult for education navigators, overlap with other similar resident service programs (such as Jobs Plus), or because of other local knowledge. After groupings and exclusions, a total of 78 AMPs were eligible for random assignment.[28]

The left panel of Exhibit 15, uses the Chicago Housing Authority as an example of how AMPs were randomized. The figure aggregates households to the AMP level and shows AMPs randomized to treatment (orange) or control (gray). It shows that AMPs can be located in similar neighborhoods and randomized to different conditions, which ideally prevents confounding between treatment status and characteristics like the quality of neighborhood schools. The right panel of the exhibit zooms in on one Chicago neighborhood, Bronzeville, and three AMPs, with each dot representing a household and its treatment status. It shows that there can be households in distinct AMPs that are located in similar neighborhoods and also shows that AMPs vary in how geographically clustered or distributed their units are.

[28] In the discussions that follow, we use the term “AMP” to refer to these 78 AMPs, even though some are technically “modified AMPs,” or AMPs grouped together for the purpose of randomization.

Exhibit 15: Map of AMP randomizations in Chicago. The *left panel* illustrates randomization at the AMP level. AMPs are placed at the mean latitude and longitude of units. The sizes of the dots are scaled to the number of age-eligible youth in each AMP. The *right panel* zooms in on three AMPs. The map shows how the randomization helped minimize potential spillovers while still resulting in neighborhoods with both treatment and control group students due to the clustering of different AMPs in the same neighborhood.



Navigators were instructed to treat as many age-eligible youth within the treatment AMP as they were able to and were instructed not to serve anyone from control AMPs. In practice, navigators were given rosters of youth living in treatment AMPs with which they could verify the eligibility of youth. If students who did not reside in a treatment AMP attempted to engage, navigators were asked to give them a list of other community resources or to refer them to other PHA staff for assistance, but navigators were instructed not to provide any personalized assistance.[29]

The following randomization procedure was used to determine which AMPs would be treated, with code for the procedure provided in Appendix Section 7.4:

1. For each PHA, AMPs that were too large or too small were removed from consideration. AMPs with fewer than 10 age-eligible individuals were removed. AMPs with more age-eligible youth than a single navigator could support were also removed.
2. AMPs were sorted by a random number, with the first AMP in the sorted list assigned to treatment.
3. Subsequent AMPs were assigned to treatment by progressing down the randomly sorted list until the final AMP assigned to treatment exceeded the maximum workload.
4. All remaining AMPs were assigned to control.

[29] It is possible at group events for youth from control AMPs to be present because navigators typically did not find it feasible to check participants against the rosters given the format.

Statistical models

The primary model estimates the intent to treat effect by applying OLS to a regression of the outcome of interest (y_a) on an indicator for treatment (T_a), where a indexes the modified AMPs used as the unit of randomization. The main specification also controls for two blocking variables (X): a dummy indicator for each grantee and a continuous measure of the number of age-eligible youth at the time of randomization. Additionally, because the assignment mechanism results in larger AMPs having slightly higher probabilities of being selected for the treatment group, the model uses inverse probability weights to account for the estimated probability of selection into the treatment and control groups.[30] Finally, all models use the Lin estimator (Lin et al., 2013), which entails:

1. Mean centering each covariate, which we will refer to as $\tilde{X}$: $\tilde{X} = X_i - \bar{X}$
2. For each model that includes covariates, regressing the outcome on the treatment, mean-centered covariates, and interaction between the two.

More formally, the analysis estimates the following linear model:

$$y_a = \beta_0 + \beta_1 T_a + \gamma \tilde{X}_a + \delta T_a \tilde{X}_a + \epsilon_a \quad (1)$$

Each outcome of interest is computed as a proportion, with the (1) numerator being the count of students who completed the FAFSA for the 2019–2020 school year and (2) the denominator being the students defined by the eligibility criteria outlined in Section 4.2.

The estimate of interest is β_1, which is the estimated effect of the intervention on AMPs randomized to treatment.

A second model includes AMP-level baseline characteristics as covariates.[31] As discussed in more detail in Appendix Section 7.6, there were differences between the AMPs randomly selected to be treated and those selected for the control group. In particular, AMPs with a higher proportion of Black residents and AMPs with higher household income were more likely to be selected into the treatment group in Seattle, and AMPs with a higher proportion of non-citizens were more likely to be selected into the treatment group in Los Angeles. These imbalances likely stem from the fact that there were a small number of AMPs in each PHA, and buildings tend to have fairly homogeneous populations. Although underlying characteristics of AMPs would expect to balance out in a large enough sample, this sample was too small to achieve perfect balance. Using covariates can bring the sample back into mechanical balance, but it also can create unstable results given the small sample issues.

[30] Probabilities were estimated via simulating the randomization process m = 1000 times. The weights used take the simulated probability of selection to treatment for each AMP and calculate the inverse probability weight as $\frac{1}{(T * p + (1 - T) * (1 - p))}$, where T is an indicator variable for being assigned to the treatment in actuality and p is the estimated probability of selection into treatment.

[31] Since PIC is updated quarterly, the characteristics were taken from the PIC file from the quarter preceding May 2017, so all covariates were measured at baseline.

The adjusted model includes covariates that are likely correlated with the college-going behavior of youth in that AMP and/or the households' openness to navigator help:[32]

- Percentage of households that self-report Black, non-Hispanic (White held out as reference category).
- Percentage of households that self-report non-Black, Hispanic.
- Percentage of households that self-report other race, Non-Hispanic.
- Percentage of households that contain non-citizens.[33]
- Mean total annual household income.

Similar to the main specification, a secondary specification uses the Lin estimator with mean-centered versions of the covariates. In the model, Z denotes the combined matrix of (1) AMP-level covariates, and (2) the blocking variables included in the above specification.

$$y_a = \beta_0 + \beta_1 T_a + \gamma \tilde{Z}a + \delta T_a \tilde{Z}a + \epsilon_a \qquad (2)$$

The first two models examine the effect of an AMP being randomized and analyze the impact using the outcomes of all students in the AMP—or the intent to treat estimand. Section 3.5 shows, however, that rather than engaging all students, navigators met with fewer than half of eligible youth in treatment AMPs. As a result, there is a good amount of non-compliance: the presence of a navigator in an AMP increased the likelihood that a youth engaged with that navigator, but a sizeable fraction of youth in treatment AMPs never engaged. Similarly, there was at least some non-compliance in the form of control students receiving navigator help.[34]

As a result, a third specification estimates the effect of SOAR on engaged students—the complier average causal effect. Engagement is defined as at least one in-person meeting between a navigator and the student. This analysis uses individual rather than AMP-level data, with individuals now indexed by i. The analysis estimates the following two-stage model:

1. A model predicting whether or not a youth engages with the navigator as a function of that youth's treatment status and baseline covariates (Z) measured at the youth level:[35]

$$\text{engage}_i = \beta_0 + \beta_1 T_i + \delta Z_i + \epsilon_i$$

[32] For instance, mixed citizenship families may face greater confusion about eligibility for aid.
[33] This includes both "eligible non-citizens"—non-citizens who are eligible to live in public housing—and "ineligible non- citizens"—non-citizens who are not eligible for housing assistance.
[34] Unfortunately, navigators may have only recorded interactions with treatment group youth and not the interactions with control group youth, which means we are able to measure compliance in the form of those assigned to treatment not receiving treatment but are less confident that we observe compliance in the form of those assigned to control receiving treatment.
[35] We used the following covariates drawn from PIC that were correlated with navigator engagement/outcomes, and that remained imbalanced between groups: gender, race/ethnicity (Black, Non-Hispanic; Hispanic; or other), total annual household income, total household members, PHA dummy.

2. A model using the predictions from stage one to estimate the Treatment on Treated effect, or the estimated effect among the youth whom the navigator engages.

$$y_i = \beta_0 + \beta_1 \widehat{\text{engage}}_i + \epsilon_i$$

Due to data limitations, the outcome variable in this first model likely had systematic measurement error—for instance, if navigators consistently *underreport* engagement, the data will undercount the number of youth served, and the estimate will potentially overestimate the treatment effect.

Inference criteria, including any adjustments for multiple comparisons

The decision rule was based on p-values and confidence intervals generated using a permutation approach that uses the randomization procedure described above, with any two-tailed p-value less than 0.05 considered statistically significant (for example, randomization inference, or RI). The analysis studying FAFSA completion and using the p-values from RI is considered confirmatory. Exhibit 16 outlines the calculation for p-values from a long-form data frame where there is:

- An observed treatment effect.
- Treatment effects from m = 1, 000 permutations.

Exhibit 16: Code snippet for randomization inference p values

```
ri_p_df =   perm_obj %>% filter(model == model_string & outcome_var ==
            outcome_name) %>%
            summarise(ri_p = mean(abs(permuted_coefs) > abs(obs_tx_coef)))
```

A separate analysis of FAFSA completion that uses p-values from linear regression and heteroskedastic-consistent standard errors rather than randomization inference serves as a robustness check on the RI-based inference, with the caveat that the small number of AMPs makes assumptions behind those p-values less credible.[36]

In addition to the primary outcome—FAFSA completion during the 2019–2020 cycle—we also pre-registered and examined the following secondary outcomes: FAFSA completion during the 2018–2019 cycle, college enrollment by selectivity/type, and receipt of a Pell Grant, all for the 2019–2020 school year only.

4.4 Non-experimental impact analysis

The non-experimental analysis of Project SOAR is intended to add to the overall description of programmatic effects but is considered exploratory given the stronger assumptions the method requires.

[36] We used the HC2 specification to estimate standard errors, which has a small sample correction; however, the p-values from the randomization inference procedure generate exact p-values that rely on fewer sample assumptions.

The effect of Project SOAR in the nonexperimental PHAs is estimated using a synthetic control method. The basic intuition behind the synthetic control method is to create a relevant comparison unit for a treated unit by using the data to create a weighted composite of other units in a donor group. The method was first introduced by Abadie and Gardeazabal (2003) and has become increasingly popular in recent years, with several authors suggesting extensions of the basic intuition. This analysis uses the more recent three-step process as described in Xu's generalized synthetic control method, implemented using `gsynth` in R (Xu 2017):

1. Model latent "factors" of outcomes using only the (untreated) donor pool and both pre- and post-treatment data.
2. Use the results from Step (1) to find the factor loading for each treatment unit.
3. Use those factor loadings to predict the counterfactual outcomes for the treated unit(s) in the post-treatment period(s).
4. Compare the treated unit(s) counterfactual values in the post-treatment year(s) to its observed values.

The result is an average treatment effect on the treated (ATT) for each year. For the FAFSA outcomes, 11 of the 13 years (the 2007–2008 to 2017–2018 FAFSA cycles) we consider pre-treatment; 2 of the 13 years (the 2018–2019 and 2019–2020 FAFSA cycles) we consider post-treatment. For the college enrollment outcomes, we consider 11 years pre-treatment (what we call the 2008–2009 to 2018–2019 college cycles) and 1 year (2019–2020) post-treatment.

Donor pool

The donor pool was initially made up of all PHAs with the exception of the non-experimental PHAs and the four experimental PHAs. As Section 7.10 describes in greater detail, the Department of Education's redaction of small cell sizes—cases where fewer than ten age-eligible students in a PHA completed the FAFSA in a given year—meant that a significant proportion of the PHAs had some years of the outcome data redacted. This means that the effective donor pool is restricted to PHAs that were similar in size to the non-experimental grantees. While these PHAs are only a small proportion of PHAs overall, Exhibit 17 shows that they house the majority of public housing residents.

Exhibit 17: Number of PHAs and residents in donor pool for synthetic control analysis

Status	N PHAs	N residents (2016; all ages) possible donor residents	% of
In donor pool (main specification)	1,137	1,791,548	88%
Excluded from donor pool	1,804	244,497	12%
Treatment	5	16,150	–

Inference criteria, including any adjustments for multiple comparisons

The non-experimental analysis uses the parametric bootstrapping implemented in `gsynth` for inference. Broadly, the procedure holds out one control unit to be treated as a fake "treatment" unit and uses the remaining control units to re-estimate the procedure and predict outcomes for the left-out unit. Uncertainty estimates are then based on differences between the predicted outcomes for that held-out control unit and its observed outcome, repeated for all control units. As with the main analysis, we use $p < 0.05$ as the threshold for statistical significance.

4.5 Preferred specification

As pre-specified, the main confirmatory specification is the intent-to-treat effect of the program on FAFSA completion during the 2019–2020 cycle, controlling only for a PHA dummy and the number of youth in the AMP at the time of randomization.

We prefer the ITT effect over the complier analysis because it more closely approximates the relevant policy question: what happens if PHAs implement Project SOAR? In other words, the ITT estimand assumes that in the course of normal program operations, some students will choose not to engage with navigators, but that decisions about program efficacy are made taking some level of non-compliance as a given. An ITT analysis cannot say whether or not the program model being tested resembles the policy ideal closely enough. For example, the ITT impact analysis does not assess whether or not there was adequate staffing or if grantees were given enough time to develop sufficiently mature programs to engage with the expected number of individuals.

The complier analysis of the effect on students who met with a navigator at least once only addresses the effectiveness of SOAR conditional on being engaged with a navigator. This estimate may help provide some idea as to what a maximum effect could be, but the estimand is comparing only the types students who are interested enough or motivated enough to engage with navigators and is not representative of the eligible population as a whole. As such, we use it to provide more context to the ITT analysis but treat it as a secondary, exploratory outcome.

Second, the focus on FAFSA completion during the 2019–2020 cycle allows for a test of more mature program operations.

Finally, the more parsimonious set of controls helps us mitigate issues in the AMP-level analysis of unstable treatment effect estimates that stem from common support issues—for instance, AMPs within a PHA having very low percentages of "Other race/ethnicity" or non-citizen residents. These issues can also lead to unstable weights in the complex weighting procedure, which involves randomly re-assigning AMPs to the treatment group to reweight the data by the inverse probability of treatment.

4.6 Deviations from the analysis plan

There were no major deviations from the analysis plan. However, the analysis did deviate in the following ways, largely due to data availability or time constraints:

1. **The analysis plan discussed using data from Seattle Public Schools (SPS) to examine academic characteristics of the students that navigators engage rather than just the demographic characteristics in PIC.** The SPS data on student academic characteristics were not available, so the analysis of variation in navigator engagement is restricted to the demographic characteristics available in PIC.
2. **Model 2 uses a slightly different set of covariates than specified due to data availability.** The analysis plan prespecified using the following three covariates which were not available for the analysis:

    - Average highest grade of education completed by household head.
    - Percentage of household heads employed full-time.
    - Percentage of families homeless at time of admission to the housing program.

3. **The synthetic control analysis does not use PHA-level covariates.** The implementation of the generalized synthetic control method in `gsynth` did not reliably allow us to adjust for covariates.[37] Therefore, we restrict the role of PHA-level covariates to comparing the demographic composition of the treatment PHAs to the composition of PHAs who were in the donor pool due to a large-enough amount of non-redacted outcomes data.
4. **We did not test the robustness of synthetic control results to methods other than the generalized synthetic control method.** We prespecified that we might compare the results from `gsynth` to the results from other synthetic control methods: the augmented synthetic control method (Ben-Michael, Feller, and Rothstein, 2018) and the original synthetic control method (Abadie and Gardeazabal, 2003). We have not yet conducted these comparisons.

5 Results

5.1 Experimental impact analysis

This section begins with the results from the primary specification of the analysis: estimating the impact of SOAR by regressing 2019–2020 FAFSA completion on an indicator for treatment and the blocking variables only (Equation 1). The primary specification is estimated at the AMP level. Exhibit 18 shows for each experimental grantee the number of AMPs and the individual sample size (i.e., number of eligible youth) in both the treatment and control groups. The number of individuals in the sample varies given the outcome year due to people moving in

---

[37] In particular, depending on the degree of missingness for those covariates in a particular donor pool, the model with covariates would sometimes produce estimates and standard errors identical to the model without covariates.

and out of their units and individuals aging in and out of the eligible age range. The exhibit shows that while the individual sample size is fairly large on its face ($n = 3,834$ age-eligible youth for the primary outcome), the sample size of AMPs is much smaller ($n = 78$). Exhibit 19 shows the raw outcome counts and percentages.

Exhibit 18: Experimental analysis: sample demographics. Demographics correspond to the analytic sample for eligible for the 2019–2020 FAFSA cycle.

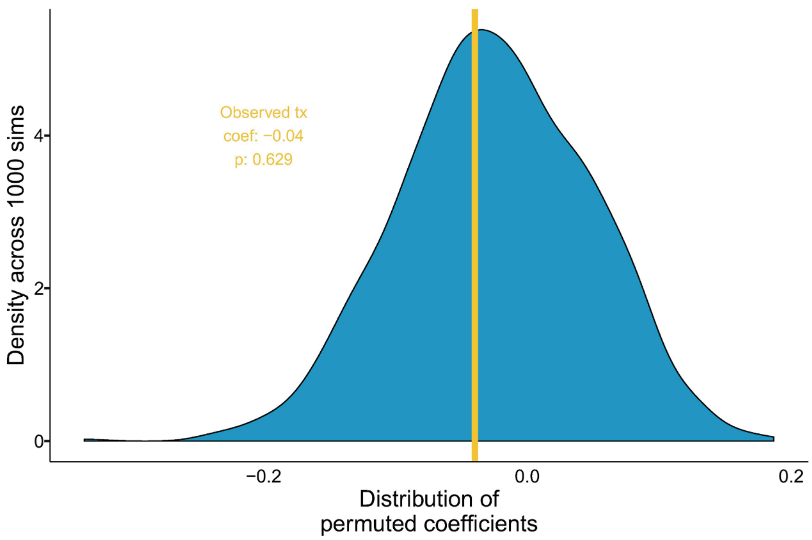
Group	PHA	Sample			Demographics		
		Count AMPs	Eligible for 18-19 cycle	Eligible for 19–20 cycle (primary outcome)	% Black	% Hisp.	Median HH income
Treatment	LA	4	627	687	14	83	17,670
Control	LA	4	490	510	14	83	18,023
Treatment	Chicago	11	529	548	95	4	12,229
Control	Chicago	15	454	460	94	5	12,896
Treatment	Philly	10	370	387	93	5	13,777
Control	Philly	17	639	675	94	4	16,201
Treatment	Seattle	9	360	369	75	4	28,650
Control	Seattle	8	215	198	65	9	26,880
Overall		78	3,684	3,834	66	30	16,911

Exhibit 19: Raw counts and percentages for outcomes. All refer to the 2019–2020 cycle and show counts and percentages before reweighting by the inverse probability of randomization. For the FAFSA counts, the denominator is residents aged 17–20; for the Pell and college counts, the denominator is shifted forward a year to residents aged 18–21.

Group	PHA	Submit FAFSA (2019)			Enroll college (2019)		
		N eligible	Count	Percentages	N eligible	Count	Percentages
Treatment	LA	687	226	32.9%	627	234	37.4%
Control	LA	510	181	35.6%	490	194	39.5%
Treatment	Chicago	548	181	33.0%	529	166	31.4%
Control	Chicago	460	160	34.7%	454	150	33.1%
Treatment	Philly	387	83	21.5%	370	87	23.6%
Control	Philly	675	171	25.3%	639	203	31.8%
Treatment	Seattle	369	108	29.4%	360	135	37.6%
Control	Seattle	198	57	28.9%	215	72	33.4%
Treatment	All non-experimental	1,091	265	24.3%	1,179	268	22.7%

Based on the main method of inference (randomization inference), Project SOAR had no effect on FAFSA completion for the 2019–2020 school year. Exhibit 20 shows the observed treatment coefficient (at the vertical yellow line) relative to the distribution of $m = 1,000$ permuted treatment coefficients. This is a basic graphical representation of a p value. A significant positive effect would be one where the treatment effect was at the right tail of the distribution—that is, where only a small fraction of the distribution of permuted coefficients were to the right of the observed estimate.[38] As the graph shows, the estimated effect is near the center of the distribution, indicating that a difference of the observed magnitude is not unusual given random chance. More specifically, the point estimate shows that treatment AMPs had an average FAFSA completion rate 4 percentage points lower than control AMPs, but the difference is not statistically significant ($p = 0.63$).

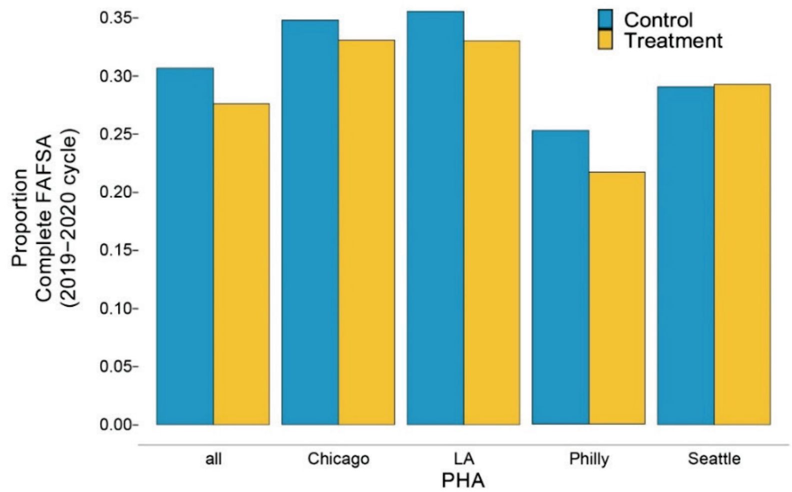
Exhibit 20: Randomization Inference results for FAFSA Completion 2019–2020 cycle (blocking variables only): Distribution of permuted treatment coefficients. The figure shows that the observed treatment coefficient is neither larger nor smaller than the majority of permuted treatment coefficients, indicating a null effect.



A descriptive examination of the rates of FAFSA completion by PHA shows that, in most instances, the control group completion rates were higher than those in the treatment group. Exhibit 21 shows the proportions of FAFSA completion by PHA, accounting for the probability of AMPs being selected into treatment, but without additional adjustment for blocking variables or other covariates.

[38] Conversely, if the treatment had a significant negative effect, the treatment effect would be at the left tail of the distribution and only a small fraction of the estimates would be to the left of the observed estimate.

Exhibit 21: Descriptive rates of FAFSA Completion (2019–2020 cycle). The proportions are reweighted by the IPT weights, but otherwise do not adjust for covariates. The figure shows that in all PHAs except for Seattle, there were slightly lower completion rates among treatment group members, with large amounts of variation in the raw levels between PHAs.



Results from the secondary method of inference support the finding of a null effect from the main specification. Exhibit 22 shows the uncertainty around the estimates from the randomization inference procedure, and Exhibit 23 shows the uncertainty around the estimates from robust standard errors calculated from parametric estimation. In both cases, the graphs display a high degree of uncertainty around the treatment level (calculated as the control mean plus the treatment effect) and support the conclusion that the difference in the two levels cannot be differentiated from zero. Exhibit 24 shows the same general results by looking at the predicted values and confidence intervals separated by each PHA. Again, the story is the same: there are very small differences between the treatment and control groups relative to the size of the uncertainty of the estimates.

Exhibit 22: Randomization Inference results for FAFSA Completion 2019-2020 cycle (blocking variables only): Proportions. Shows observed control mean and observed treatment mean. 95% confidence intervals on control mean are based standard error of mean; 95% confidence interval (CI) on treatment mean are from adding the control mean to the 2.5th and 97.5th percentile of distribution of permuted treatment coefficients from randomization inference.

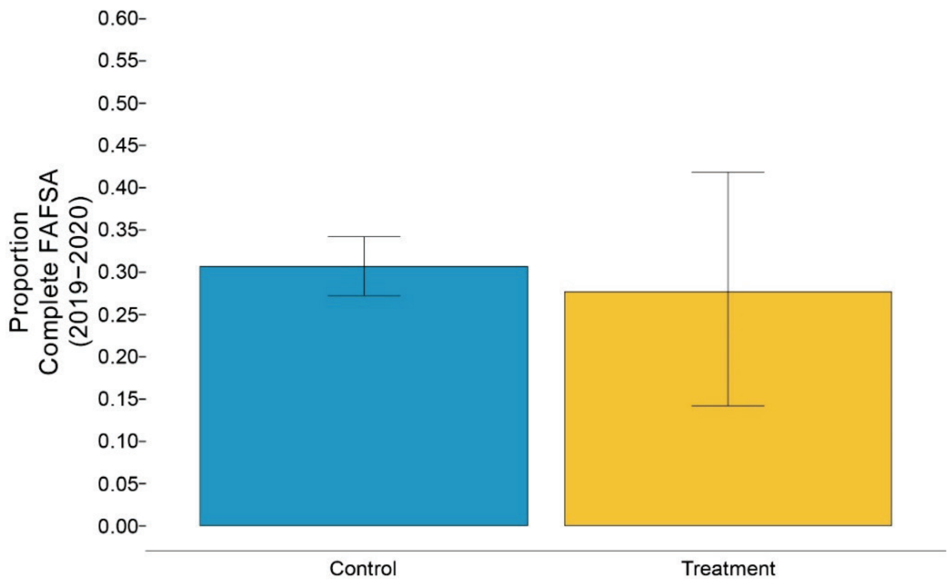


Exhibit 23: Parametric results for FAFSA Completion 2019–2020 cycle (blocking variables only): Combined across PHAs. Shows observed control mean and for treatment, the control mean plus the treatment coefficient. 95% CI on control mean are based standard error of mean; 95% CI on treatment mean are based on the standard error of the mean and the standard error of the estimated treatment effect.

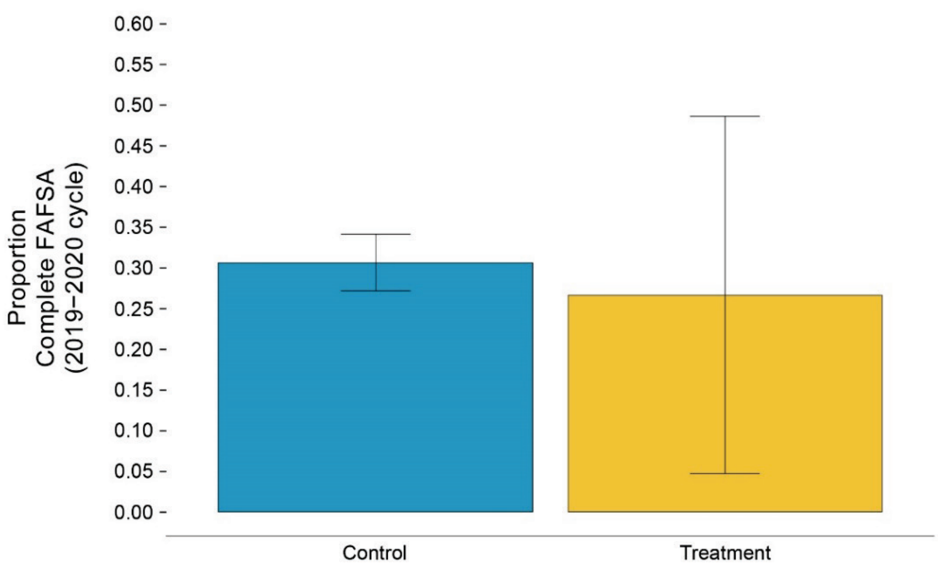
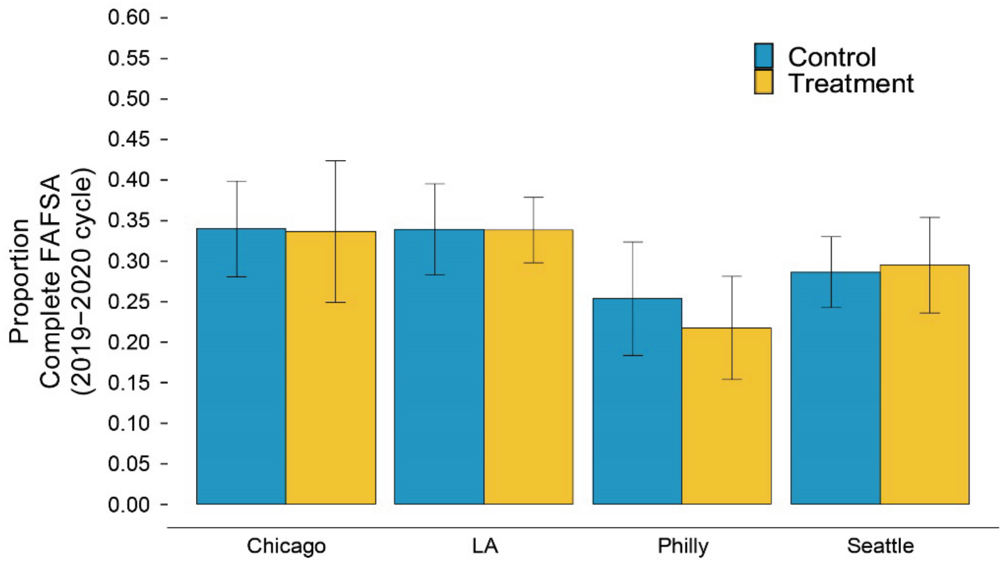


Exhibit 24: Parametric results for FAFSA Completion 2019–2020 cycle (blocking variables only): Separate by PHA. Shows predicted values and 95% CI for each PHA. Other covariates in the model are set to their PHA-specific means.



The secondary model specifications adjusting for covariates also produce insignificant estimates; however, the point estimates are much larger. Exhibit 25 shows the observed positive treatment coefficient on 2019 to 2020 FAFSA completion relative to the $m = 1000$ permuted treatment coefficients.[39] The estimated effect of SOAR when adjusting for additional covariates is an increase in the rate of FAFSA completion of 62.2 percentage points, but the result is not statistically significant ($p = 0.445$) given the very high degree of uncertainty in the estimates.

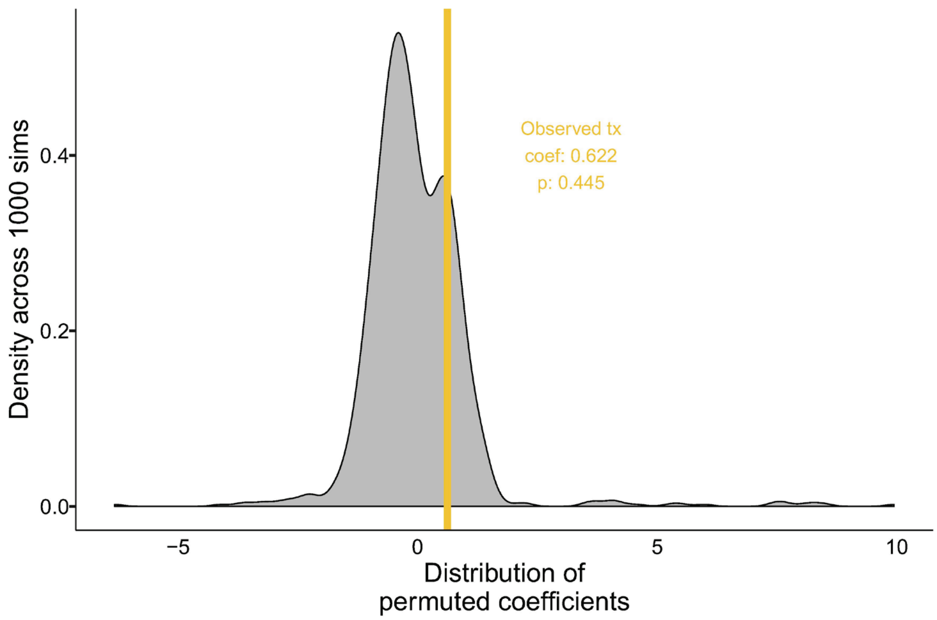
The high degree of uncertainty likely stems from the small sample size and the homogeneous nature of AMPs. Because there are so few AMPs, introducing covariates in effect creates a problem of very small cells available for comparison (for example, think of comparing treatment and control AMPs within Milwaukee with a high proportion of Hispanic, non-White residents to each other). Given this high degree of variability, it is just as likely that even if we were to randomly pick 78 AMPs, we would be about as likely to see a similar difference in FAFSA completion—or even a difference of *negative* 62 percentage points—even though we know those AMPs are not the ones being offered SOAR. In other words, while the point estimate seems large, it should not be taken as evidence that SOAR is highly effective. The variability in these results is important to keep in mind when interpreting the results of the complier analysis described below.

The results for the secondary outcomes of interest are consistent with the FAFSA result. SOAR had no significant impact on any of the secondary outcomes. Exhibit 49 in Appendix Section 7.9

[39] Observant readers may notice that the X-axis extends well beyond the limits of what is a possible effect size due to the limitations of using linear probability models. In short, OLS models with outcomes bounded by zero and one may produce estimates outside of those bounds.

shows the point estimates, parametric standard errors, and randomization inference-based and parametric p values for the secondary specification of the impact on 2019–2020 FAFSA completion (Equation 2) and the college enrollment and Pell receipt outcomes. In all cases, estimated effects are insignificant, which is perhaps not surprising given that without a positive effect on FAFSA completion, it would be unexpected to see positive effects on college enrollment and extremely surprising to see a positive effect on Pell Grant receipt, which is dependent on FAFSA completion.

Exhibit 25: Randomization Inference results for FAFSA Completion 2019–2020 cycle (blocking + additional covariate adjustment): Distribution of permuted treatment coefficients. The figure shows that the observed treatment coefficient is larger than the majority of permuted coefficients, but that there is a high degree of instability in these coefficients and that the difference is not statistically significant.



5.2 Experimental impact analysis: complier effect

This section describes the analysis on students who engaged in at least one in-person interaction with a navigator. Exhibit 26 shows the proportion of compliers by PHA and treatment status, with compliance defined as at least one in-person meeting between a navigator and a resident. The figure shows wide variation in the proportion of treatment group residents who met with a navigator at each PHA, ranging from about 15 percent (Los Angeles) to over 60 percent (Philadelphia). In addition, in Seattle, the different groupings of AMPs into treatment and control groups—one where students in the same AMP could be assigned to different treatment groups depending on which school they were zoned to attend—might have resulted in more meetings with students outside the analytic sample.

Exhibit 26: Rates of compliance by PHA and group assignment. The denominator of each is the number of age-eligible and residentially-eligible students in the group. The numerator of each is the number of these eligible residents who met at least once in person with a navigator.

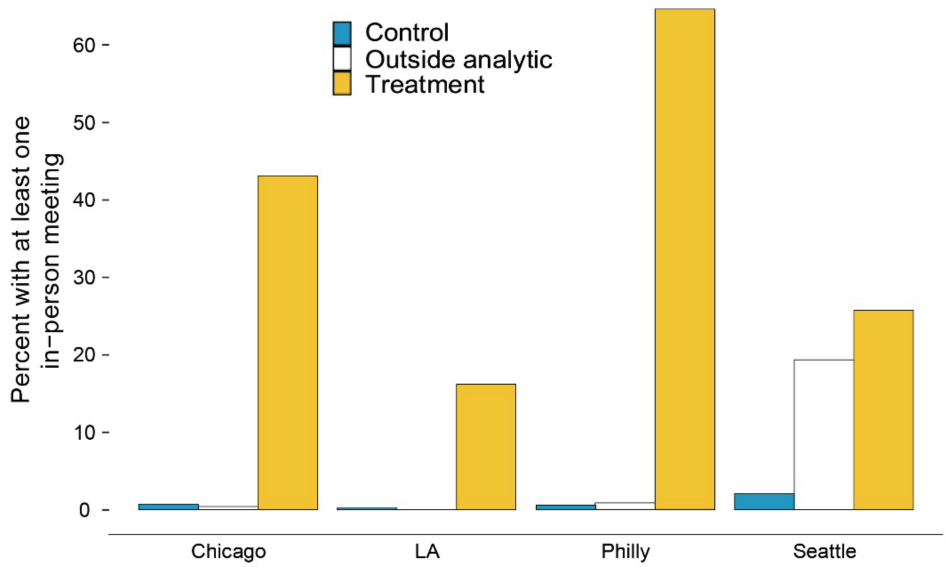


Exhibit 27 shows the results of the complier analysis for the main outcome, restricted to students who:

- Could be matched to the interaction tracker data, and
- Having been matched, were coded to either:

1. `1 == complier`: student was matched to the navigator's participant tracker AND the student had at least one in-person meeting with a navigator.
2. `0 == non-complier`: student was matched to the navigator's participant tracker and the student had no in-person meetings with a navigator—they may have had no meetings at all or meetings that were not in-person.

The exhibit shows that those who met with a navigator were more likely to complete FAFSA, but the results are not significant at the $p < 0.05$ level. The result deserves a closer look because the point estimate is large to the point of not being possible. The interpretation of the coefficient is that the average estimated effect was a 111 percentage point increase in the rate of FAFSA completion. This out-of-bounds estimate is a result of using a linear probability model that estimates Ordinary Least Squares using a binary outcome bounded by zero and one. The linear nature of the estimator allows for estimates to lie outside of the zero to one range. That it does exceed one is somewhat expected, given that the covariate-adjusted ITT effect was 62 percentage points and the take-up rate is low. The complier estimator is, in effect, rescaling the ITT estimate based in part on the engagement rate, so it is not surprising to see the estimate approximately double when the engagement rate is about 50 percent.[40] In this case, we recommend interpreting the estimate more as a sign of the high degree of uncertainty in the estimates and less as a signal of the magnitude of SOAR's impact. In other words, the estimate is not significant, even though it is very large. Appendix Section 7.8 presents a robustness check with a different definition of compliance, which also results in an insignificant estimate.

Exhibit 27: Effect of treatment on compliers: main definition of compliance. The exhibit shows results from the two stage least squares method of analyzing complier effects discussed in Section 4.3.

	Dependent variable: **FAFSA Complete (2019–2020)**
Treatment (complier instrument)	1.114 (1.011) p = 0.271
Male	–0.033 (0.093) p = 0.719
Hispanic (ref: Black)	–0.032 (0.087) p = 0.713
Other (ref: Black)	–0.154 (0.113) p = 0.175
Total annual income	0.00000 (0.00000) p = 0.053∗
Chicago (ref: LA)	0.012 (0.129) p = 0.927
Philly (ref: LA)	–0.417 (0.210) p = 0.048∗∗
Seattle (ref: LA)	0.292 (0.418) p = 0.485
Total HH members	–0.038 (0.026) p = 0.144
Constant	–0.310 (0.760) p = 0.684

[40] When conducting analyses of the impact on compliers using the two-stage least squares method discussed in Section 4.3, researchers emphasize the importance of using a linear probability model at the first stage, since the properties of the 2SLS when using a nonlinear model at the first stage are poorly understood.

Observations	925
Residual Std. Error	0.591 (df = 915)

Note: $*$p<0.1; $**$p<0.05; $***$p<0.01

5.3 Non-experimental impact analysis: Synthetic Control Method

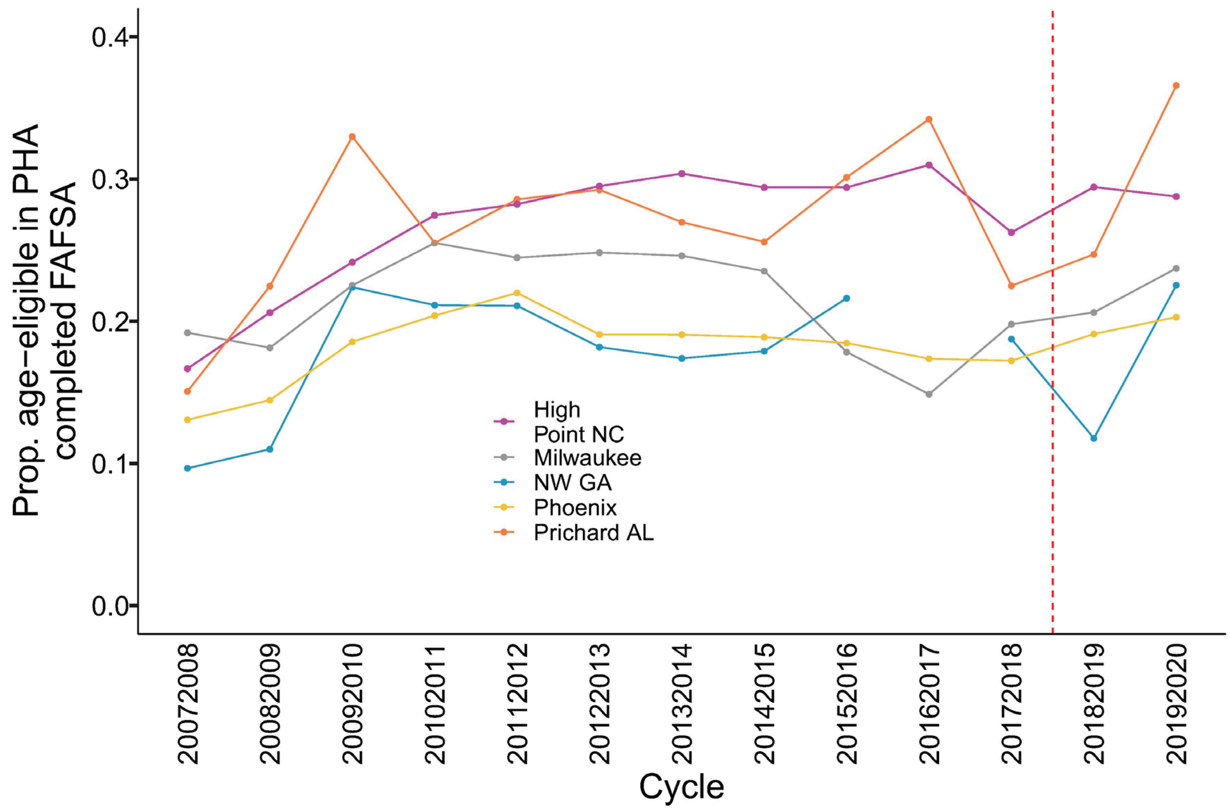
This analysis examines the results for the five non-experimental PHAs that made navigators available to all residents aged 15–20, rather than restricting eligibility to students living in certain AMPs.[41]

Exhibit 28 shows trends over time in FAFSA completion in the five non-experimental treatment PHAs.[42] The figure shows the trends in FAFSA completion for each PHA going back to the 2007–2008 FAFSA cycle and extending to the two cycles after the start of Project SOAR, which is demarcated by the vertical, dotted line. A visual analysis suggests some upward movement in the final year of analysis (the 2019–2020 FAFSA cycle), but it also shows some year-to-year variation in FAFSA completion rates over time at some PHAs.

[41] Appendix Section 7.10 discusses the construction of the synthetic control donor pool and how the Department of Education's policy of redacting data for cells with fewer than ten observations affect which PHAs are in the "donor pool" to serve as the comparator group.

[42] For the remainder of this section, we refer to these as the treatment PHAs, except where necessary to distinguish between the four experimental treatment PHAs and five non-experimental treatment PHAs.

Exhibit 28: Trends in FAFSA completion: non-experimental treatment PHAs. Each dot on the graph represents the completion rate among those who met the eligibility criteria (age-eligible and a resident of the PHA at some point during the cycle). The missing dot in the 2016–2017 cycle for NW GA is due to the redaction of < 10 cell count.



While the PHAs did not have an especially high number of eligible students (for example, NW GA, the smallest PHA in the study, had approximately 89 eligible residents at the time of the grant awards), neither were they so small that one would expect large swings in the rates of FAFSA completion from 1 year to the next. Despite this, both Prichard and Northwest Georgia see swings of about 10 percentage points between years. Even Milwaukee, with several hundred eligible youth, saw a drop in the rate of FAFSA completion by nearly 10 percentage points between the 2014–2015 and 2016–2017 FAFSA cycles.[43] Overall, the year-to-year fluctuations can make it more difficult to model a precise synthetic control unit as a comparison.

The result of the synthetic control method analysis is a positive but statistically insignificant effect on FAFSA completion in the 2019–2020 cycle. Exhibit 29 shows results for the main synthetic control specification, with Exhibit 30 listing the estimates and p-values. For the figures, each bar represents the average treatment effect on the treated (ATT) and 95 percent

[43] Exhibit 51 decomposes the between-year variation into changes in the number of age-eligible youth residing in the PHA (denominator for completion rates) and changes in the number of students who submit the FAFSA (numerator). The exhibit shows both changes in the number of eligible residents and changes in the number of those completing the FAFSA, suggesting some amount of fluidity on both margins.

confidence intervals. The blue bars show the 11 pre-treatment FAFSA cycles. If the match works well, these bars should be centered around zero, preferably with small error bars, which would indicate that the procedure is both closely matching the pre-treatment trends of the grantees and also doing so with little uncertainty. The yellow bars show the two post-treatment FAFSA cycles where navigators had either started providing services (2018–2019 cycle) or were engaged in providing more robust services (2019–2020 cycle).

The main specification includes PHAs as donors even if they have some redacted FAFSA counts; they are included as donors for the years in which the counts are not redacted. That specification shows a modest, positive estimated increase in FAFSA completion of approximately 3 percentage points for the 2019–2020 FAFSA cycle, but the result is not statistically significant ($p = 0.28$).

The secondary specification includes PHAs as donors only if they have no redacted FAFSA counts. In addition, since one of the treatment PHAs had redaction (Northwest Georgia), it is excluded from the analysis. The results, presented in Exhibit 31, show a similar modestly positive estimate of an increase in the FAFSA completion rate of about 4 percentage points for the focal post-treatment year, and the result is borderline significant ($p = 0.08$).

Exhibit 29: Synthetic control treatment effect on FAFSA completion by year. The results are positive in the focal treatment year (2019–2020 FAFSA cycle) but not statistically significant ($p = 0.28$).

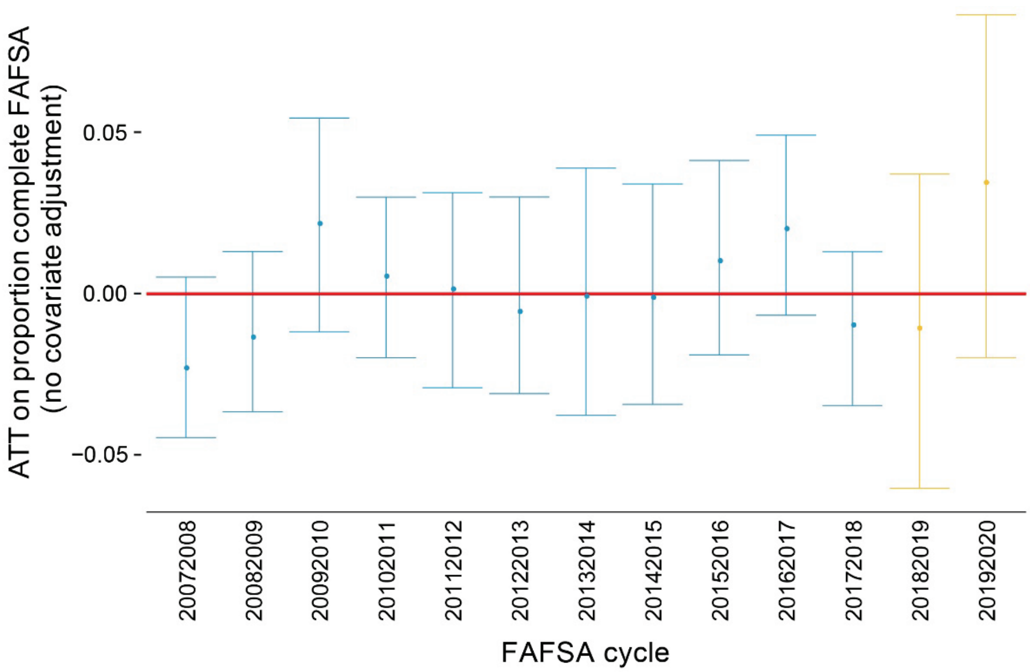
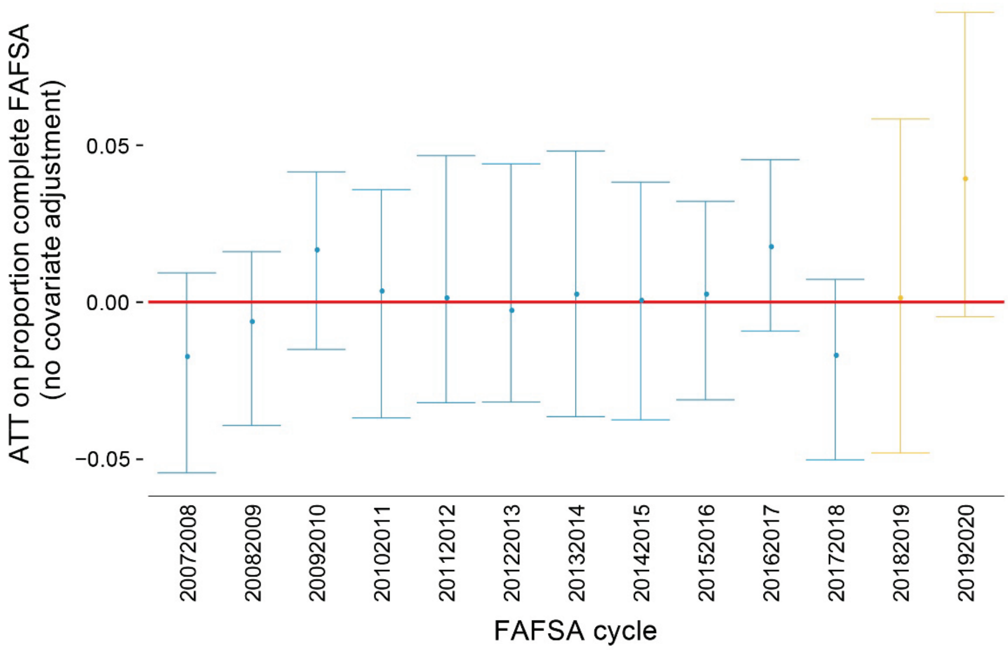


Exhibit 30: Synthetic control results. The *PHA-year* specification corresponds to the specification where we retain PHAs in the donor pool as long as they have at least one non-redacted year of FAFSA data, with the PHA only serving as a donor during the observed years (Exhibit 29). The *PHA* specification corresponds to the specification that removes all PHAs with *any* redacted FAFSA cycles (Exhibit 31). The highlighted year was the main pre-registered primary outcome of interest.

Cycle	Spec.	ATT	CI.lower	CI.upper	p.value
Pre-treatment years					
2007–2008	PHA-year	-0.02	-0.04	0.01	0.09
2007–2008	PHA	-0.02	-0.05	0.01	0.31
2008–2009	PHA-year	-0.01	-0.04	0.01	0.27
2008–2009	PHA	-0.01	-0.04	0.02	0.51
2009–2010	PHA-year	0.02	-0.01	0.05	0.25
2009–2010	PHA	0.02	-0.02	0.04	0.35
2010–2011	PHA-year	0.01	-0.02	0.03	0.72
2010–2011	PHA	0.00	-0.04	0.04	0.86
2011–2012	PHA-year	0.00	-0.03	0.03	0.93
2011–2012	PHA	0.00	-0.03	0.05	0.77
2012–2013	PHA-year	-0.01	-0.03	0.03	0.92
2012–2013	PHA	-0.00	-0.03	0.04	0.82
2013–2014	PHA-year	-0.00	-0.04	0.04	0.90
2013–2014	PHA	0.00	-0.04	0.05	0.87
2014–2015	PHA-year	-0.00	-0.03	0.03	0.93
2014–2015	PHA	0.00	-0.04	0.04	0.92
2015–2016	PHA-year	0.01	-0.02	0.04	0.60
2015–2016	PHA	0.00	-0.03	0.03	0.88
2016–2017	PHA-year	0.02	-0.01	0.05	0.18
2016–2017	PHA	0.02	-0.01	0.05	0.19
2017–2018	PHA-year	-0.01	-0.03	0.01	0.47
2017–2018	PHA	-0.02	-0.05	0.01	0.16
Post-treatment years					
2018–2019	PHA-year	-0.01	-0.06	0.04	0.60
2018–2019	PHA	0.00	-0.05	0.06	0.93
2019–2020	PHA-year	0.03	-0.02	0.09	0.28
2019–2020	PHA	0.04	-0.00	0.09	0.08

Appendix Section 7.11 shows additional results for the secondary outcomes of Pell receipt and college enrollment, which all show null estimated effects.

Exhibit 31: Synthetic control treatment effect on FAFSA completion by year: removal of PHAs with any redaction. The figure shows the ATT for each of the treatment years. The result is positive and borderline significant ($p = 0.08$). In addition to changes in the donor pool for this model (PHAs with no redaction in FAFSA completion), the model also excludes the one treatment PHA (NW GA) with a redacted FAFSA cycle.



While the estimated results for FAFSA completion are not strictly significant, it is possible there is suggestive evidence of movement in the positive direction. Both specifications show estimated treatment effects for the focal post-treatment year that are fairly large relative to the estimates in the pre-treatment years (for example, they are larger than any of the predicted differences in the pre-treatment years), providing some evidence that we are not likely to see estimates of a similar size entirely by chance, but the confidence intervals given the estimation procedure are too wide for us to consider the result statistically significant. Even if we assume there is a modest increase in FAFSA completion, there is no indication of any positive effect (modest or otherwise) on college enrollment or Pell Grant receipt. In sum, the evidence from the non-experimental impact analysis suggests that SOAR did not achieve its primary goals.

6 Discussion

Given the results, we cannot say with confidence that Project SOAR improved FAFSA completion or other postsecondary educational outcomes. Past intensive interventions like the Accelerated Study in Associate Programs (Scrivener et al. 2015; Miller et al., 2020), One Million Degrees (Bertrand et al., 2019), and the New Hampshire Scholars program (Carrell and Sacerdote, 2017) have shown effects of greater than 6 percentage points improvement in FAFSA completion. There is no signal of similarly sized impacts from SOAR. Even though the experimental estimates were estimated imprecisely due to the limited number of AMPs used for randomization, the results as a whole suggest that SOAR was less effective than those interventions among the eligible SOAR population as a whole.

6.1 Limitations

There are a few limitations and features of the study to keep in mind when interpreting the results.

Eligibility criteria:

The decision to adopt very broad eligibility rules based only on age made it more difficult to understand how effective Project SOAR might be among college-interested individuals in need of assistance. In particular, two sets of groups could attenuate the estimated impact. The first group is made up of 15- to 20-year-olds who are not considering college and are unlikely to reconsider their plans even if contacted by a navigator. The second group is made up of 15- to 20-year-olds who would attend college without additional assistance or are already attending. Neither of these groups is likely to be helped by navigators, but both are included under the broad eligibility criteria.

Answering a question about the effectiveness of SOAR among all 15- to 20-year-olds (or 17- to 20-year-olds) may miss important effects on the subset of youth who are most in need of assistance. A more policy-relevant question may be to ask how effective the intervention is with respect to individuals who are on the margin of applying to college (perhaps identified along the lines of [Carrell and Sacerdote, 2017]), yet any such targeting of services would depend on PHAs collecting additional educational information about residents. To the extent that many SOAR grantees had to navigate data-sharing agreements with local school districts, there could be a blueprint for how to strengthen ties between PHAs and local schools to identify students who are academically on track for college but who require extra assistance to do so.

Local flexibility:

The main goal of an impact evaluation is to understand if a given program is effective, which assumes to some extent that the program in question can be easily defined. The emphasis on providing local flexibility, while allowing for adaptation and experimentation, makes it difficult to define what exactly Project SOAR is in the context of an analysis that pools across four and five distinct grantees. As the implementation analysis showed, grantees made different decisions as to where to focus limited resources. In many ways, the demonstration was a test of the average effect of nine different college access programs. While it is possible some local models worked well, this evaluation was not designed (or powered) to be able to examine grantees individually. As such, these results do not suggest all efforts were unsuccessful, just as they are not well positioned to identify specific strategies that seem to be more (or less) promising. While some level of variation is beneficial in program design, it also is helpful to define in detail what makes up the core program model. The ASAP program provides one good example of a core program philosophy with some minor local variation (such as providing a small stipend for gas/groceries in Ohio instead of providing a transit pass in New York [Miller et al., 2020]). Structuring future grants to better identify core program features or more directly plan to identify and compare local variation can help to narrow relevant evaluation questions.

Program maturation:

The period of performance for the grant may have had an effect on performance. Grantees may not have been able to achieve program maturity given the 2-year grant period. It took grantees most of the first year to develop inroads with the community. Navigators were busy during this period learning and updating their approaches, and services provided during the first year of the grant were likely of different quality than those provided over the second year of the grant. By the time navigators were building more recognition in the community, grantees were already under pressure to make plans for the grant's termination. In some cases, this may have created staffing problems. Additionally, most grantees observed that residents were reluctant to participate in a program that they thought would disappear in a short while (as has happened with many other new initiatives in the past). Designing demonstrations with longer grant periods may allow for better program development and steady-state operations to emerge. Combining an impact analysis with a formative analysis in the early stages of the grant could help to identify which particular features of the program seem more or less effective.

Resource allocation:

Estimates about the resources needed to assist students effectively may have been wrong. HUD made certain assumptions about the number of eligible students each navigator could reasonably be expected to recruit and assist, assuming a certain level of responsiveness from students. The assumptions may have been based on program models that required less time and energy devoted to recruitment than the PHA context required (for example, interventions focused in high schools do not have to contend with the resource-intensive process of going door-to-door to proactively educate people about the program), which may have made it difficult for navigators to spend time both recruiting new participants and working with students who were already engaged.

Non-compliance:

There likely were some individuals in the control group who meaningfully engaged with navigators. An early site visit to one of the grantees suggested that PHAs were having some trouble finding a clear demarcation between what was a SOAR service and what was a business-as-usual service. For example, they were actively recruiting treatment participants for college visits, but if a control person asked to go and there were empty seats, they included the control student. Most grantees suggested that they drew a clear line around in-person assistance. In theory, the complier analysis could adjust for these interactions, but only if interactions with control group individuals were reliably tracked in the interaction tracker, which we have reason to doubt (Exhibit 26). Early reports suggested that navigators may not always have recorded control group students who were served in the interaction tracker. To the extent that there was more interaction with control group individuals than was measured in the data, the effect would tend to be biased downwards.

Generalizability:

The four experimental PHAs were ones that (1) applied for the grant and (2) had a large enough resident population that randomization was feasible. In turn, these PHAs have features that place limits on the settings to which the results generalize. First, the experimental PHAs are some of the largest in the country, meaning that the results from the experiment generalize best to other large PHAs. Second, the corresponding school districts all have some form of school choice, which could affect whether the most motivated students need an on-site navigator or whether they instead applied to selective high schools with robust college counseling. Finally, the experimental PHAs are located in places where there are other ongoing efforts to promote FAFSA completion at either the school district, city, or state level. For instance, in fall 2019, Illinois passed a "universal FAFSA completion" law that requires students to apply for FAFSA in order to graduate from high school.[44] While this legislation does not impact the students in the study, since it goes into effect for students starting in the 2020–2021 school year, it shows that we are studying the impact of navigators in settings where policymakers are generally interested in leveraging a variety of tools to promote FAFSA completion.

Limitations of the experimental analysis:

The original design was based on a power analysis which included Milwaukee as one of the experimental grantees. The loss of Milwaukee in the experimental component decreased the power of the study to detect significant effects. While we pre-registered that the study was powered to detect a 6 to 7 percentage point change using the randomization inference procedure and approximately a 5 percentage point change using parametric standard errors, the inclusion of PHA dummies in the actual analysis led to a decrease in the effective power. Some of the PHAs had fewer than 10 AMPs randomized to each group, meaning we were powered to detect something larger than a 6 to 7 percentage point change. While some of the most effective interventions have seen effects of 10 percentage points or higher, those interventions generally involved more intensive advising combined with other types of financial and non-financial supports that SOAR did not provide. Given that, we would reasonably expect SOAR to have a smaller effect, which was something we were not well powered to detect.

Limitations of the synthetic control method:

The final limitation applies to the non-experimental analysis, which showed a positive but non-significant effect of SOAR on FAFSA completion. One assumption of the synthetic control method is that units in the donor pool are untreated. Nearly all states in which PHAs are located have programs helping low-income students with FAFSA and other elements of the college

[44] Students are also allowed to submit waivers seeking an exemption from the requirement.

application process; similarly, nearly all school districts where children residing in those PHAs attend have some form of navigator-like assistance. Because we consider the *unique* element of the present intervention to be the physical presence of a navigator at the PHA, rather than assistance provided at one's local school or nearby nonprofits, we do not expect that other PHAs have interventions that share this feature. However, the presence of other interventions, and our inability to (1) understand the complete range of interventions that are present, and (2) exclude PHAs with similar ongoing interventions, means that some PHAs selected to be a part of the synthetic comparison will have a postsecondary initiative which could bias the estimated effect towards zero, assuming such programs increase postsecondary activity.

6.2 Conclusion

Project SOAR was a program that the grantees and communities very much appreciated and thought of as valuable for students who would slip through the cracks of other educational programs. Even so, grantees were unable to build their grants into mature models during the course of the grant's period of performance. Even the grantees that used existing college access program models had to contend with implementation challenges like figuring out how to get navigators permanent meeting space that impeded quick adoption. These ongoing challenges suggest the model was not at a point where it was robust enough for this type of impact evaluation.

Outside of questions of program maturity, the null effects could result from other factors unrelated to the direct effectiveness of navigators at providing assistance to students, including: (1) navigators not assisting a large enough proportion of residents, (2) navigators assisting residents who were unlikely to ever complete the FAFSA or apply to college, and (3) navigators assisting residents who were always going to complete the FAFSA and apply to college. Without knowing more about who is interested in college and who needs help, it is difficult to direct navigators to where their services can make the biggest difference. Even if navigators can identify the group of interested individuals, it would not help with the limitations of an ITT analysis unless there is additional data collected at baseline. For example, something as simple as an application on which students express interest in navigator assistance could refine the target population. Students could express interest and then be randomized, or express interest following randomization and then be randomized to either a higher or lower intensity of outreach. However it is accomplished, narrowing the pool of targeted students can help better align scarce resources with student needs and may mechanically make evaluation easier.

The current body of evidence on college access interventions suggests that SOAR could be a valuable program model. Other intensive, in-person advising efforts have been shown to be effective. Before evaluating any similar efforts in the future, the program model should be more defined, sufficient resources should be in place to support navigators, and more effort should be made to identify and spend the most effort on the students who are both college-interested and in need of additional assistance.

7 Appendix and supplementary materials

7.1 Examples of the different data structures

Exhibit 32 uses simulated data to show the structure of the AMP-level data. We use the AMP-level data for the main confirmatory analysis that measures the causal impact of the navigators on FAFSA completion.

Exhibit 32: Example of data structure for AMP-level data (used for experimental analysis)

PHA	AMP	treat	FAFSA_rate	perc_black
Chicago Housing Authority	1	0.00	0.63	0.35
Chicago Housing Authority	2	0.00	0.46	0.29
Chicago Housing Authority	...	0.00	0.50	0.24
Chicago Housing Authority	4	1.00	0.53	0.23
Philadelphia Housing Authority	1	1.00	0.54	0.35
Philadelphia Housing Authority	2	0.00	0.54	0.27
Philadelphia Housing Authority	...	1.00	0.53	0.11
Philadelphia Housing Authority	4	1.00	0.54	0.34
Housing Authority of the City of Los Angeles	1	0.00	0.58	0.17
Housing Authority of the City of Los Angeles	2	1.00	0.60	0.31
Housing Authority of the City of Los Angeles	...	1.00	0.52	0.20
Housing Authority of the City of Los Angeles	4	1.00	0.49	0.40
Seattle Housing Authority	1	0.00	0.58	0.36
Seattle Housing Authority	2	1.00	0.53	0.27
Seattle Housing Authority	...	0.00	0.62	0.32
Seattle Housing Authority	4	0.00	0.46	0.27

Similarly, Exhibit 33 uses simulated data to show the structure of the individual-level data.[45] We use the individual-level data for the descriptive analysis of whom navigators serve, and to adjust the main causal estimates to measure the effect of the treatment on students that navigators met with (rather than effect of the treatment on all eligible students).

Exhibit 33: Example of data structure for individual-level data (used for descriptive engagement analysis and adjusting causal analysis for the proportion of youth engaged)

PHA	AMP	treat	Youth_id	Met with navigator	FAFSA
Chicago Housing Authority	1	0.00	7120	0	1
Chicago Housing Authority	2	0.00	4263	0	0
Chicago Housing Authority	...	0.00	4517	0	0
Chicago Housing Authority	4	1.00	7001	1	1
Philadelphia Housing Authority	1	1.00	1879	0	0
Philadelphia Housing Authority	2	0.00	4708	0	0
Philadelphia Housing Authority	...	1.00	4667	1	1

[45] For simplicity, we omit covariates.

PHA	AMP	treat	Youth_id	Met with navigator	FAFSA
Philadelphia Housing Authority	4	1.00	9816	0	1
Housing Authority of the City of Los Angeles	1	0.00	4235	0	0
Housing Authority of the City of Los Angeles	2	1.00	5840	0	1
Housing Authority of the City of Los Angeles	...	1.00	6950	0	1
Housing Authority of the City of Los Angeles	4	1.00	1900	1	1
Seattle Housing Authority	1	0.00	5875	0	1
Seattle Housing Authority	2	1.00	4396	1	1
Seattle Housing Authority	...	0.00	8246	0	0
Seattle Housing Authority	4	0.00	6706	0	1

Exhibit 34 uses simulated data to show the structure of the PHA-level data. We use the PHA-level data for the non-experimental analysis of navigator impact. The data show average rates of FAFSA completion over time and are used to create a synthetic control for the five non-experimental grantees.

Exhibit 34: Example of data structure for PHA-level data (used for quasi-experimental analysis of effect of navigators on residents of sites that did not randomize)

PHA	FAFSA_rate	perc_black
City of Phoenix Housing Department	0.53	0.19
High Point Housing Authority	0.52	0.13
Housing Authority of the City of Milwaukee	0.62	0.17
Northwest Georgia Housing Authority	0.54	0.20
Prichard Housing Authority	0.57	0.19

7.2 Additional details: quantitative analysis of implementation

Here, we provide additional details on the quantitative analysis of program implementation that we discuss in Section 3.5. First, the templates for interaction trackers given to navigators included definitions for the following fields:

- Converser: e.g., student versus parent.
- Medium and Mode.
- Topics: e.g., application process or financial literacy.
- Purpose: e.g., outreach and education; follow-up or check-in.
- Obstacles: e.g., financial; documentation.
- Engagement level: uncontacted, engaged, or disengaged.

Due to data reliability, our primary analyses look at two dimensions: first is medium and mode; second is who the converser was. Exhibit 35 shows the specific definitions for these. Within the same definition, Exhibit 36 shows that the navigators' hand-inputting of data led to multiple variations of the same entry—for instance, five different forms of describing one-on-one counseling.

Exhibit 35: Definitions for key interaction tracker elements

Medium/mode	
Texting	A single text message or text message conversation between the navigator and the conversant taking place during the specified day
Phone call	A phone conversation or voicemail between the navigator and the conversant taking place during the specified day
Email	A single email or email exchange between the navigator and the conversant taking place during the specified day
Social Media	A single email or email exchange between the navigator and the conversant taking place during the specified day
Webinar	An attended webinar scheduled for the specified date
One on one/counseling	An in-person, individual interaction involving only one participant (but possibly including guardians, school officials, or others)
Small Group/workshop	An in-person interaction involving a smaller number of participants and allowing for substantial individual participant-navigator communication
Large group/event	A large, in-person interaction involving several participants with limited opportunity for individual participant-navigator communication
Letter	Individualized communication delivered through the postal system
Literature mailing	Delivery of pre-existing literature or materials through the postal system
Converser	
Participant	The FAFSA eligible AMP resident being assisted by the navigator
Parent/Guardian	A primary caregiver for FAFSA eligible AMP resident
Participant and Parent/Guardian	BOTH the primary caregiver AND the FAFSA eligible AMP resident are present for the interaction
School Official w/wo	A teacher, guidance counseled, or similar official related to the FAFSA eligible AMP resident is present for the interaction

Exhibit 36: Variation in interaction trackers' inputted data within the same field. The left panel shows variation within the converser field; the right panel shows variation within the medium/mode field.

Other (PHA Added)
parent
Parent
Parent / Guardian
Parent/Guardian
Parent/Participant
participant
Participant
PARTicipant
Participant and Parent / Guardian
participant and Parent/Guardian
Participant and Parent/Guardian
Participant w/ Parent
Participant/Parent
Particpant
Provider/School w/ participant
School Official
School Official w/wo Participant and Guardian
School Official with Participant and Guardian
School Official without Participant and Guardian Sibling

1:1
Email
Large Group (>15)
Large Group / Event
Large group/event
Large Group/Event
Letter
Literature mailing
Literature Mailing
Meeting
One-on-One/Counseling
One on One / Counseling
One on one counseling
One on one/counseling
Other
Other (PHA Added)
OTher (PHA Added)
Phone
Phone call
Phone Call
Small Group (<=14 or)
Small Group (14 or less)
Small Group / Workshop
Small Group/workshop
Small Group/Workshop
Social Media
Text
Texting
Webinar
Workshop

Exhibit 37: Top words in free-text notes on one-on-one counseling: experimental PHAs

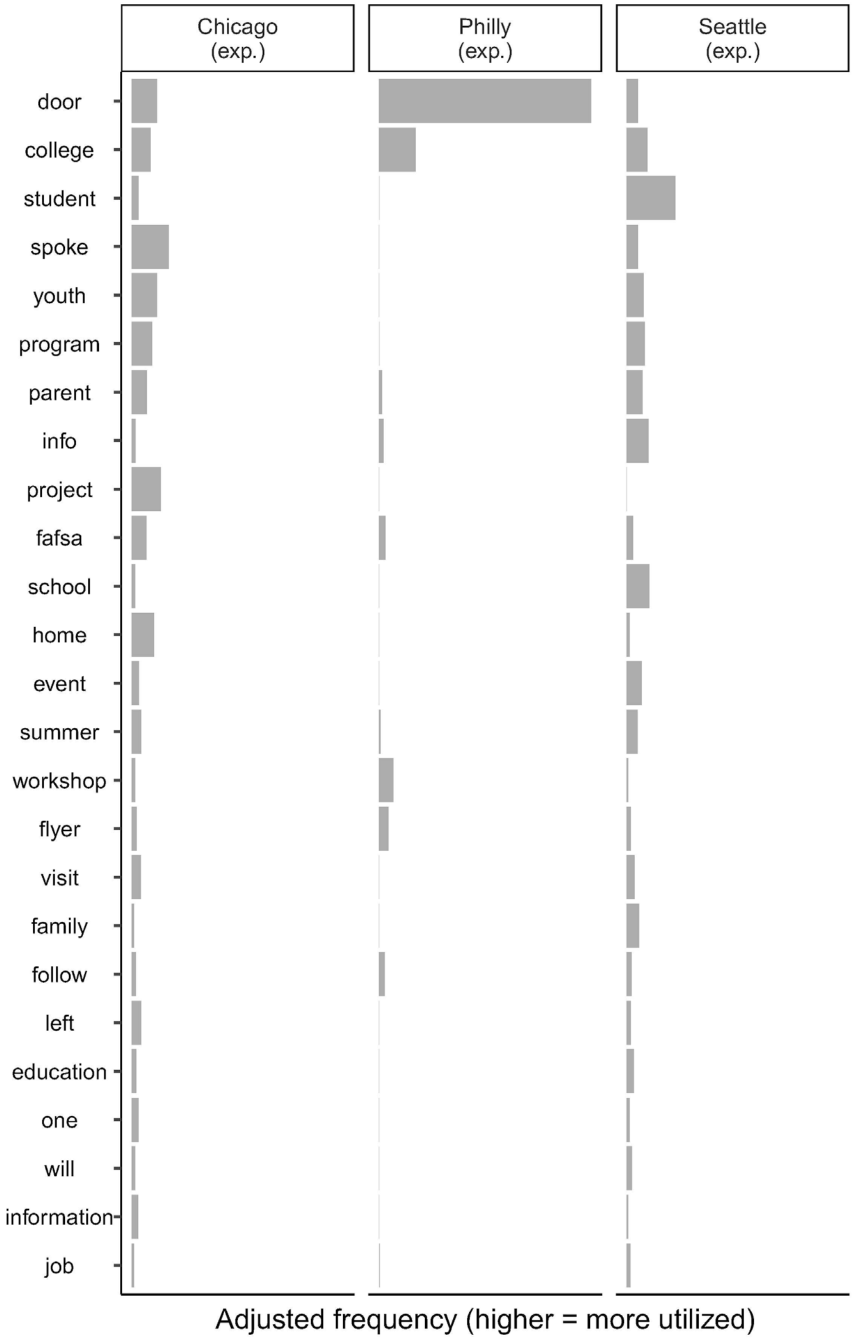


Exhibit 38: Top words in free-text notes on one-on-one counseling: non-experimental PHAs

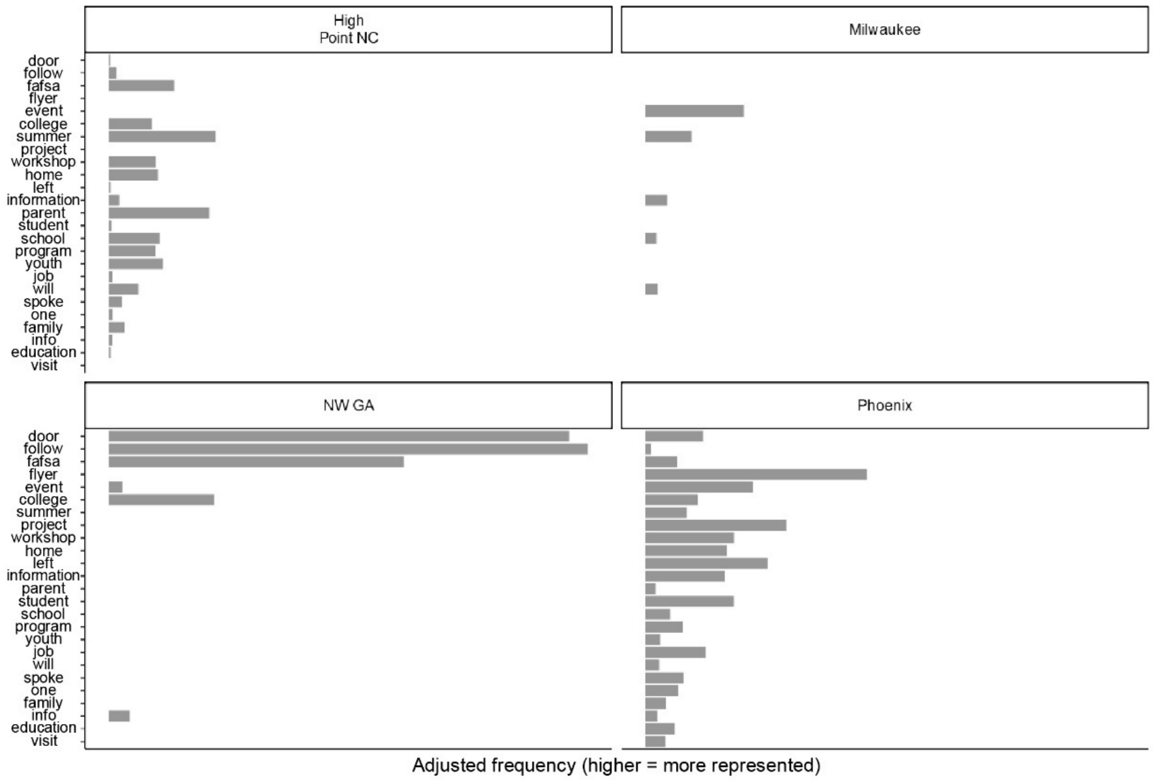
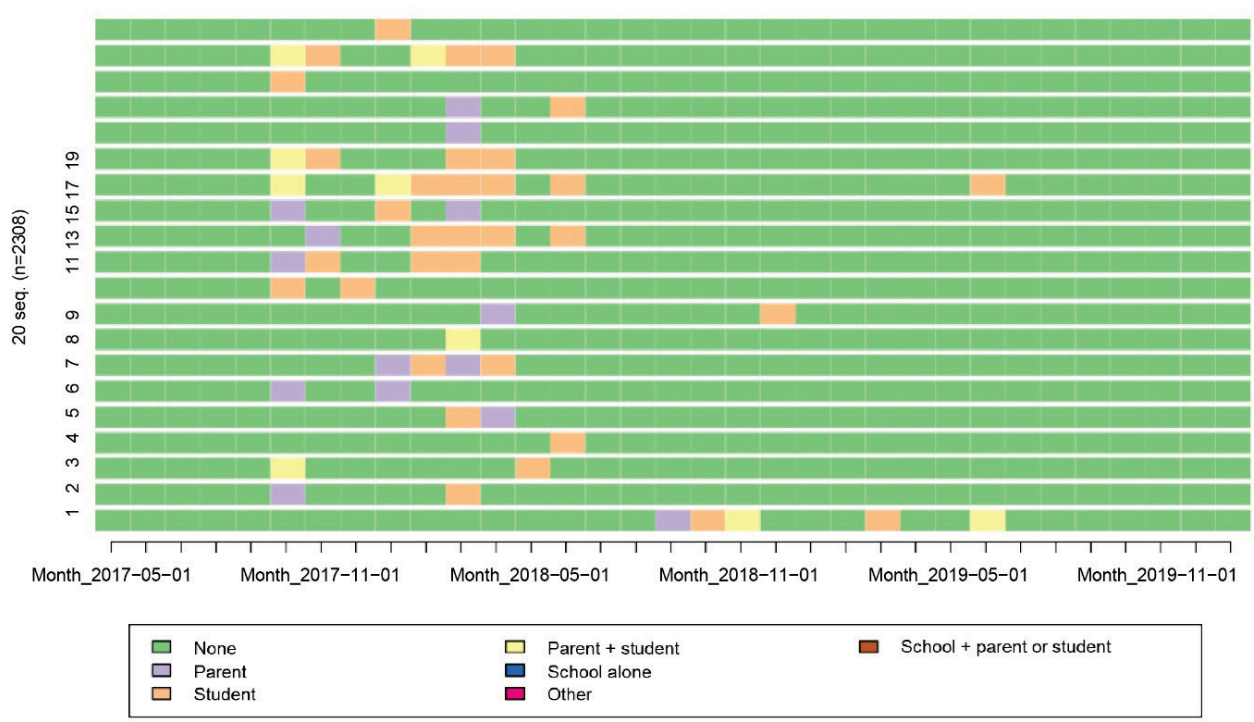


Exhibit 39: Illustration of sequences with 20 randomly-chosen students



7.3 Details on construction of analytic sample

Exhibit 40: Code snippet for defining eligibility to appear in a cycle

```
## create dummy dataframe of start and end dates for different FAFSA cycles
fafsa_cycles <- data.frame(year = years) %>%
         mutate(fafsa_start = ifelse(year <= 2015,
                                     sprintf("%s-01-01", year),
                                     sprintf("%s-10-01", year-1)),
                fafsa_end = sprintf("%s-06-30", year+1),
                fafsa_effective_start = sprintf("%s-09-01", year),
                fafsa_effective_end = sprintf("%s-08-31", year+1),
                fafsa_name = sprintf("FAFSA_Cycle_SY_%s%s",
                                     year,
                                     year+1),
        # oldest people are 20 at start of period
               oldest_dob_fafsa = ymd(fafsa_effective_start) - years(20),

        # youngest people are 17 at end of period
                youngest_dob_fafsa = ymd(fafsa_effective_end) - years(17))

## code to return those eligible within a given cycle
return_elig_cycles <- function(row, cycles_touse = fafsa_cycles,
                               parallelize = TRUE){


  in_cycle <- ifelse(one_dob <= cycles_touse$youngest_dob_fafsa & # not too young
                 one_dob >= cycles_touse$oldest_dob_fafsa & # not too old

    # participation started before the end of the FAFSA period
                 one_start <= cycles_touse$fafsa_effective_end &

    # if participation ended, it was after the start of the period
                 one_end >= cycles_touse$fafsa_effective_start, 1, 0)

  which_cycles <- cycles_touse$fafsa_name[which(in_cycle == 1)]

}
```

7.4 Replicating the random assignment process

The following code can be used to replicate the random assignment process:

```
# this function takes a dataset with PHA and AMP identifiers and completes
##the assignment algorithm with
# a set of predefined parameters for staff size and workload.
##To create permutations, the input list needs to be
# randomly sorted again before the function is run

# Set the list of PHAs and the number of navigators and max workload
for each
state <- c("CA", "IL", "PA", "WA") nav <- c(3, 3, 2, 3)
ml  <- c(150,150,150,82)
pha.params <- cbind.data.frame(state,nav,ml)

# Assume individual takeup rate ("rate") for offered services:
```

```
rate <- 0.5

assignment <-
function(input, pha.params){
    # Create a container dset called "substates" (just initialize to input)
    substates <- input[1, ]
    # Want the dset to be empty, so empty it
    substates <- substates[-1, ]
    # Now we have a dset with all the same variables
    as input, # but empty; we'll fill it
    # up with data once the random sorting and
    allocation # have allowed us to make
    treatment
    # assignments.

    for (s in c("IL", "PA", "CA", "WA")){
    # Deal only with one state at a time:
    st <- input[input$state == s, ]

    # Number of AMPs in PHA s:
    (st_n <- dim(st)[1])

    # Number of navigators in PHA s:
    (nn <- pha.params$nav[pha.params$state == s])

    # Calculate the total load for AMP1 outside the loop:
    st$tot.load[1] <- st$tot.served[1]
    # Assign the first AMP to treatment outside the loop:
    st$treat[1] <- 1

    # Initialize index varible i:
    i <- 2

    # Begin while loop, calculating as long as the total workload is
    less than or # equal to the maximum load per navigator times the
    number of navigators: while (st$tot.load[i-1] <=
    pha.params$ml[pha.params$state==s]*nn){
    # Calculate running total load
    st$tot.load[i] <- sum(st$tot.served[1:i])
    # Assign to treatment as long as the while loop is still going
    st$treat[i] <- 1
    # Increment i
    i <- i + 1
    }
    # Build the "substates" data by stacking finished "st" data
    frames on top of # each other.
    substates <- rbind(substates, st)
    }

   return(substates)
}
```

7.5 Additional results: experimental analysis. Secondary specification for main outcome

Section 5.1 focused on experimental results from our primary specification, which controls only for the PHA dummy and the number of youth in the AMP (Equation 1). The present section focuses on experimental results from the secondary specifications (Equation 2), which controls for both those blocking variables and the following covariates:

- Race/ethnicity: Percent of youth who are Non-Hispanic Black; Hispanic; Non-Hispanic other.
- Percent of youth who are citizens.
- Mean total annual income of the household in which the youth resides.

7.6 Additional results: experimental analysis. Differences in demographics between PHAs and across treatment conditions

The model adjusting for covariates is relevant because (1) AMPs within a PHA differ along youth attributes that past research shows are relevant for FAFSA completion and (2) some of these attributes remain imbalanced within a PHA even after the inverse probability of treatment reweighting discussed in Section 4.3 Exhibit 41 shows that although most of the variation in race/ethnicity is between PHAs (most notably, LA as majority Hispanic, Seattle as a mix of Black and "Other," Chicago and Philadelphia as majority Black), there is remaining within-PHA variation. Similarly, Exhibit 42 shows variation not only between PHAs (e.g., Seattle's income distribution is higher than Chicago's) but also variation at the modified AMP level within a PHA.

Exhibit 41: Race/ethnicity comparison across PHAs and AMPs. Each dot represents one modified AMP used in randomization.

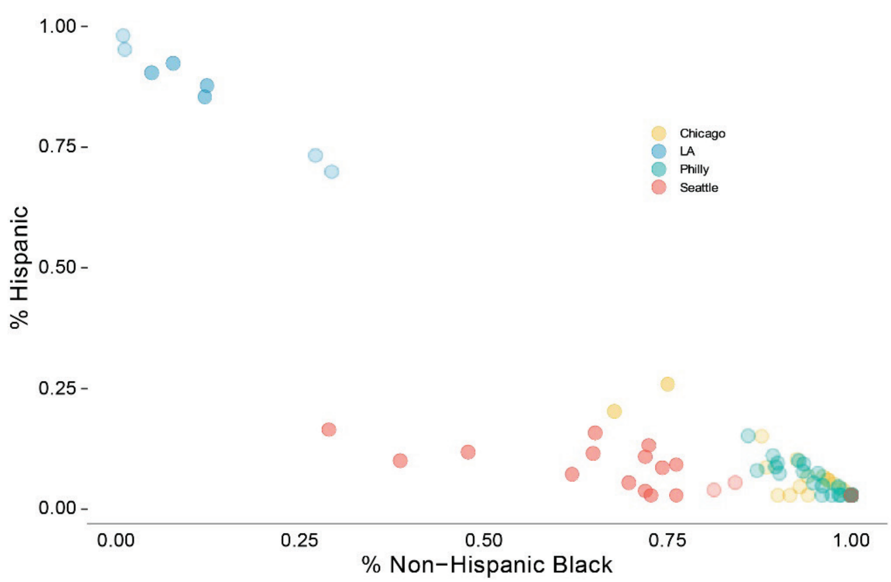
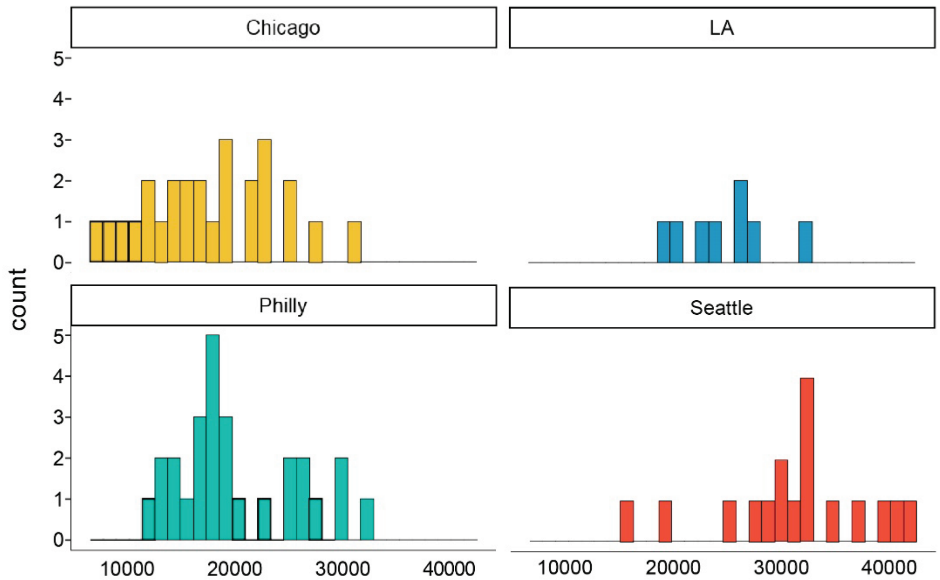


Exhibit 42: Household income comparison across PHAs and AMPs. Each bar represents one modified AMP used in randomization.



Important for our purposes, however, are differences in the demographics between AMPs randomized to treatment and AMPs randomized to control. Exhibit 43 shows the overall balance across PHAs. As expected, the randomization process resulted in the treatment group having larger AMPs, but these differences persist after reweighting for randomization probabilities. The treatment AMPs also have a slightly lower percentage of non-Black, non-Hispanic race/ethnicity.

Since each of the models (Equation 1; Equation 2) controls for a PHA dummy, more important are differences in the treatment versus control group composition within the same PHA. Exhibit 44 shows those differences. LA is the only PHA with a large non-citizen population, that variable remains imbalanced (and higher in the treatment AMPs). Seattle has a higher percent Black demographic in the treatment AMPs and a higher median household income. Overall, the differences (1) motivate the use of a specification that includes additional covariate controls, but (2) show that the combination of a PHA dummy with control variables may lead to highly unstable estimates because some PHAs have very low proportions of certain demographic groups.

Exhibit 43: Differences between treatment and control AMPs: Full sample; raw differences since all variables are aggregated to the AMP level, the figure shows either the mean proportions (binary variables) or mean values (continuous variables) across treatment and control AMPs. All are reweighted by the inverse probability of treatment weighting (IPTW).

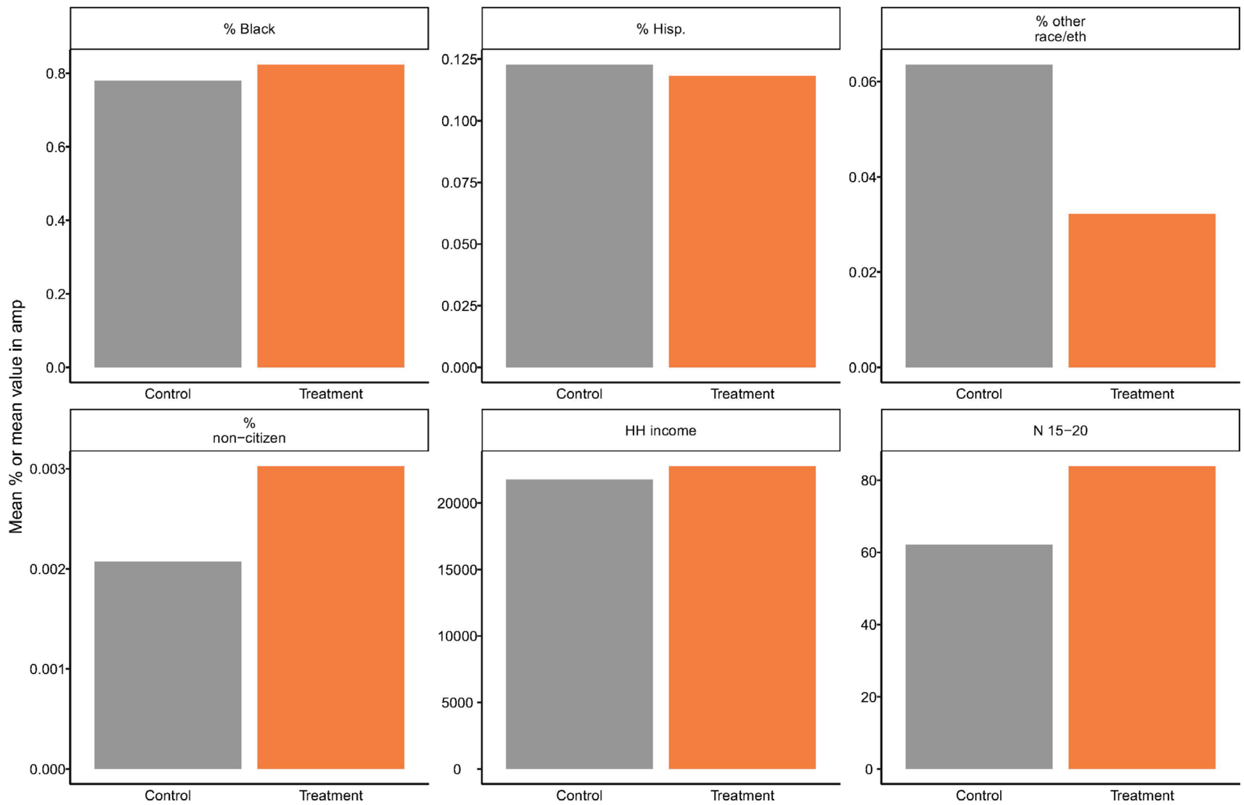
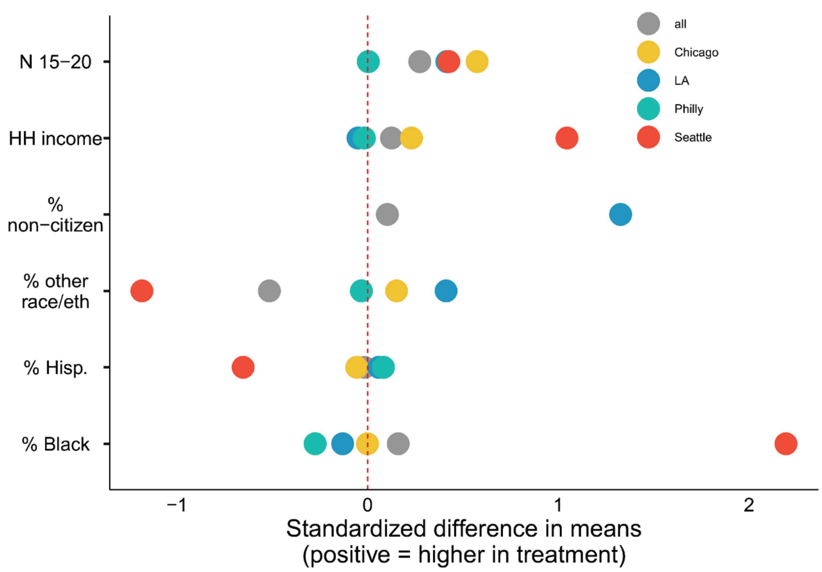


Exhibit 44: Differences between treatment and control AMPs: full sample and by PHA; standardized difference in means. Each dot either represents one PHA or all PHAs (gray). All estimates are reweighted by the IPTW.



7.7 Additional results: experimental analysis. Results from models with additional covariate adjustment

Exhibit 25 in the main text shows the observed positive treatment coefficient on 2019 to 2020 FAFSA completion relative to the m = 1000 permuted treatment coefficients. Here, Exhibit 45 shows the rates based on constructing confidence intervals using those permuted treatment coefficients, which are wide due to the variability in the estimates depending on the treatment permutation. Exhibit 46 shows the rates based on parametric standard errors across all PHAs, and Exhibit 47 shows the rates based on parametric standard errors specific to each PHA. Overall, the results show that, after adjusting for covariates associated with FAFSA completion and that remain imbalanced across treatment groups, the treatment group completed the FAFSA at higher rates. but the differences are not statistically significant and are highly uncertain due to minimal within-PHA variation in some demographic traits.

Exhibit 45: Randomization Inference results for FAFSA Completion 2019–2020 cycle (blocking + additional covariate adjustment): proportions. Shows observed control mean and observed treatment mean. 95% confidence intervals on control mean are based standard error of mean; 95% CI on treatment mean are from adding the control mean to the 2.5th and 97.5th percentile of distribution of permuted treatment coefficients from randomization inference.

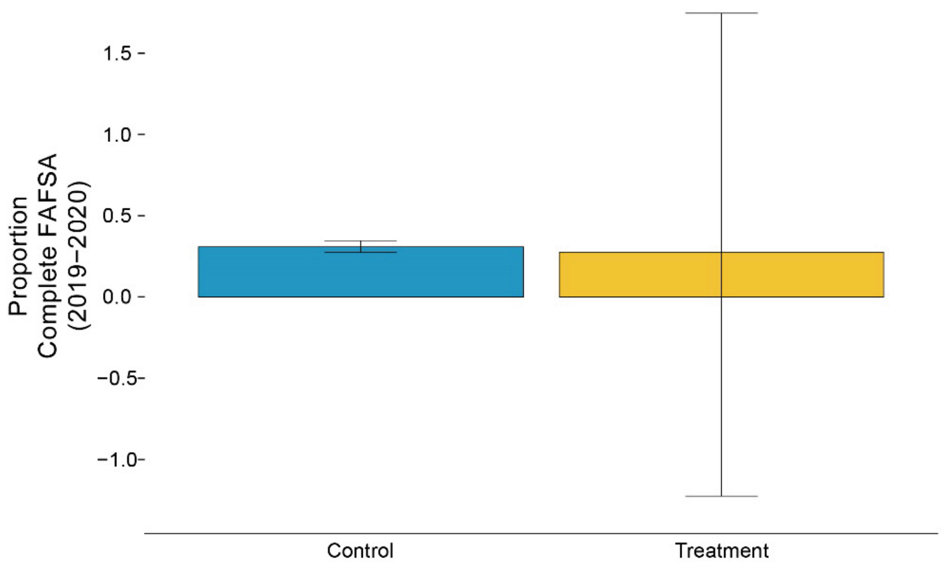


Exhibit 46: Parametric results for FAFSA Completion 2019–2020 cycle (blocking + additional covariate adjustment): combined across PHAs. Shows observed control mean and for treatment, the control mean plus the treatment coefficient. 95% CI on control mean are based standard error of mean; 95% CI on treatment mean are $var_t + var_{int} - 2 * covar(t, int)$.

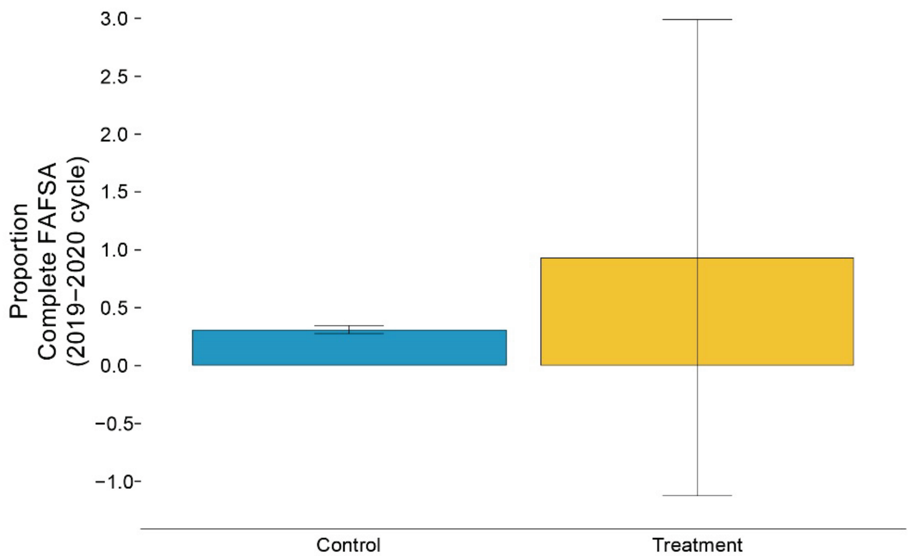
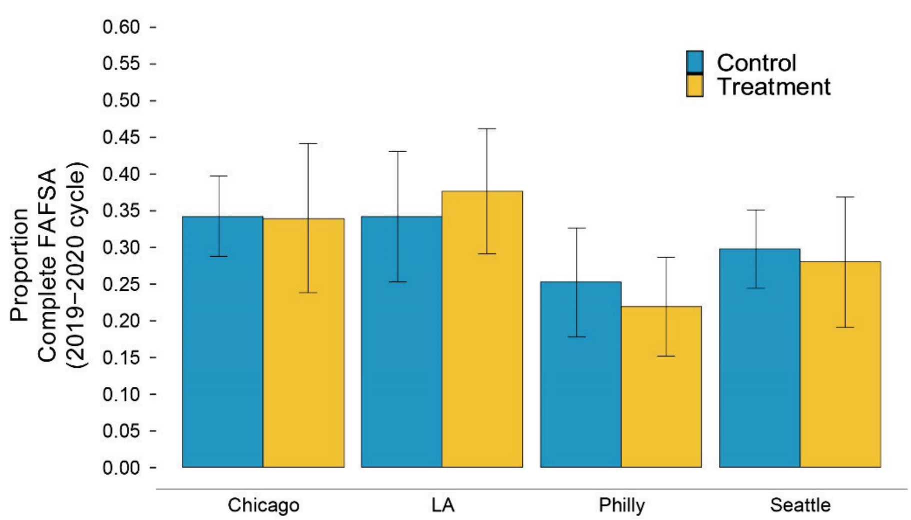


Exhibit 47: Parametric results for FAFSA Completion 2019–2020 cycle (blocking + additional covariate adjustment): separate by PHA. Shows predicted values and 95% CI for each PHA. Other covariates in the model are set to their PHA-specific means.



7.8 Additional results: experimental analysis. Alternate definition of compliance.

We also analyzed a supplementary form of compliance that includes all age and residentially eligible residents, rather than residents who could be matched to the participant lists in the trackers. For this robustness check, compliance was coded as the following:

1. `1 == complier`: youth is matched to the navigator tracker AND the youth has at least one in-person meeting with a navigator

2. `0 = non-complier`: youth is not matched to the navigator tracker OR the youth is matched but has no in-person meetings with a navigator

The reason this version of compliance is less preferred than the version discussed in Section 5.2 is that the zeroes contain a mixture of those whose compliance status is unknown because they are never matched to the tracker and those whose compliance status is known to have never met in person (since they matched). Exhibit 48 presents the results from this specification, which are close to zero and non-significant.

Exhibit 48: Effect of treatment on compliers: alternate definition of compliance

	Dependent variable:
	FAFSA_Complete_2019–2020
complier	–0.027 (0.038) p = 0.478
genderM	–0.118 (0.015) p = 0.000***
raceHispanic	0.037 (0.027) p = 0.175
raceOther	–0.012 (0.035) p = 0.723
tot_annual_income	0.00000 (0.00000) p = 0.002***
PARTICIPANT_CODEIL002	0.008 (0.029) p = 0.786
PARTICIPANT_CODEPA002	–0.095 (0.030) p = 0.002***
PARTICIPANT_CODEWA001	0.006 (0.031) p = 0.844
tot_household_members	–0.014 (0.004) p = 0.0004***
Constant	0.399 (0.033) p = 0.000***
Observations	3,787
Residual Std. Error	0.631 (df = 3777)

Note: $*p<0.1$; $**p<0.05$; $***p<0.01$

7.9 Additional results: experimental analysis. Secondary outcomes

Exhibit 49 summarizes the results for the secondary outcomes, which include FAFSA completion during the 2018–2019 cycle during which navigators were still at the early phases of implementation and the Pell receipt and college enrollment outcomes. The secondary results largely follow the main ones in terms of (1) a close to zero and negative point estimate in the model that just controls for blocking variables, and (2) a positive but imprecise point estimate in the model that controls for other covariates.

Exhibit 49: Experimental results: secondary outcomes

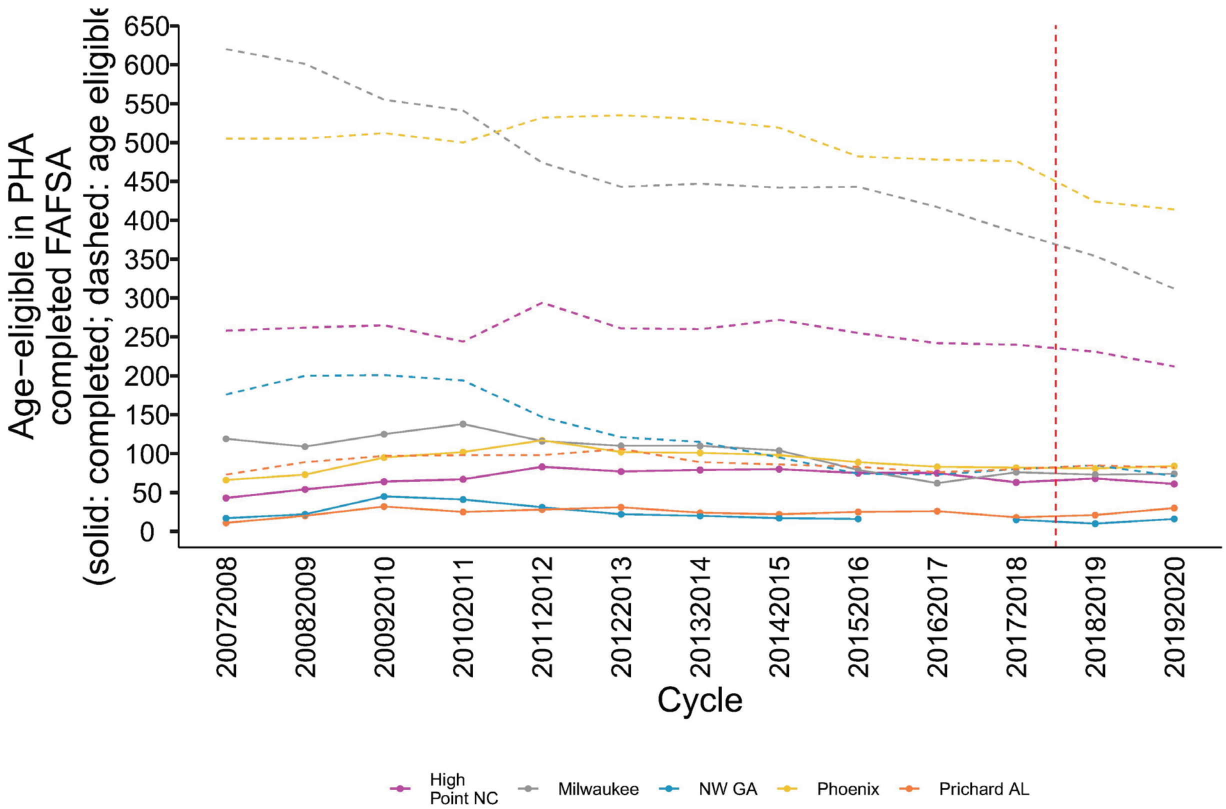
Outcome	Specification	β	SE (robust)	p-value (RI)	p-value
Complete FAFSA (2018-2019)	Blocking	-0.10	0.07	0.80	0.37
Complete FAFSA (2018-2019)	Blocking+dem	-0.20	0.10	0.34	0.80
Enroll college	Blocking	0.02	0.07	0.33	0.80
Enroll college	Blocking+dem	0.46	0.03	0.28	0.42
Enroll Tier I college	Blocking	-0.07	0.07	0.86	0.39
Enroll Tier I college	Blocking+dem	0.02	0.03	0.38	0.98
Enroll Tier II college	Blocking	-0.07	0.08	0.37	0.34
Enroll Tier II college	Blocking+dem	0.15	0.56	0.42	0.80
Enroll Tier III	Blocking	-0.03	0.77	0.80	0.28
Enroll Tier III	Blocking+dem	0.29	0.51	0.77	0.39
Enroll Tier IV	Blocking	0.01	0.28	0.30	0.87
Enroll Tier IV	Blocking+dem	0.15	0.42	0.72	0.79
Receive Pell	Blocking	-0.03	0.13	0.08	0.32
Receive Pell	Blocking+dem	-0.23	0.69	0.98	0.12

7.10 Additional results: synthetic control analysis. Details on analysis and construction of the donor pool.

The main text of Section 5.3 shows variability in FAFSA completion rates over time. Because PHA-level completion rates are a dual function of (1) the number of age-eligible students residing in a PHA, and (2) the count of those students who complete the FAFSA, overall changes in the rate stem from a mix of each of the two sources.

Exhibit 50, an alternative way of presenting the FAFSA completion rates presented in the main text of Exhibit 29, shows the variability in overall rates comes from a mixture of variation in each source.

Exhibit 50: Trends in FAFSA completion: non-experimental treatment PHAs (separating counts of completion from counts of age and residentially eligible). The figure shows general declines in the 17–20 populations in some PHAs.



The generalized synthetic control method uses control units—in our case, PHAs that were neither an experimental grantee nor a non-experimental grantee—to model trends in FAFSA completion and impute counterfactual, post-treatment outcomes for the treated units. This modeling is complicated by the fact that our outcomes are selectively redacted; PHAs who had fewer than 10 students complete the FAFSA in a given year[46] have a missing completion count.

To be in the donor pool, PHAs need at least 1 year of non-redacted FAFSA data. Exhibit 51 shows the number of cycles we observe for each PHA—for the majority of PHAs, which were significantly smaller than the grantees, all cycles are redacted (the large bar at 0). Exhibit 52, a heatmap of this redaction in the pre- versus post-treatment cycles, shows that the redaction occurs in both types of cycles rather than differentially pre- or post-treatment. In sum, this means that smaller PHAs are excluded from the donor pool. Our estimated treatment effects generalize best to the types of PHAs that, similar to the five grantees, have high-enough FAFSA counts to make it into the donor pool.

[46] Or receive a Pell grant or enroll in college, for those outcomes.

Exhibit 51: Redaction of FAFSA counts due to small cell sizes. The figure shows a high count of PHAs that, due to their low number of youth residents, had fewer than ten students complete the FAFSA across many cycles, leading to redaction.

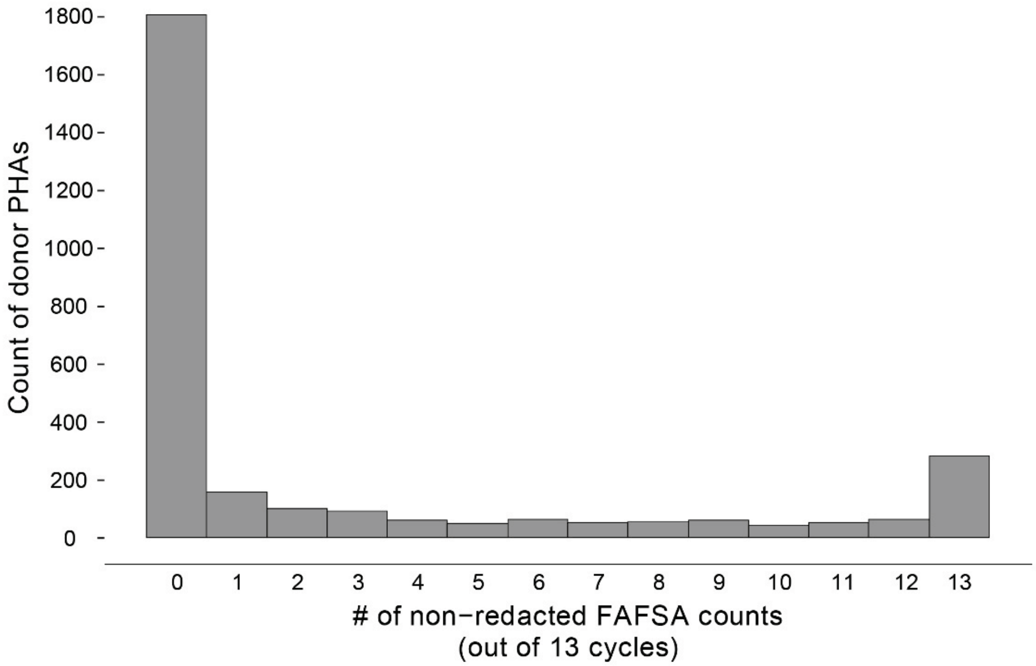
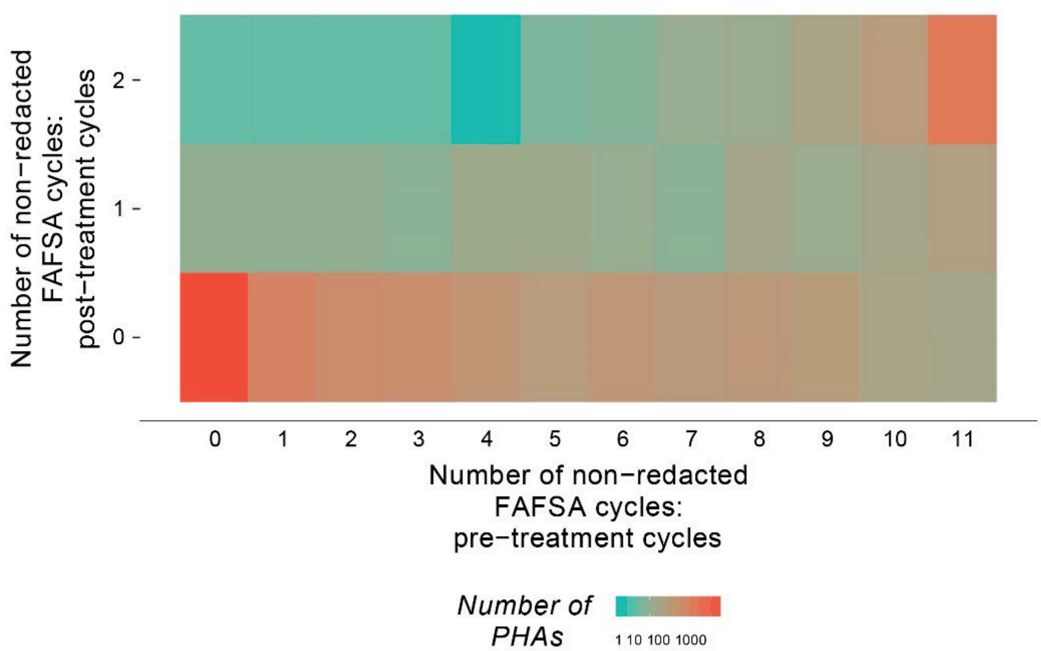


Exhibit 52: Redaction of FAFSA counts due to small cell sizes: pre- versus post-treatment years. The figure shows that the majority of PHAs have 0 non-redacted values in both the pre-treatment FAFSA cycles and the post-treatment FAFSA cycles. This decreases the concern that removing PHAs with redaction from the donor pool induces post-treatment bias.



The following figures provide some insight into PHAs that fall into the following four categories:

1. Non-experimental treatment PHAs. These are the five grantees discussed in Exhibit 2. Four of the five PHAs had complete counts across all 13 cycles; one (NW GA) had one redaction (2016–2017 cycle).

2. PHAs with non-redacted FAFSA completion data across all 13 cycles. These PHAs are ones with thirteen complete cycles of FAFSA completion and are included in both of the specifications discussed in Section 5.3.

3. PHAs with non-redacted FAFSA completion data for at least one of the 13 cycles but that have redaction for at least one cycle. These PHAs are included in the secondary specification discussed in Section 5.3 (PHA-year, which includes PHAs in the donor pool for the years in which they have complete FAFSA data).[47]

4. PHAs with redacted FAFSA completion data for all 13 cycles. These PHAs are not included in the synthetic control analysis donor pool.

Exhibit 53, drawing on the Picture of Subsidized Households data discussed in Section 4.1, shows that the PHAs removed from the donor pool due to fully-redacted data are significantly smaller than the focal treatment PHAs (`people_total`), have more elderly residents (`pct_age62plus`), and have a much lower percentage of minority residents. In contrast, the PHAs that remain in the donor pool are more similar to the treatment PHAs in terms of total residents, minority composition, and fewer elderly residents. Exhibit 54 shows the spatial relationships between the PHAs of each type.

[47] In Exhibit 53, this group is labeled "Included main spec.: excluded PHA-level deletion."

Exhibit 53: Comparison of donor pool PHAs with treatment PHAs: resident attributes. The bars represent the mean across that group of PHA. Attributes are from the Picture of Subsidized Housing data using the year 2016, a pre-treatment year that corresponds roughly to the year that PHAs would be applying to SOAR.

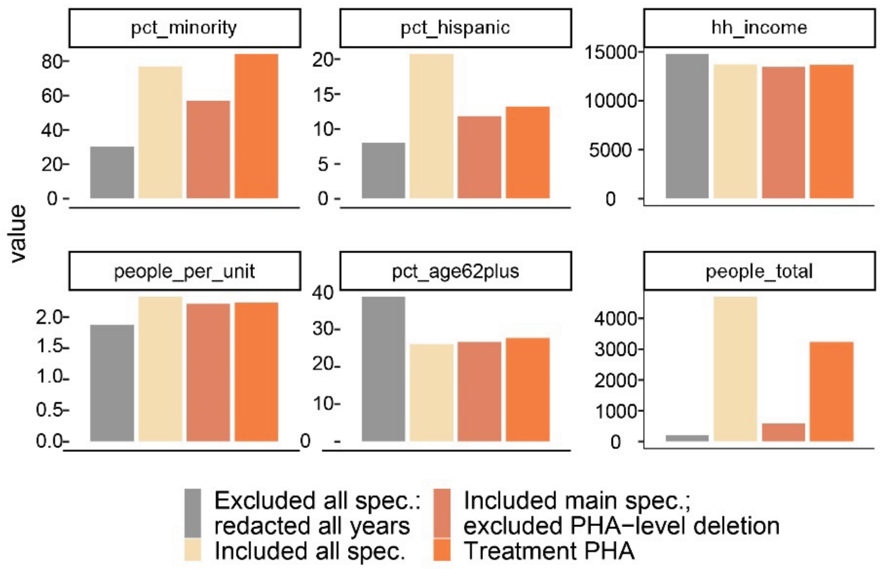
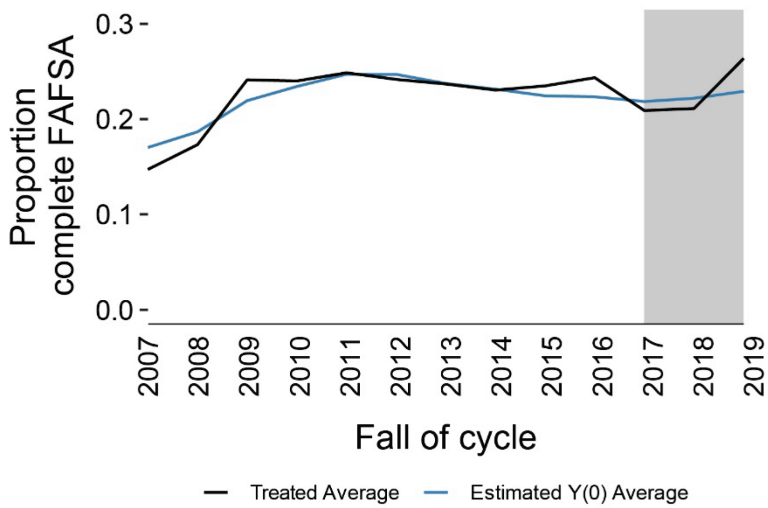


Exhibit 54: Comparison of donor pool PHAs with treatment PHAs: locations. The gray dots represent PHAs that are not in the donor pool due to redaction; the yellow dots represent PHAs always in the donor pool; the ochre dots represent PHAs sometimes contributing to the donor pool depending on the year. The map of these PHAs in relation to the focal treatment PHAs shows that, rather than being spatially proximate, the donor PHAs might represent those in other larger suburbs/cities within the state.

Finally, Exhibit 55 presents another way of visualizing the analytic results of the synthetic control model focused on FAFSA completion. Rather than the treatment effect by year, it shows the match between the observed FAFSA completion rates in the treated PHAs (treated average) and the predicted completion rates in the synthetic control (estimated Y_0 average). In line with the main results, there is an uptick in the treated PHAs relative to the counterfactual controls in FAFSA completion cycle 2019–2020.

Exhibit 55: Observed trends in treated PHA versus counterfactual trends based on donor PHAs.

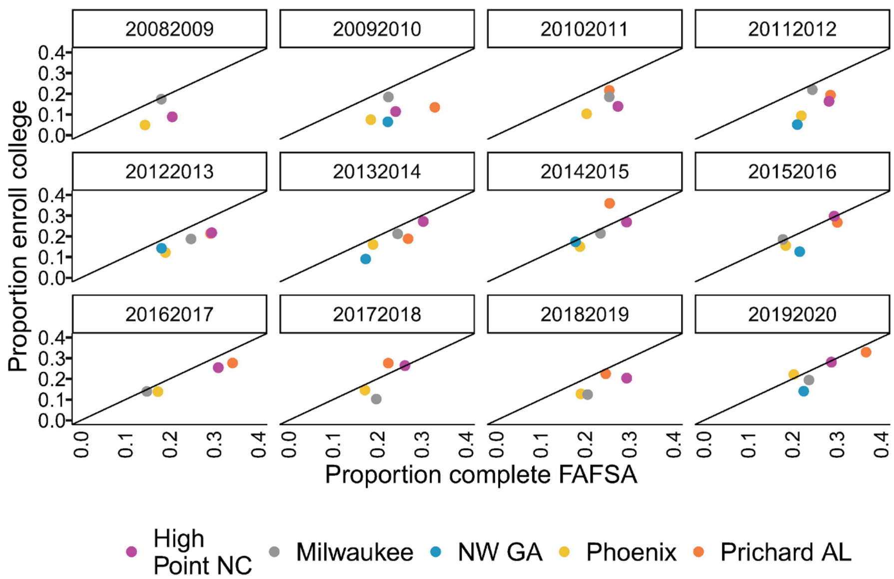


7.11 Additional results: synthetic control analysis. Secondary outcomes.

The main outcome is whether the student completes the FAFSA, an important input to college attendance for low-income students. Here, we analyze results for the secondary outcomes related to college enrollment and Pell receipt.

Exhibit 56, focusing on the non-experimental treatment PHAs, shows that generally, a lower percentage of students enroll in college than complete the FAFSA.

Exhibit 56: FAFSA completion versus college enrollment. Each dot represents the rate for one PHA-cycle.



The generalized synthetic control model showed no significant impact on either Pell receipt or college enrollment in the post-treatment year: 2019–2020. Exhibit 57 presents the results for Pell receipt and shows a point estimate close to zero that is not statistically significant. Exhibit 58 presents the results for college enrollment, which also show a close to zero and non-significant effect during the post-treatment year.

Exhibit 57: Synthetic control treatment effect on Pell receipt by year

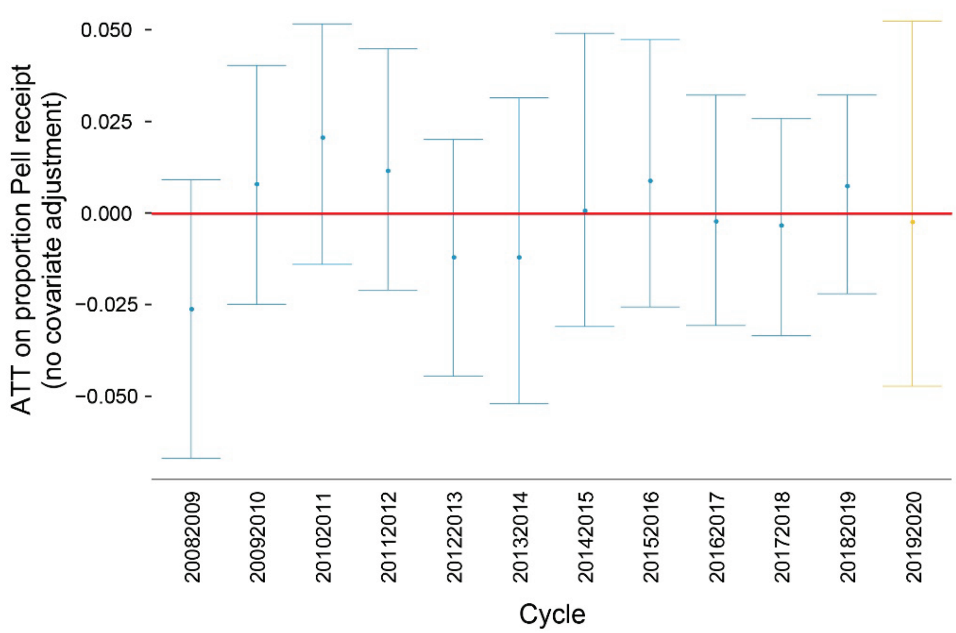
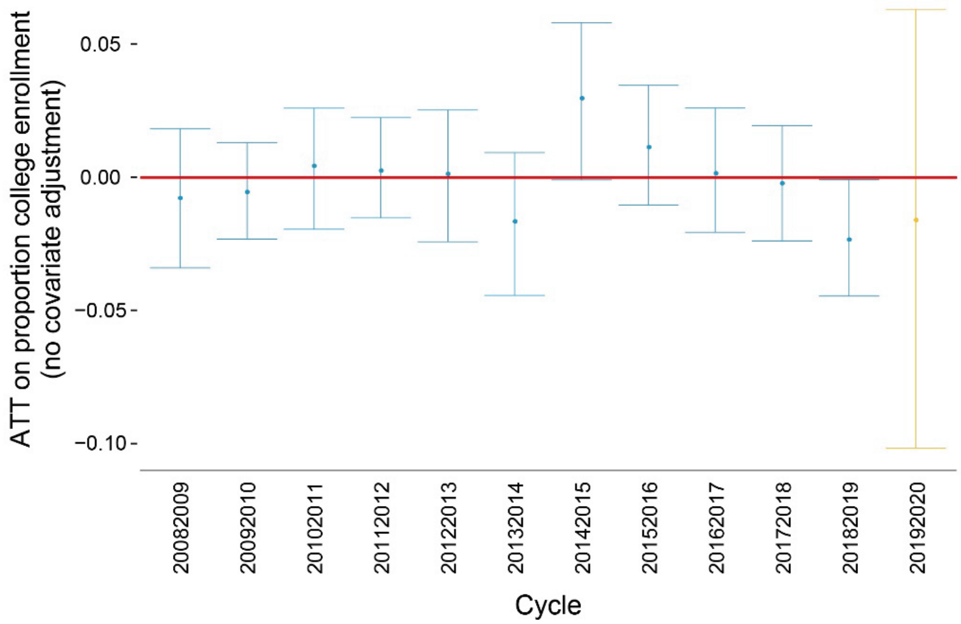


Exhibit 58: Synthetic control treatment effect on college enrollment by year



References

Abadie, Alberto, and Javier Gardeazabal. 2003. "The Economic Costs of Conflict: A Case Study of the Basque Country." *American Economic Review* 93 (1): 113–132.

Ben-Michael, Eli, Avi Feller, and Jesse Rothstein. 2018. "The Augmented Synthetic Control Method." *arXiv preprint arXiv:1811.04170*.

Bertrand, Marianne, Kelly Hallberg, Kenny Hofmeister, Brittany Morgan, and Emma Shirey. 2019. *Increasing Academic Progress Among Low-Income Community College Students: Early Evidence From a Randomized Controlled Trial*. Technical report. University of Chicago Poverty Lab.

Bettinger, Eric P, Bridget Terry Long, Philip Oreopoulos, and Lisa Sanbonmatsu. 2012. "The Role of Application Assistance and Information in College Decisions: Results from the H&R Block FAFSA Experiment." *The Quarterly Journal of Economics* 127 (3): 1205–1242.

Bird, Kelli A, Benjamin L Castleman, Joshua Goodman, and Cait Lamberton. 2017. "Nudging at a National Scale: Experimental Evidence from a FAFSA Completion Campaign." *EdPolicyWorks working paper no. 55*.

Carrell, Scott, and Bruce Sacerdote. 2017. "Why do College-Going Interventions Work?" *American Economic Journal: Applied Economics* 9 (3): 124–51.

Castleman, Benjamin L, and Lindsay C Page. 2015. "Summer Nudging: Can Personalized Text Messages and Peer Mentor Outreach Increase College Going Among Low-Income High School Graduates?" *Journal of Economic Behavior & Organization* 115:144–160.

_____. 2016. "Freshman Year Financial Aid Nudges: An Experiment to Increase FAFSA Renewal and College Persistence." *Journal of Human Resources* 51 (2): 389–415.

Chetty, Raj, John N Friedman, Emmanuel Saez, Nicholas Turner, and Danny Yagan. 2017. *Mobility Report Cards: The Role of Colleges in Intergenerational Mobility*. Technical report. National Bureau of Economic Research.

Clark, Brian, and Ying Shi. 2020. "Low-Income Female Students and the Reversal of the Black-White Gap in High School Graduation." *AERA Open* 6 (2): 2332858420915203.

Deming, David J, Claudia Goldin, Lawrence F Katz, and Noam Yuchtman. 2015. "Can Online Learning Bend the Higher Education Cost Curve?" *American Economic Review* 105 (5): 496–501

Enamorado, Ted, Benjamin Fifield, and Kosuke Imai. 2019. "Using a Probabilistic Model to Assist Merging of Large-Scale Administrative Records." *American Political Science Review* 113 (2): 353– 371.

Lin, Winston. 2013. "Agnostic Notes on Regression Adjustments to Experimental data: Reexamining Freedman's Critique." *The Annals of Applied Statistics* 7 (1): 295–318.

Ma, Jennifer, Sandy Baum, Matea Pender, and C.J. Libassi. 2019. *Trends in College Pricing 2019*. Technical report. College Board.

Ma, Jennifer, Matea Pender, and Meredith Welch. 2019. *Education Pays 2019: The Benefits of Higher Education for Individuals and Society*. Technical report. College Board.

Miller, Cynthia, Camielle Headlam, Michelle Manno, and Dan Cullinan. 2020. *Increasing Community College Graduation Rates with a Proven Model: Three-Year Results from the Accelerated Study in Associate Programs (ASAP) Ohio Demonstration*. Technical report. MDRC.

Office of Evaluation Sciences (OES). 2016. *Increasing FAFSA Completion by HUD-Assisted Youth*. Washington, DC: General Services Administration, OES. https://oes.gsa.gov/projects/hud-youth-fafsa/.

_____. 2017. *Increasing FAFSA Renewal Rates*. Washington, DC: General Services Administration, OES. https ://oes.gsa.gov/projects/increasing-fafsa-renewal-rates.

_____. 2019a. *Increasing FAFSA Completion Among Public Housing Residents: NYCHA*. Washington, DC: General Services Administration, OES. https://oes. gsa.gov/projects/increasing-fafsa-completion-nycha/.

_____. 2019b. *Increasing FAFSA Completion Among Public Housing Residents: Seattle and King County*. Washington, DC: General Services Administration, OES. https://oes.gsa.gov/projects/increasing-fafsa-completion-seattle-and-king-county/.

Page, Lindsay C, Benjamin L Castleman, and Katharine Meyer. 2020. "Customized Nudging to Improve FAFSA Completion and Income Verification." *Educational Evaluation and Policy Analysis* 42 (1): 3–21.

Page, Lindsay C, Stacy S Kehoe, Benjamin L Castleman, and Gumilang Aryo Sahadewo. 2019. "More than Dollars for Scholars: The Impact of the Dell Scholars Program on College Access, Persistence, and Degree Attainment." *Journal of Human Resources* 54 (3): 683–725.

Phillips, Meredith, and Sarah J Reber. 2019. *Does Virtual Advising Increase College Enrollment? Evidence from a Random Assignment College Access Field Experiment*. Technical report. National Bureau of Economic Research.

Scott-Clayton, Judith. 2012. *Information Constraints and Financial Aid Policy*. Technical report. National Bureau of Economic Research.

Scrivener, Susan, Michael J Weiss, Alyssa Ratledge, Timothy Rudd, Colleen Sommo, and Hannah Fresques. 2015. *Doubling Graduation Rates: Three-Year Effects of CUNY's Accelerated*

Study in Associate Programs (ASAP) for Developmental Education Students. Technical report. MDRC.

Xu, Yiqing. 2017. "Generalized Synthetic Control Method: Causal Inference with Interactive Fixed Effects Models." *Political Analysis* 25 (1): 57–76.

U.S. Department of Housing and Urban Development
Office of Policy Development and Research
Washington, DC 20410-6000

U.S. DEPARTMENT OF HOUSING AND URBAN DEVELOPMENT

EQUAL HOUSING OPPORTUNITY

August 2021