Analysis Plan

Project title: Project SOAR
Project code: 1732

1 Project Description

Project SOAR (Students + Opportunities + Achievements = Results) is a new 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 provides grant funding to nine public housing authorities (PHAs) to deploy counselors—or "education navigators"—to provide individualized assistance to public housing youth between the ages of 15 and 20 to make postsecondary education enrollment plans, and transition to and succeed in postsecondary programs.[1]

The main objective of the navigators is to increase Free Application for Federal Student Aid (FAFSA) completion among eligible residents. Additional objectives include increasing financial literacy, increasing the number of college applications, reducing summer melt (when students are accepted to a program and pay a deposit but fail to matriculate), and increasing enrollment in postsecondary education and training programs. Some of these objectives will not be observed (e.g., financial literacy) and will not be a part of the impact analysis. OES is working with HUD and the Department of Education (ED) to evaluate the effectiveness of Project SOAR. The remainder of this document describes the planned evaluation, including data and statistical models.

1.1 Project SOAR Design

Project SOAR targets multiple behaviors related to pursuing a post-secondary education. Each grantee, a public housing authority (PHA), will hire between one and three navigators to help residents between 15-20 years of age. Each navigator is expected to assist approximately 100-125 students over the course of a year. While the primary goal of the intervention is to help students complete the FAFSA, there are four total objectives defined in the grant:

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

Nine Public Housing Authorities were selected for the demonstration and were funded to employ education navigators to carry out tasks in support of the four main program objectives. Table 1 lists each housing authority, the number of navigators awarded to each, and whether the PHA is part of the experimental evaluation—where navigators were randomly assigned to individuals—or part of a non-experimental component. The table also includes the number of

[1] The decision to focus on residents aged 15-20, rather than older ages that might reflect the higher rates of non-traditional college-going among low-income individuals, was based on (1) the desire to use some eligibility criteria to make navigators' task of assisting residents more feasible, and (2) the assumption that a higher proportion of 15-20 year olds were interested in college-going than, for instance, 25-30 year olds.

[2] 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, like help registering for courses and figuring out their living arrangements.

Table 1: PHAs in study

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

15-20 year olds who were living in a public housing development within the PHA at the time of the grant award and the number of youth each PHA proposed to serve.[3]

Figure 1 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 SY 2018-2019 FAFSA season on October 1, 2017.

Given the differing speed of the hiring process, each grantee effectively began providing services at different points in time between August and November 2017.[4] The grant supported navigators through April 2019, which means that the navigators' services could help students with both the 2018-2019 FAFSA season and the 2019-2020 FAFSA season. We focus on the 2019-2020 FAFSA season because, by then, navigators had a year to support students with the application process.[5]

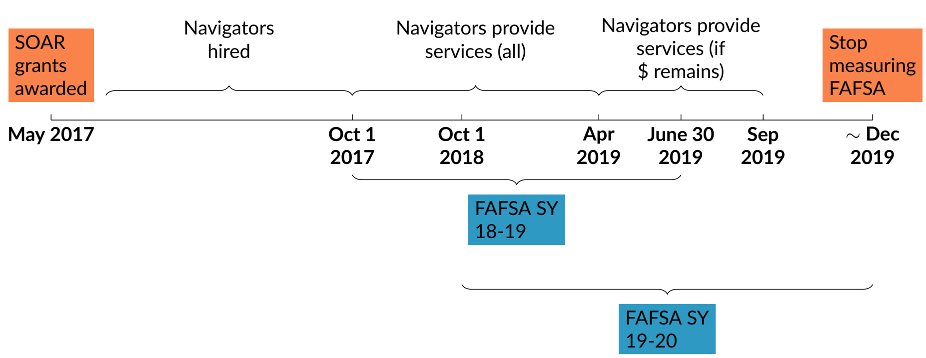

Figure 1: Timeline for navigators serving residents and FAFSA application cycles

[3] The number of navigators was roughly scaled to the size of the eligible student population that PHAs reported to HUD when they applied for the grants. PHAs also differed in whether the navigators were employed full time or part time. Nevertheless, there was substantial variation across PHAs in the number of navigators per eligible student, which led to an experimental design that utilizes the oversubscribed sites to randomly assign the navigators.

[4] As a result, the end date of hiring and start date of services in Figure 1 is an approximation and varies across PHAs.

[5] Grantees were allowed to continue providing services until September 30, 2019 if they had unobligated funds after April. 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. We estimate that the match will take place in February 2020, and that this date will be sometime in December 2019 or January 2020.

1.2 Project SOAR Evaluation

There will be three separate analyses. Here, we preview each analysis, before describing each in more detail.

Analysis one (descriptive engagement): descriptive analysis of whom navigators engage (all PHAs)

Whether a navigator improves a student's postsecondary outcomes depends on at least two factors. First, the navigator needs to engage a student. Second, among the students whom the navigator engages, the quality of the engagement has to change the student's behavior, leading to improved FAFSA completion and progress on other postsecondary outcomes. Navigators have a broad pool of age-eligible (15-20) students they aim to assist, and different students might be more or less receptive to assistance. The first analysis will examine the characteristics of the students whom navigators (1) try to contact and (2) successfully engage. As we discuss in greater detail in Section 2.1, PHAs varied substantially in how they tracked these interactions. Therefore, our analysis focuses on comparing engagement *within* a PHA across students, rather than between PHAs.

Analysis two (experimental): experimental impact analysis (4 PHAs)

Four of the grantees—Chicago, Los Angeles, Philadelphia, and Seattle—had particularly high youth to navigator ratios to the extent that navigators would not be able to attempt to serve everyone. As a feature of the grant, these four PHAs were chosen to randomly select a group of youth who would be offered the services of education navigators (the treatment group) and another group of youth who would not be eligible for SOAR services (the control group). These four PHAs comprise the experimental component of the evaluation. The analysis will compare the postsecondary outcomes of the youth selected for the treatment group to outcomes of the youth selected for the control group.

Analysis three (quasi-experimental): quasi-experimental impact analysis (5 PHAs)

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 services in these PHAs because there was not excess demand (i.e., caseloads were small enough that navigators could be reasonably expected to engage with all eligible students), and these PHAs were excluded from the experimental component of the impact evaluation. One PHA (Milwaukee) had similar 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.[6] In each of the five PHAs not participating in the experimental component, navigators attempted to provide assistance to all residents in the targeted age range. The analysis will use quasi-experimental methods to estimate the impact of Project SOAR by comparing the postsecondary outcomes of students in these five PHAs to students in similar PHAs that were not selected for the grant.

2 Data and Data Structure

This section describes variables that will be analyzed, as well as changes that will be made to the raw data with respect to data structure and variables.

In the Appendix, we outline examples of the three main types of data structures:

1. Individual-level data: we use these data for the analysis of which residents navigators engage (descriptive engagement analysis) and a version of the experimental analysis that adjusts the estimates for the fact that, in practice, navigators only served a subset of all treatment group youth.[7]
2. AMP-level data: we use these data for the main causal estimates of the impact of navigators on college outcomes (experimental analysis).
3. PHA-level data: we use these data for the quasi-experimental estimates of the impact of navigators on college outcomes (quasi-experimental analysis).

[6] 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.

[7] As we discussed earlier, this is a mix of who navigators contact and which youth are interested in their services.

2.1 Data Sources

There are multiple sources of data that will be used for the three different analyses. This section describes the main data sources, indicates which analyses they will support, gives a short narrative on how they will be used, and highlights limitations that affect our conclusions.

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.

- *Analyses*: Descriptive, Experimental, Quasi-experimental
- *What these data help us investigate*: PIC contains some information that can help us understand if navigators interacted with some types of students more than others; for instance, whether navigators served younger or older students. The data also allow us to calculate AMP-level attributes to control for baseline demographic characteristics of the public housing residents that remain imbalanced following randomization.
- *Limitations*: Residents' interest in attending college likely affects whom navigators end up serving. While PIC contains demographic variables that are likely correlated with a resident's likelihood of attending college (e.g., age; race/ethnicity; gender; disability), it does not collect youth residents' grades, test scores, or other academic outcomes relevant for college going.

Enterprise Data Warehouse and Analytics (EDWA): EDWA is a data warehouse maintained by the Department of Education which contains information on student's interaction with the Federal post-secondary educational system. In particular, EDWA contains information on FAFSA completion, post-secondary enrollment, and Federal student aid. A memorandum of understanding (MOU) between HUD and ED allows 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 will provide aggregate-level data back to HUD in a form that cannot be re-identified. ED may also agree to analyze the person-level data and report summary statistics from regressions.

- *Analyses*: Experimental, Quasi-experimental
- *What these data help us investigate*: These data include the main outcomes of the study. They will be used to estimate if there were any changes in FAFSA completion or other postsecondary outcomes for students eligible for SOAR services.
- *Limitations*: Navigators served residents during two FAFSA completion cycles (Figure 1). One began shortly after navigators were hired (October 1st, 2017), ending June 30th, 2019. The other began when navigators were about a year into their tenure, starting October 1st, 2018 and ending June 30th, 2020. As Figure 1 highlights, due to the timing of our analysis, with the match conducted in February of 2020, we do not observe the full second cycle of FAFSA completion. Instead, we will likely only observe FAFSA completion up through December 2019 or January 2020.

College Type: in addition to examining whether the student enrolls in college, the intervention might also impact the type of college students enroll in. 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 degrees. Following prior work (Chetty et al. 2017; Deming et al. 2015), we will group colleges into five tiers of selectivity:[8]

1. Highly selective colleges: these encompass tiers 1 through 4 in the Barron's rating system, or about 200 colleges.

[8] 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. For instance, Chetty et al distinguish between Ivy-Plus and other universities. If there are still inadequate cell sizes to preserve privacy in one of the tiers, we will first combine the Highly selective colleges and selective colleges tier into one tier. Then, we will combine the four-year and two-year public and not-for-profit college tiers. This would then leave the distinction between students enrolling in either a two or four-year public/not-for-profit college (tiers one - four) versus students enrolling in a two or four-year for-profit college.

2. Selective colleges: these encompass tiers 5 through 6 in the Barron's rating system, or about 1000 colleges.
3. Non-selective four-year colleges: tiers 7 and 8.
4. Non-selective two-year public and not-for-profit colleges: tier 9.
5. Non-selective private, for-profit colleges, two or four-year: tiers 10 and 11.

- *Analyses*: Experimental, Quasi-experimental
- *What these data help us investigate*: If there are general increases in postsecondary enrollment, these data will help determine what types of schools the enrollment increases are concentrated in. As a note, we will only include this analysis if the Department of Education agrees to tabulate counts separately by these tiers.
- *Limitations*: College enrollment is a function of (1) which colleges a student applied to, (2) which colleges accepted the student's application (if applicable), and (3) which college a student chose to enroll in. By observing enrollment, we miss possible impacts on outcomes like the breadth of colleges a student applied to.

SOAR data systems: For Project SOAR, HUD developed a data tracking tool which navigators were instructed to use to document program activities, with grantees expected to use it consistently across sites. We expect the tracking tool to include a roster of residents and (1) the date of an interaction between a resident and a navigator, (2) the type of interaction, and (3) the length of the interaction. The tool contains a unique key for each individual for linking to PIC, and via PIC to information held by the Department of Education. However, as we outline in the limitations, initial submissions suggest that grantees vary substantially in how they record these activities.

- *Analyses*: Descriptive, Experimental
- *What these data help us investigate*: the tracking tool helps us with the descriptive analysis of which students navigators help and with the experimental analysis that looks at effects adjusted for compliance.
- *Limitations*: We do not yet have the complete data set, but initial submissions suggest that (1) the data may be incomplete (e.g., interactions are left unrecorded) and (2) different grantees may use different terminology to describe the same interaction (e.g., differences in what counts as a contact attempt). We will use the fields that seem most consistently recorded for our secondary analysis that measures the impact of navigators on residents actually served (part of the experimental analysis). For the descriptive engagement analysis, we will be cautious in the conclusions we draw; for instance, we will avoid comparing engagement across PHAs that may result more from reporting inconsistencies than actual differences in engagement. Instead, we will focus on within-PHA comparisons. All analyses relying on the SOAR data systems are contingent on the data being of sufficient quality. If we find that data are unusable, we may not be able to complete some of the descriptive analyses as planned and will indicate in the final report how and why the analyses diverted from the analysis plan.

Site visit summaries: An initial set of site visits were conducted in July and August 2018 by teams of between two and four staff from OES and HUD. 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 interactions with potential SOAR participants, group events, and visits to several of the housing developments. There was a second set of site visits conducted in 2019. OES will work with HUD to incorporate any learnings from the site visits to add to the descriptive analysis.

- *Analyses*: Descriptive
- *What these data help us investigate*: Interview data will shed light on how PHAs designed their SOAR programs to address local challenges, what navigators perceived to be the major barriers preventing their students from enrolling in a postsecondary program, and the challenges navigators faced in delivering services to eligible students.

- *Limitations*: The site visits included interviews with grantee staff and navigators, but did not include interviews with youth residents or their parents. Therefore, our discussion of barriers that youth face will focus on navigators' perspectives on these barriers, rather than residents' direct reports.

Seattle Public School (SPS) District Student Information System (tentative): While PIC provides some information on students, it lacks important information on students' academic characteristics, such as the students' grades and whether or not they are on track to graduate. To gain a richer portrait of whom navigators serve, albeit one confined to a single PHA, we will attempt to partner with SPS to get academic records for SHA students. Ideally, the data will contain end-of-year grades and cumulative attendance information. This analysis is contingent on executing a data sharing agreement with SPS.

- *Analyses*: Descriptive
- *What these data help us investigate*: Past FAFSA research focuses on interventions targeted towards students with a higher-than-average propensity towards attending college. Descriptively, these data allow for an investigation of whether navigators focus on similar students or end up serving students who struggle more with attendance and academic outcomes and thus may fall through the cracks of other college access programs.
- *Limitations*: SPS has a flag for which students were enrolled in SOAR, but this flag is limited to the 2017-2018 school year (SOAR's first year). Therefore, the analysis will be limited to residents who engage with the navigator during the first year of the program's implementation. Additionally, it is not clear how enrollment is defined, meaning it is unclear the level of engagement students had with navigators.[9] Availability of these data is contingent on the approval of a Data Sharing Agreement.

2.2 Outcome Variables to be Analyzed

This section describes the different outcomes measures for the three analyses and how each will be created.

Descriptive study of whom navigators engage

We will construct three outcome variables for the descriptive study of whom navigators engage, subject to data availability and data quality:[10]

- *Navigator attempted to engage student (yes or no)*: this reflects some personal outreach method by the navigator to the student, which might include visiting the family's apartment or texting or emailing the student. This does not include a generic communication sent to all eligible residents, which we will ascertain by excluding outreach attempts where the interaction tracker lists all residents as recipients of the outreach.
- *Navigator engaged student (yes or no)*: this reflects that the navigator and the student or his or her parent had at least one in-person meeting, which includes group events.
- *Quantity of interactions among students whom the navigator engaged (continuous measure)*: this is a continuous version of the previous variable, so captures both personal meetings between the navigator and either a student or parent, as well as group sessions.[11]

Impact analysis of navigator assistance

The primary outcomes 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, with Figure 1 showing the relationship with the timeline for navigator services. We will restrict the analytic sample for the impact analysis to the population of students who are at a stage in their high school or post-graduate careers where FAFSA completion is appropriate.[12]

[9] Due to staffing changes at SHA, we were unable to find how the indicator was originally developed in the file send to SPS for data matching.

[10] More precisely, if one of the data elements seems to be collected too unreliably to learn useful insights—for instance, if the trackers seem to mainly record successful interactions rather than contact attempts—we will only analyze the other outcomes.

[11] If this data seems reliably collected, we may include an exploratory analysis of dosage effects in the main impact analysis.

[12] The analysis will only include FAFSA completions through approximately December 31, 2019. While individuals still can complete the FAFSA until June 30, 2020 it is unlikely that those who are attending a program in fall of 2019 would complete the FAFSA after enrolling in

1. *Main analytic sample for confirmatory analysis*: students who are 17 years old or older at some point between October 1st, 2018 and June 30th, 2020. This roughly corresponds to students who were high school seniors or graduated during the 2019-2020 FAFSA cycle. The main reason for this inclusion criteria is that although students aged 15 and 16, who are likely high school sophomores or juniors, can receive navigator services, students would not need to complete the FAFSA until the fall of their senior year. Therefore, the measure restricts the analytic sample, and the denominator of the FAFSA completion rate, to students who would be completing the FAFSA for the first time as a high school senior and 19 and 20 year olds who may be completing the FAFSA either because they are entering a post-secondary program for the first time or because they are already enrolled and need to renew the FAFSA for the next year.
2. *Secondary analytic sample for robustness check/non-confirmatory analysis*: chronological age is a rough proxy for whether or not the student is a high school senior, but variation across districts in age cutoffs for enrollment complicate whether 17 year olds are likely to be high school seniors. In addition, on the upper end of the age window, FAFSA completion rates drop off over time, peaking at ages 17 and 18 and declining thereafter. Therefore, a secondary analysis will use the school cutoff criteria at the school districts that correspond to the four experimental PHAs to include students in the analytic sample if their birth date is 1) after the relevant school cutoff to be a senior before June 30, 2020 but 2) makes it likely that they are a senior during the 2018-2019 or 2019-2020 school years (therefore, excluding 19 or 20 year olds who graduated before these years).[13]

Secondary outcomes include rates of post-secondary enrollment, institution type (public, private, or proprietary), program length (2-year or 4-year), program selectivity, and Pell Grant receipt. Other exploratory outcomes could include persistence and eventually repayment outcomes.[14]

3 Statistical Models and Hypotheses

3.1 Descriptive analysis of whom navigators engage

The descriptive analysis will use both administrative data and data gathered from staff interviews to provide better context to the other analyses. The two sources of data may conflict or corroborate one another. The main goals of the descriptive analysis are to provide understanding of how PHAs approached program implementation and what features of the local context were helpful or presented challenges. Additionally, the analysis will seek to understand more about the students who were engaged with SOAR. That includes trying to understand who navigators contacted, who ended up interacting with navigators, and what kinds of interactions took place.

3.2 Experimental analysis of navigator impact

This section describes the statistical models and hypothesis tests that will make up the analysis for the four PHAs participating in the randomized component of the study.

Random assignment process

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

OES 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

a program.

[13] We will only use this analytic sample for the main experimental analysis, and not for the synthetic control analysis. That is because it is infeasible to 1) match each PHA to a school district, and then 2) find the school cutoff dates for each matched district.

[14] We will post an updated pre-analysis plan if we obtain persistence and repayment data. One can imagine effects on persistence and repayment that emerge even if there are null effects on certain college-going outcomes. For instance, even if treatment and control students attend colleges at similar rates, if treatment students attend a college that is a better financial deal—for instance, a local community college rather than a private, for-profit college—we might see improvements in repayment that are not reflected in the more general college-going outcomes.

for various reasons—for example, due to geography that would make travel difficult for education navigators, overlap with other similar resident service programs (e.g., Jobs Plus), or other local knowledge. After grouping and exclusions, a total of 77 AMPs were eligible for random assignment.

Figure 2, focusing on the Chicago Housing Authority as an example, aggregates households to the AMP level and shows AMPs randomized to treatment (orange) or control (blue). 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. Figure 3 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 versus distributed their units are.

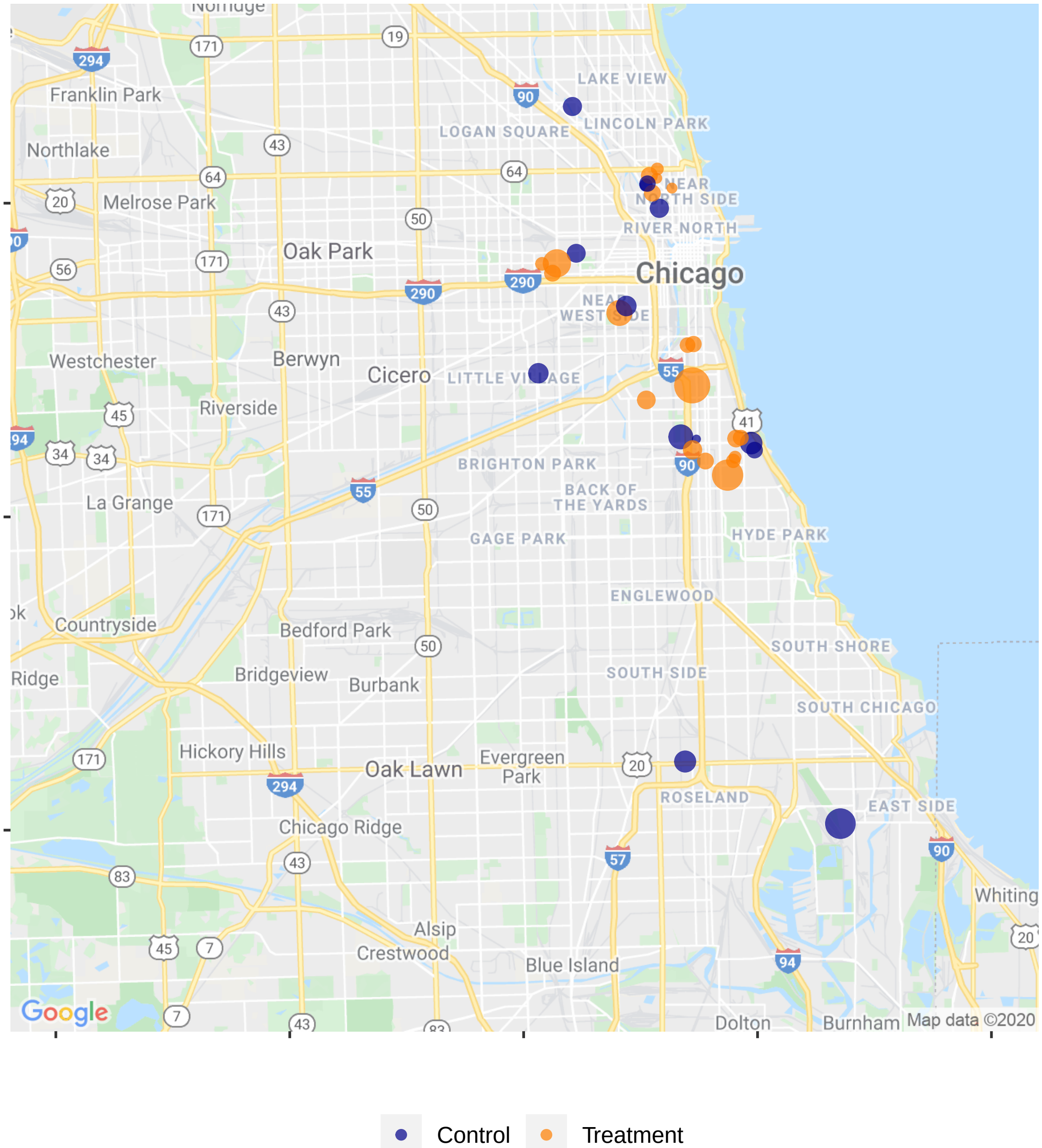

Figure 2: Map of AMP randomizations in Chicago at the AMP level. AMPs are placed at the mean latitude and longitude of units. The size of the dots are scaled to the number of age-eligible youth in each AMP.

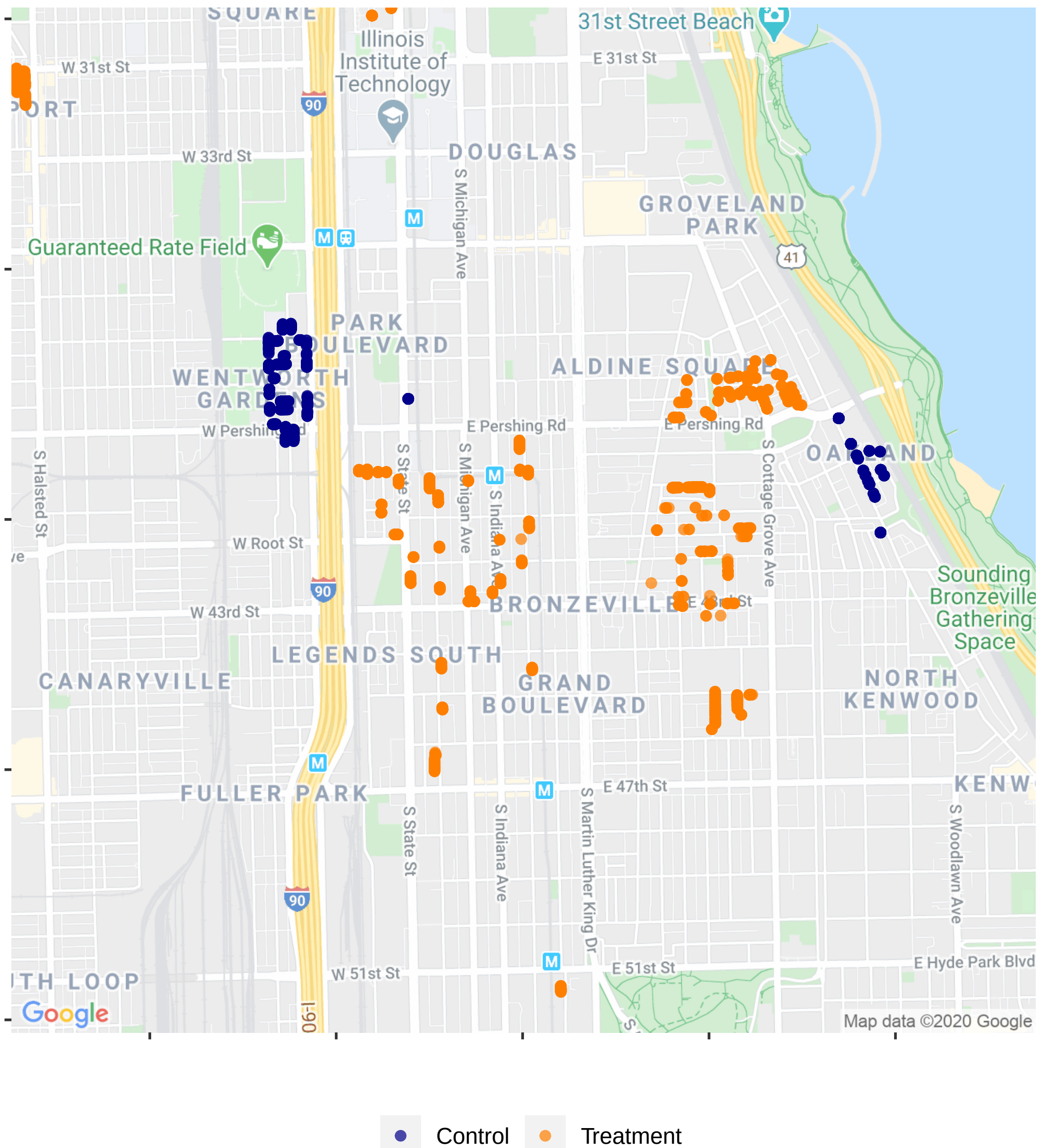

Figure 3: Map of AMP randomizations in Chicago at the unit level zooming in on 3 AMPs. The map shows how the randomization helped minimize potential spillovers while still resulting in neighborhoods with both treatment and control group students due to the clustering of different AMPs in the same neighborhood.

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 a youth 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 engage with other

PHA staff for assistance, but they were instructed not to provide any personalized assistance.[15]

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

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, then
3. Subsequent AMPs were assigned to treatment by progressing down the sorted list until the final AMP assigned to treatment exceeded the maximum workload, then
4. All remaining AMPs were assigned to control.

Statistical models

The primary model, estimating the intent to treat effect, will apply OLS to a regression of the outcome of interest (y_a) on an indicator for treatment (T_a) and a series of blocking variables (X) including grantee dummies and indicators for the size of each AMP, where a indexes the modified AMPs used as the unit of randomization. Additionally, because the assignment mechanism results in larger AMPs having slightly higher probabilities of being selected for the treatment group, the model will use inverse probability weights to account for the estimated probability of selection into the treatment and control groups.[16] Finally, all models will use the Lin estimator (Lin 2013), which involves:

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, we will estimate the following using a linear model:

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

Each outcome of interest will be computed as a percentage, 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 we outlined earlier, which include (1) restricting to those 17 and older for the main specification, (2) restricting to high school seniors and above based on school cutoff dates in the alternative specification.

The estimate of interest is β_1, which is the estimated effect of the intervention on AMPs randomized to treatment. The roster of eligible students used for calculating AMP-level outcomes will include all students who were eligible to receive SOAR services between October 1, 2017 and March 31, 2019.

A second model will include AMP-level baseline characteristics.[17] These are in a matrix which will include AMP-level means of the following covariates to control for potential imbalances caused by (1) AMP-level randomization, (2) potential differences in the demographic composition of residents in different AMPs. We chose covariates that are likely correlated with the college-going behavior of youth in that AMP and/or the households' openness to navigator help:[18]

[15] 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.

[16] Probabilities will be estimated via simulation. The weights used will 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.

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

[18] For instance, mixed citizenship families may face greater confusion about eligibility for aid.

- Percentage of households that self report Black race/ethnicity (white held out as reference category)
- Percentage of households that self report Hispanic/Latino race/ethnicity
- Percentage of households that self report Other race/ethnicity
- Total number of residents in the AMP
- Average household income of residents
- 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
- Percentage of household members who are ineligible noncitizens

Similar to the main specification, we will use the Lin estimator that uses the mean-centered versions of these covariates. We use Z to denote the combined matrix of 1) AMP-level covariates, and 2) the blocking variables we include in the above specification.

$$y_a = \beta_0 + \beta_1 T_a + \gamma \tilde{Z}_a + \delta T_a \tilde{Z}_a + \epsilon_a$$

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. Yet, the descriptive analysis will likely show that, rather than treating all students, workload issues and challenges like family reluctance to engage, means that only some fraction of youth in treatment AMPs actually meet with navigators. As a result, we expect to observe non-compliance: the presence of a navigator in an AMP increases the likelihood that a youth will engage with that navigator, but some fraction of youth in treatment AMPs will not engage. Similarly, the physical structure of buildings in a PHA, and the fact that some navigators held group events like field trips from which it was difficult to exclude control students, led to non-compliance in the form of some control students receiving navigator help.[19]

As a result, a third model will examine the effect of the AMPs in the presence of this non-compliance. For this, our preferred specification will require individual rather than AMP-level data, with individuals now indexed by i. We will estimate the following two models (subject to data availability and quality):

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:[20]

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

[19] 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 may not be able to observe compliance in the form of those assigned to control receiving treatment.

[20] We will use the following covariates drawn from PIC that we expect to be correlated with navigator engagement/outcomes, and that may remain imbalanced between groups: dummy indicators for Black, hispanic, or other; household income; highest grade of education completed by household head; whether household head is employed full time; whether the family was homeless prior to PHA entry; whether there is an undocumented household member. We will not use student absenteeism and achievement covariates for this analysis, because those will be limited to one PHA.

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

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

Due to the data limitations we discuss earlier, the outcome variable in this first model may have systematic measurement error—for instance, if navigators consistently *underreport* engagement, we underestimate the number of youth served and end up with potential overestimates of the treatment effect. However, given that preliminary data suggests that navigators may have served fewer than 50 percent of age-eligible residents, some adjustment is important.

Inference criteria, including any adjustments for multiple comparisons

The decision rule will be 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 (randomization inference, or RI). The analysis studying FAFSA completion, and using the p-values from RI, will be considered confirmatory. The below code outlines the calculation for p-values once we have a long-form data frame where there is:

- An observed treatment effect
- Treatment effects from m = 500 permutations

```
## function to return two-side p-values
get_ri_p <- function(coef_data, obs_coef_name,
                     ri_coef_name){
    ri_coef_name = sym(ri_coef_name)
    obs_coef_name = sym(obs_coef_name)
    ri_results = coef_data %>%
        summarise(upper_p = mean(!!ri_coef_name >= !!obs_coef_name),
                  lower_p = mean(!!ri_coef_name <= !!obs_coef_name),
                  two_sided_p = 2*min(upper_p, lower_p))
    return(ri_results)

}


## example with data simulated to have strong
## treatment effect
get_ri_p(coef_data = coef_data, obs_coef_name = "obs_coef_sig",
         ri_coef_name = "ri_coef_null")

## example with data simulated to have null treatment
## effect
get_ri_p(coef_data = coef_data, obs_coef_name = "obs_coef_null",
         ri_coef_name = "ri_coef_null")
```

We will also conduct a secondary analysis of FAFSA completion that will use p-values from linear regression and heteroskedastic-consistent standard errors rather than randomization inference, with the caveat that the small number of AMPs makes assumptions behind those p-values less credible.[21]

[21]We will use the HC2 specification to estimate standard errors, which has a small sample correction; however, the p-values from the random-

Secondary outcomes will be treated as exploratory and are not fully specified in this analysis plan. Potential secondary outcomes of interest include rates of post-secondary enrollment, institution choice, and financial aid outcomes.

Imported variables

The analytic dataset will be in the structure of having one observation per AMP and columns for relevant variables. The dataset will include outcome variables obtained from ED, baseline covariates obtained from PIC, and navigator-youth engagement variables from the SOAR data system, to the extent that these variables are available.

Transformations of variables

We expect HUD to provide individual-level covariates, and we will calculate AMP-level means and percentages to be included in the AMP-level regression models.

Outcome data provided by ED will all be counts at the AMP level. Where possible given ED's restrictions on identifiability of data, we will ask for counts by subgroup, such as gender and age. From the counts provided, we will create rates for each of the key outcomes, for example the percent of eligible individuals who completed the FAFSA, enrolled in a post-secondary institution, and received a Pell Grant For outcomes variables that do not lend themselves to rates, we will calculate AMP-level means.

Data exclusion

It is anticipated that data will be relatively complete and accurate. Any individual not matching a record in EDWA is assumed to not have an interaction with the federal post-secondary educational system. While this will almost certainly incorrectly include some individuals who do not match because of mis-entered SSNs or name mismatches, prior matches between HUD and ED show the rate of matching errors is low. The high successful match rate is due to both sources of identifying information being stringently verified. The same is true for all data in PIC used for eligibility determinations—for example, household composition and household income. As such, it is not anticipated data will need to be excluded.

Treatment of missing data for covariates

In the case of individual level-missing data, we will impute the characteristics using multiple imputation with $m = 20$ replicates,[22] and AMP-level aggregates will be created based on the mean across these replicates. The code below provides an example of imputation with simulated data that has missingness in household income and where household income is correlated with the highest level of education for the household head:

```
library(Amelia)
## creating simulated data
n_students = 3000
indiv_data_forimp = data.frame(household_id = sample(as.character(seq(from = 1111,
                                            to = 9999, by = 1)),
                                            n_students, replace = FALSE),
                        amp = sample(1:77, n_students, replace = TRUE),
                        hh_highested = runif(n_students, min = 7, max = 16))


n_lessthanhs = nrow(indiv_data_forimp %>% filter(hh_highested < 12))
n_hsorhigher = nrow(indiv_data_forimp %>% filter(hh_highested >= 12))
```

ization inference procedure will generate exact p-values that rely on fewer sample assumptions.

[22] We will use Amelia in R.

```
## add missingness on household income (cor with hh highest education)
indiv_data_forimp = indiv_data_forimp %>%
              mutate(hh_income_obs = ifelse(hh_highested < 12,
                          rnorm(n_lessthanhs, mean = 5000, sd = 1000),
                          rnorm(n_hsorhigher, mean = 10000, sd = 2000)),
              missing_indicator = sample(c(0, 1), size = n_students,
                          replace = TRUE,
                          prob = c(0.95, 0.05)),
              hh_income_wmiss = ifelse(missing_indicator == 1,
                          NA, hh_income_obs)) %>%
              dplyr::select(-hh_income_obs, -missing_indicator)

## first impute at individual level
n_replicates= 20
indivlevel_impute <- amelia(indiv_data_forimp, m=n_replicates,
                    parallel = "multicore",
                    noms = c("amp"),
                    idvars = c("household_id"))


## aggregate to the amp level
## since not getting se, just take
## mean over all replicates
## since rubin's rules just apply to se's/parameter estimates
indivlevel_impute_df = indivlevel_impute$imputations
amplevel_summaries = do.call(rbind.data.frame, indivlevel_impute_df) %>%
            group_by(amp) %>%
            summarise_if(is.numeric,
                         funs(mean))
```

Limitations

Power: 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.

For the permutation-based inference, we are powered to detect a 6-7 percentage point change with 80 percent power. Figure 4 summarizes the p-value distribution on the treatment effect at different increases in FAFSA completion; we see that the p-values move towards peaking towards 0 once the effect size becomes approximately 6 percentage points.

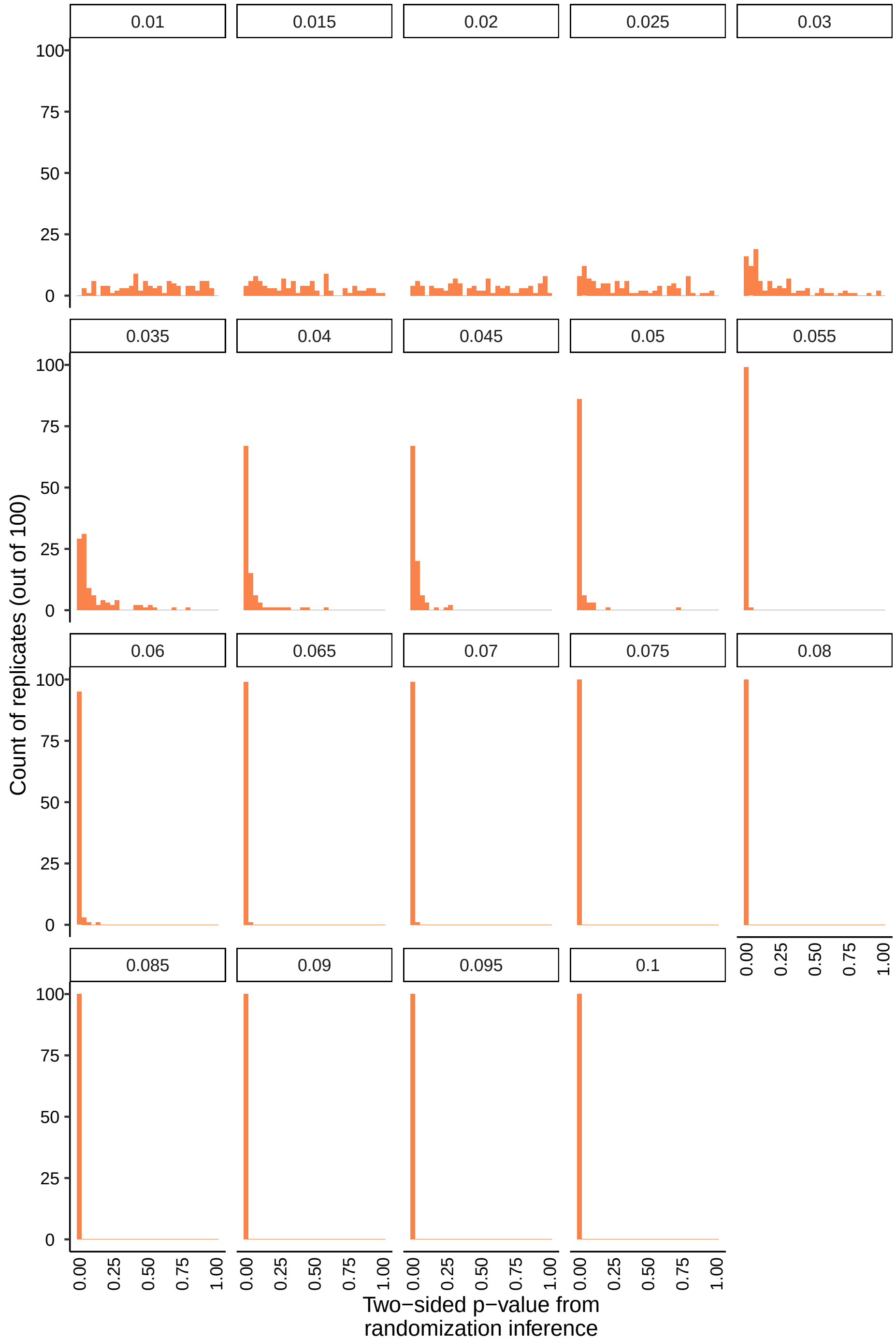

Figure 4: Average power to detect effect at different increases in FAFSA completion (randomization inference-based p value)

For the linear regression-based inference, we are powered to detect a slightly smaller 5-6 percentage point change with 80 percent power, which is smaller than the effect sizes in other studies of the effect of help on FAFSA completion. Figure 5 shows the power at different increases in FAFSA completion, averaged across 1000 simulations, while Figure 6 shows the p-value distribution on the treatment effect at different increases in FAFSA completion, with the full distribution across simulations.

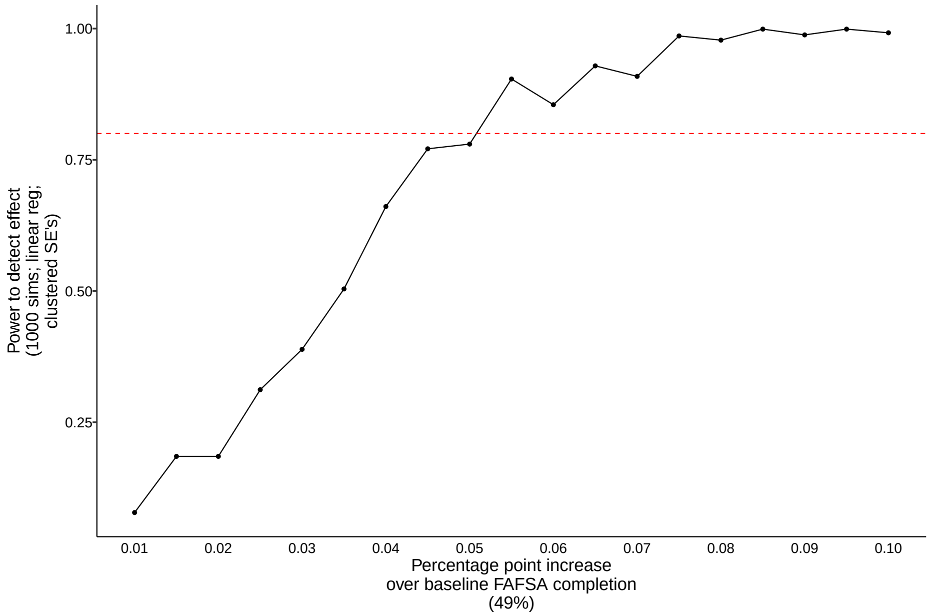

Figure 5: Average power to detect effect at different increases in FAFSA completion

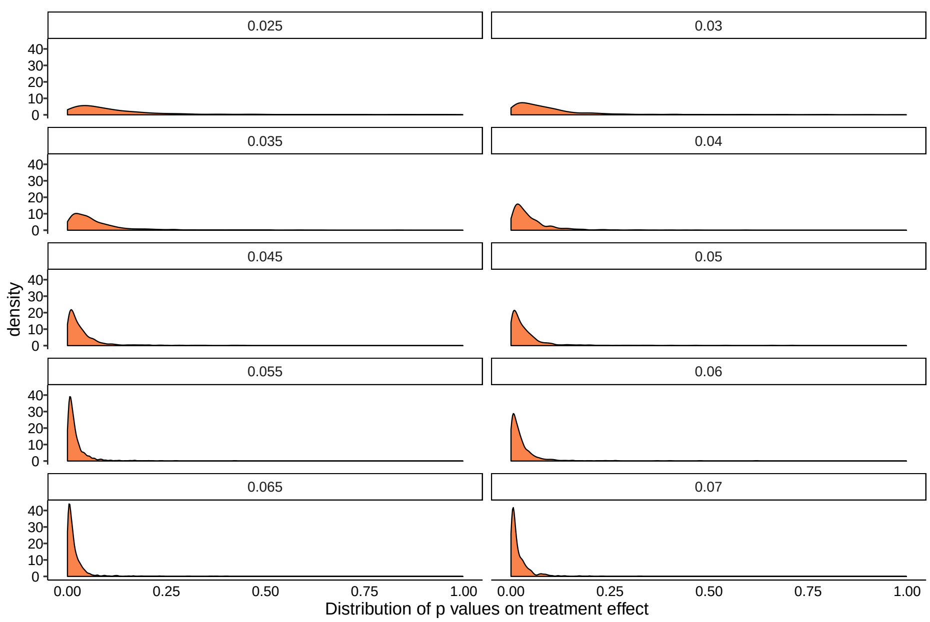

Figure 6: Distribution of p values on treatment effect across simulations

Implementation challenges: Recruitment has been a challenge for most or all grantees. Delays in hiring meant navigators were unable to develop any presence in the communities before the first FAFSA season opened on Oct 1, 2017. Based on interviews conducted over the summer of 2018, many navigators felt they were only just beginning to gain visibility and trust among residents.

Non-compliance: There is undoubtedly some non-compliance among individuals in the control group. 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 normal 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 they drew a clear line around in-person assistance, which also lends itself to focusing only on FAFSA completion as a primary outcome. We do not view these incidents as the result of spillover (i.e., the control person's decision is influenced because of their relationship with someone in the treatment group) but rather of non-compliance (e.g., the person is in a semi-public space and learns about the opportunity independent of those in the treatment group). There are also questions about how much information can be included in a referral to other services (i.e., would students find out about these services otherwise?) without being considered part of the treatment services.

Our third analysis, which predicts engagement with a navigator as a function of treatment status, leads to estimates that could adjust for this non-compliance. Navigators were instructed to track all of their interactions with students and their parents, including with students and parents from control AMPs, which would in theory provide the necessary data; however, early reports suggest that navigators may not always have recorded control group students who were served in the data tracker. Without accurate data, it may not be possible to understand the extent of non-compliance.

Program maturation: There likely are significant differences in program operations year over year even outside of the normal programmatic changes that would be expected. Navigators in the grantees that did not subcontract out to existing programs were learning and building skills over the course of the grant, and services provided during the first half of the grant were likely of different quality than those provided over the second half of the grant. There also were challenges replacing departing staff in the second year of the grant given the limited time the position will be available. The implementation analysis will help describe some of these challenges in more detail.

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 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 of 2019, Illinois passed a "universal FAFSA completion" law that requires students to apply for FAFSA in order to graduate high school.[23] 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.

Exploratory analysis

Given some challenges of program maturation and getting to steady state operations, a set of exploratory analyses will focus on the first and second year cohorts separately.

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
```

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

```r
# 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)
}
```

4 Non-experimental Analysis of Navigator Impact

4.1 Statistical Models & Hypothesis Tests: Synthetic Control Analysis

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

The effect of Project SOAR in the non-experimental PHAs will be 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 potential 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. The approach here will generally rely on the more recent three step process as described in Xu's generalized synthetic control method (Xu 2017):

1. Modeling the relationship of available covariates and outcomes using only the (untreated) donor pool and both pre- and post-treatment data.
2. Using the predicted outcome values from the model created in Step (1) to construct the synthetic comparison for the treated units in the pre-treatment period.
3. Using the weighted synthetic comparison created in Step (2) to predict the counterfactual outcomes for the treated unit(s) in the post-treatment period(s).

Donor pool

The donor pool will be made up of all PHAs with the exception of the four PHAs with AMP-level randomization in Table 1.[24] These PHAs will vary in attributes like size, FAFSA completion rates, and demographic composition. During the synthetic control construction process, PHAs that have very different attributes for these pre-treatment characteristics and/or trajectories of FAFSA completion in the pre-treatment period will be unlikely to be selected as part of a synthetic control for a focal PHA.

Inference criteria, including any adjustments for multiple comparisons

The synthetic control method allows for a simple difference in means estimator. Unfortunately, not all synthetic control methods have well described methods for assessing the statistical properties of the estimates. We will rely on the bootstrapping method for determining p-values and confidence intervals described by and implemented via the gsynth R package as our reported values (Xu 2017); however, these results should be interpreted with caution.

Transformations of variables

Data from ED will be aggregate counts at the PHA level, which will be converted into rates (average number of eligible individuals completing an action) or means.

[24] Since there are over 2000 PHAs, each with many years of beneficiary data, if there are processing time issues for creating the donor pool data from PIC uisng all PHAs, we will restrict to a smaller sample of PHAs that potentially (1) excludes smaller PHAs likely to have volatile completion rates (e.g., PHAs with fewer than 50 age-eligible youth), and (2) randomly samples from the remaining.

Imported variables

Additional PHA-level covariates measure the demographics of either (1) residents aged 15-20 during the synthetic control time window or (2) all public housing residents in the PHA. Depending on the data matching process, we will either create PHA-level aggregate characteristics (e.g., percent of different race/ethnicity groups) from the PIC data described above, which would allow the PHA-level characteristics to reflect characteristics of the youth eligible for FAFSA completion. Or, if the data matching process is better completed with data already aggregated to the PHA level, we will use the Picture of Subsidized Households public data for both the pre- and post-intervention period. These data include averages of various resident characteristics including household income, household composition, and race and ethnicity, but the characteristics reflect all residents rather than those aged 15-20. The PHA-level characteristics will be joined to the outcome data and used in the modeling.

Limitations

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 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 post-secondary initiative which could bias the estimated effect towards zero, assuming such programs increase post-secondary activity.

The selection of the synthetic unit can be sensitive to the choice of method specification and functional form. As such we also will benchmark the results of our main results by showing the results of other common synthetic control method approaches, for example the augmented synthetic control method described by (Ben-Michael, Feller, and Rothstein 2018) (R package: augsynth) or the original synthetic control method described by (Abadie and Gardeazabal 2003) (R package: synth) to give a sense for how sensitive the results are to the particular method.

5 Appendix

In this section, we describe the three different data structures.

5.1 AMP-level data (confirmatory analysis of effect on FAFSA completion)

Table 2 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.

Table 2: 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

5.2 Individual-level data (exploratory analysis of whom navigators engage)

Similarly, Table 3 uses simulated data to show the structure of the individual-level data.[25] We use the individual-level data for the descriptive analysis of whom navigators serve, and to adjust the main causal estimates to measure the treatment on those served (rather than treatment on those eligible).

[25] For simplicity, we omit covariates.

Table 3: 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	metwith_nav	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
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

5.3 PHA-level data (exploratory analysis of non-experimental PHAs)

Table 4 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 will show average rates of FAFSA completion over time and will be used to create a synthetic control for the five non-experimental grantees.

Table 4: 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

References

Abadie, Alberto, and Javier Gardeazabal. 2003. "The Economic Costs of Conflict: A Case Study of the Basque Country." *American Economic Review* 93 (1): 113–32.

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

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

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.

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

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