Analysis Plan

Project Name: Does Reducing Documentation Burden Broaden Access to Emergency Rental Assistance? Quasi-experimental evidence from Virginia

Project Code: 2301

Date Finalized: February 17, 2023

Note: This analysis plan was written after receiving outcome data and is therefore not blind to outcomes. It is intended to document and justify the analytic choices.

Analysis Plan Summary

This evaluation is part of the Office of Evaluation Sciences (OES) American Rescue Plan Act of 2021 (ARP) portfolio. The ARP was designed to address immediate needs related to the pandemic, with a specific focus on addressing historically disparate outcomes across race, class, and geography that were further exacerbated by the pandemic. As federal programs are innovating and finding new ways to achieve these goals, the OES portfolio of evaluations will measure whether ARP-funded interventions are working as intended and share lessons learned.

In support of the ARP Equity Learning Agenda, OES is working with agency partners to better understand how to improve awareness, access, and allocation of ARP programs and resources, focusing on ARP programs with equity goals. This set of evaluations will be intentional and strategic in building evidence to understand the role of ARP programs and supported interventions in improving outcomes for historically underserved populations.

This analysis plan describes a quasi-experimental evaluation of the impact of simplifying documentation requirements when applying for Emergency Rental Assistance. We examine the effects of a “fact-specific proxy” (FSP) introduced by Virginia’s Department of Housing and Community Development (VA DHCD) to broaden and streamline access to assistance. The FSP used the applicant’s ZIP code as a proxy for income eligibility, simplifying the requirement of documenting income eligibility for some applicants and not others. Simplifying income eligibility verification represents a substantial documentation burden reduction. Our general goal in the project is to ask: to what extent does simplifying the individual requirement to document income eligibility for applicants in relevant zip codes increase applications (especially among underserved groups) and reduce processing times? We analyze application data aggregated to the ZIP code level in order to answer this question.

Project Description

The Consolidated Appropriations Act (2021) and the American Rescue Plan Act (2021) created the Emergency Rental Assistance (ERA) Programs (known as ERA1 and ERA2, respectively), making approximately $46B in funding available to cities, counties, tribal communities (for ERA1), the District of Columbia, U.S. Territories, and states ("grantees") to assist households that experience financial hardship to pay rent or utilities, with the goal of preventing eviction or housing instability in the wake of the pandemic. The program provided financial assistance to renters and landlords for rent, utilities, and other housing related expenses. Renters had to meet eligibility criteria[1] to receive assistance, outlined as follows:

1. At risk of housing instability or homelessness;
2. Experience of financial hardship due, directly or indirectly, to COVID-19 (ERA1); or experience of financial hardship during or due, directly or indirectly, to COVID-19 (ERA2);
3. Have income that falls below an area-specific threshold

Grantees had latitude in how they could design their programs, and notably took advantage of "program flexibilities" that were highlighted by the US Department of the Treasury for improving the application process to quickly and equitably distribute ERA. Some examples included simplifying application forms, incorporating self-attestation of income (or self-certification), using fact-specific proxies (FSP), using categorical eligibility, and adding additional prioritization tiers for those with highest needs. These innovations offered promising opportunities to learn what works to reduce documentation burdens for underserved groups to increase program access and/or successful receipt of funds.[2] The ARP Equity Learning Agenda identifies learning opportunities about ERA program flexibilities: "*To what extent did low income renters benefit from the administrative flexibilities (such as self-attestation) that Treasury made available to Emergency Rental Assistance grantees?*"

Virginia's Department of Housing and Community Development (VA DHCD) was an early adopter of program flexibilities to reduce applicants' documentation burdens. In this project, we focus on one type of documentation burden commonly imposed on applicants: documenting income eligibility through the upload of paystubs, bank account statements, proof of receipt of other benefits, or, for those with no bank accounts or income, uploading signed documents. Administrative burdens such as these may result in higher rates of incomplete applications and slower distribution of funds, leading to inequitable outcomes especially among underserved groups.

The VA DHCD used program flexibilities to simplify the requirement of documenting income eligibility for some applicants. Importantly, all applicants were still subject to the two other eligibility criteria outlined above — experiencing housing instability and financial hardship during or due to the pandemic. However, these criteria were straightforward to establish: tenants could

[1] See Table A1 in appendix for a full description of these criteria.

[2] Such documentation burdens constitute what the Office of Management and Budget considers administrative burdens. Others (in addition to time spent on applications and paperwork) include factors like time spent traveling to in-person visits, answering notices and phone calls to verify eligibility, navigating web interfaces, and collecting any documentation required to prove eligibility.

self-report that they were experiencing housing instability and financial hardship on their application.[3] By contrast, income eligibility verification took more work, as we describe below, so its simplification represents a substantial documentation burden reduction. Our general goal in the project is to ask: to what extent does streamlining the requirement for providing individual income eligibility documentation broaden and quicken access to assistance?

Status quo before the program change

VA DHCD's Rent Relief Program (RRP) started in June 2020 and ended in May 2022, with funding from the state's Housing Trust Fund, the Coronavirus Relief Fund, ERA1, and ERA2. In December 2020, the program evolved from a '30+ door' model, with a multitude of different vendors processing applications, to a '2 door' model, in which one vendor processed tenant-initiated applications and another processed landlord-initiated applications. From December 2021 to May 2022, a single vendor processed all applications.

The main change we propose to study took place in the middle of the program, starting with the development and sharing of a list of FSP-eligible ZIP codes with vendors that processed VA's ERA applications on June 10, 2021, and leading to the simplification of required individual income documentation for tenants living in FSP-eligible ZIP codes in the subsequent weeks. Before this program change, tenants and landlords in all areas of Virginia that did not have their own ERA programs[4] and who were considering whether or not to apply for ERA faced several types of income-related verification requirements. In addition to a written attestation of their income in the application form, tenants who applied needed to further document their income eligibility using official and recent documents uploaded to the online system (e.g., a pay stub, W-2, other wage statement, tax filing, or bank statement). Landlords applying on their tenants' behalf would need to collect these documents from eligible tenants and submit them alongside their landlord application. Applicants claiming that they had zero income or that they did not have a bank account and were only paid with cash were required to document this by downloading, signing, and uploading a document attesting to this claim. After the implementation of the FSP, VA DHCD simplified the requirement to individually document income eligibility, but only in certain ZIP codes and only for certain households. We explain how this worked in detail below.

Quasi-experimental design

We leverage the fact that the requirement to upload proof of income eligibility was removed via a "fact-specific proxy" or FSP to identify the causal impact of this simplification on access to relief. In particular, we use our understanding of how the FSP simplified income eligibility verification for some potential applicants and not others to draw an analogy to an "ideal experiment" in which potential applicants are randomly assigned to have or not have the requirement to upload proof of income eligibility.

---

[3] A screenshot of the relevant portion of the tenant application is shown in Figure A1.

[4] Chesterfield county and Fairfax County ran their own ERA programs and are therefore excluded from this study. VA DHCD communicated with these programs to ensure they did not duplicate benefits.

While the FSP did *not* randomly assign the simplification of income eligibility verification, it emulates important features of this ideal experiment. This program change introduced two forms of variation in potential applicants' exposure to income eligibility documentation requirements:

1. Temporal variation: some tenants and landlords who applied before the implementation of the FSP would not have had to document income eligibility if they had applied after; *and*
2. Geographic variation: after implementation of FSP, otherwise-similar applicants in non-FSP ZIP codes were required to document income eligibility, whereas their counterparts in FSP ZIP codes were not.

We now turn to describing this program change in greater detail.

How the fact-specific proxy (FSP) worked

In February 2021, the US Department of the Treasury made changes to the guidance for the ERA program, to provide additional flexibility with respect to documenting the eligibility of households. This program change allowed the use of program flexibilities such as an FSP (see detailed timeline in Figure 1). The VA DHCD began developing a list of ZIP codes to use in implementing a FSP, and alerted vendors to its forthcoming implementation with additional training and guidance (though they were not aware of which ZIP codes would be eligible). On June 10, 2021, the VA DHCD shared the list of eligible ZIP codes with vendors that processed its ERA applications. Soon after, vendors began using the ZIP code list to shorten the income documentation review process, and communicated the change over email to applicants. Over the following two weeks, VA DHCD's vendor simplified income documentation in FSP ZIP codes on the tenant-facing online application portal, using the applicant's location of residence (e.g., ZIP code) as a proxy for requiring the upload of documents to establish applicant-level income eligibility. If the median household income in a ZIP code, estimated through the Census' 2019 5-year American Community Survey, fell below the 2021 HUD-defined statewide low-income limit for three-person households, then all households with three or fewer members in that ZIP code were considered presumptively income eligible and did not have to upload additional documentation of income eligibility, beyond a written attestation as part of the application form. The "treatment" in our analogy to an experiment is the removal of the need to upload additional income eligibility documents via the FSP. The specific rules determining what forms of income eligibility were required following the implementation of the FSP are outlined on Figure 2. Tenants considering applying could not have known whether they lived in an FSP eligible ZIP code until after 6/10/21, after which the online application was changed such that eligible households did not have to upload additional documentation of income eligibility. The list of FSP eligible ZIP codes was not made public prior to this date, by either VA DHCD or their vendors. This treatment is assigned to potential applicants (tenants and landlords) at the ZIP-code level, through the designation of all tenant households with three or fewer members in some ZIP codes as income eligible because the median income of three-person households in those ZIP codes falls below the HUD-determined low income threshold.

Figure 1. Implementation of Rent and Mortgage Relief Program

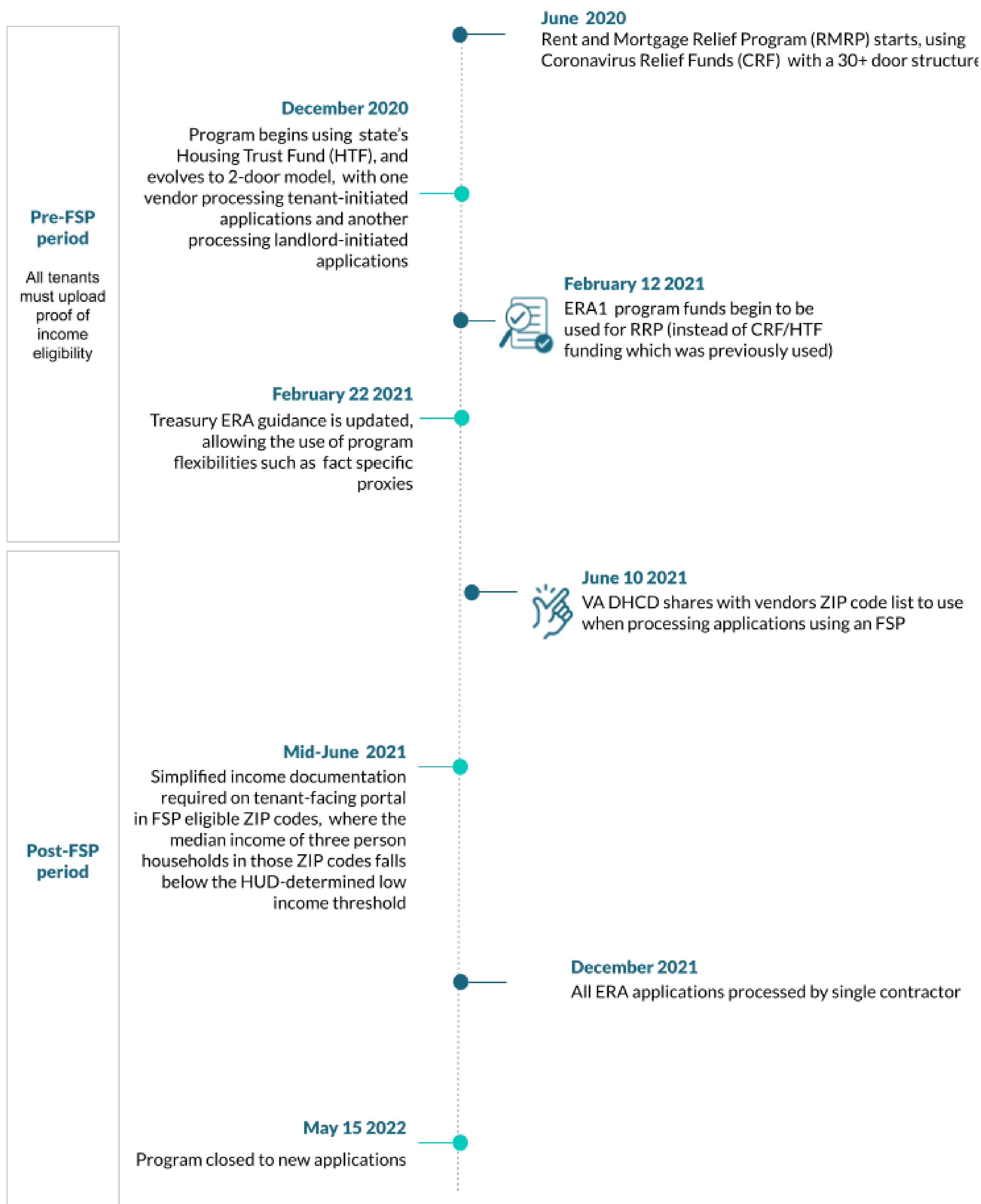


Figure 2. Applicant eligibility for relief from simplified income documentation after program change

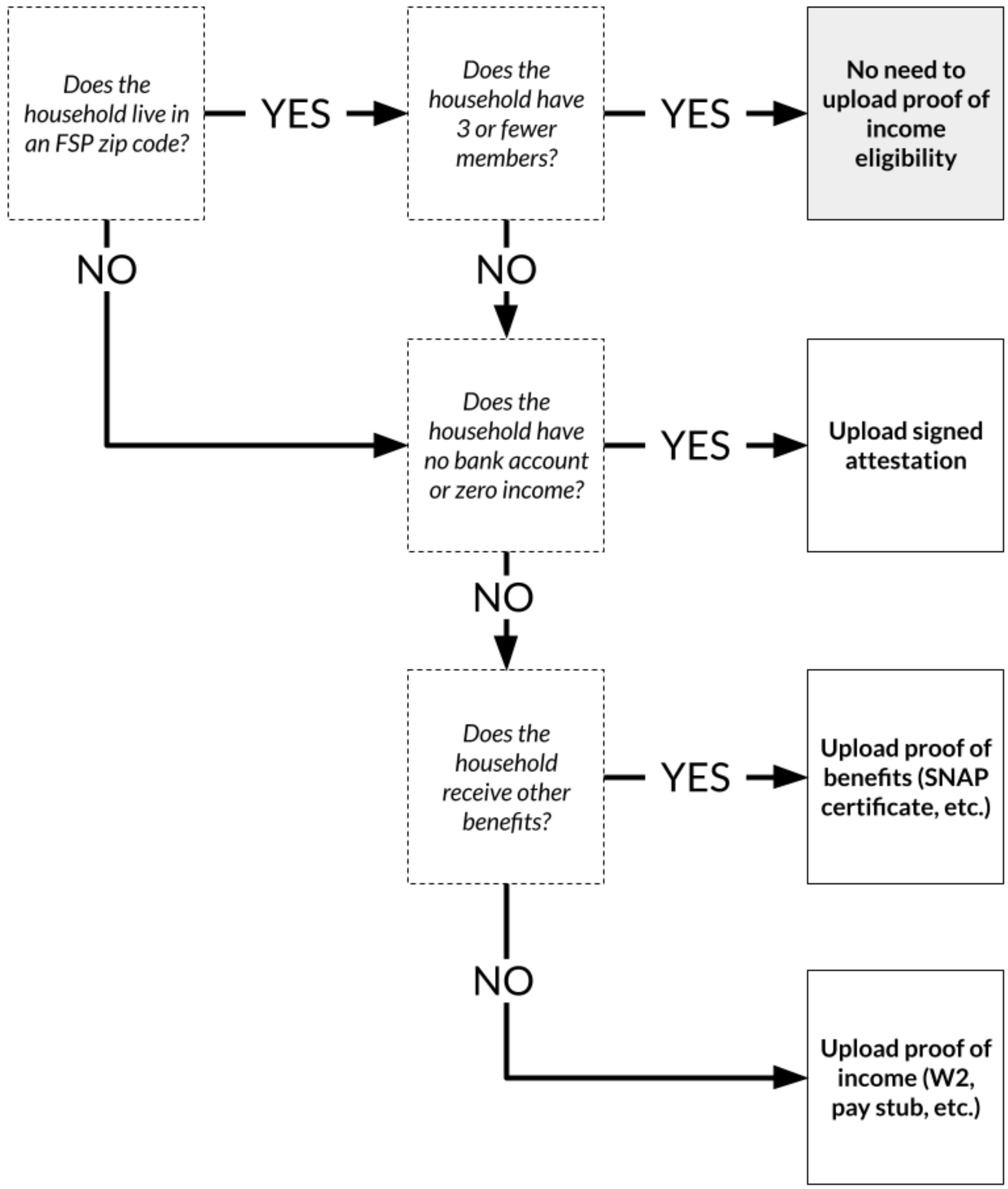
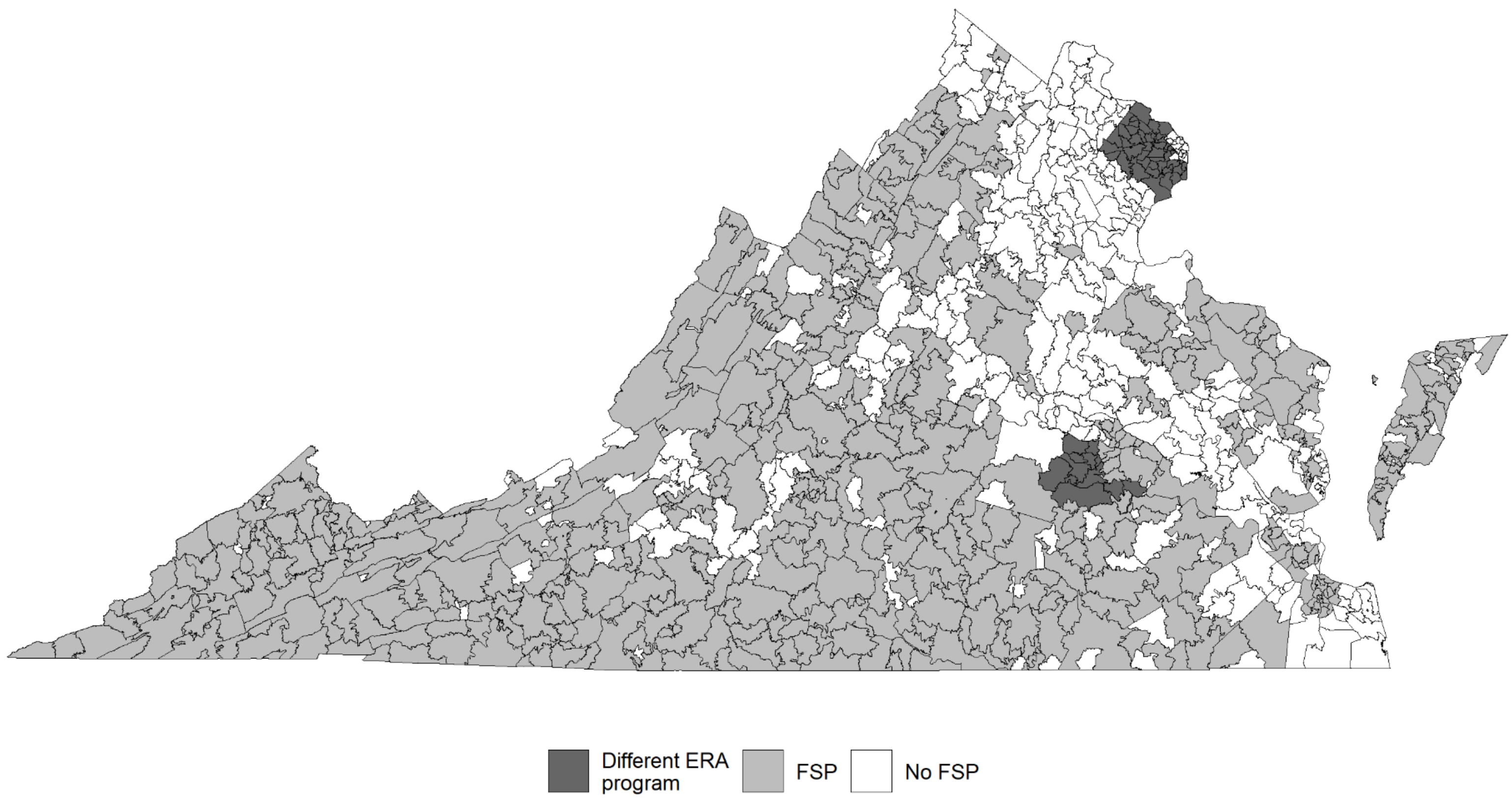


Figure 3. Geographic distribution of FSP and non-FSP eligible ZIP codes



All households with three or fewer members in 500 ZIP codes out of 820 eligible ZIP codes, as shown in Figure 3, were deemed presumptively income eligible for the Virginia Rent Relief Program managed by the VA DHCD. The 500 ZIP codes were selected since the median income for all households was less than the 2021 HUD statewide low-income limit for three-person households.[5]

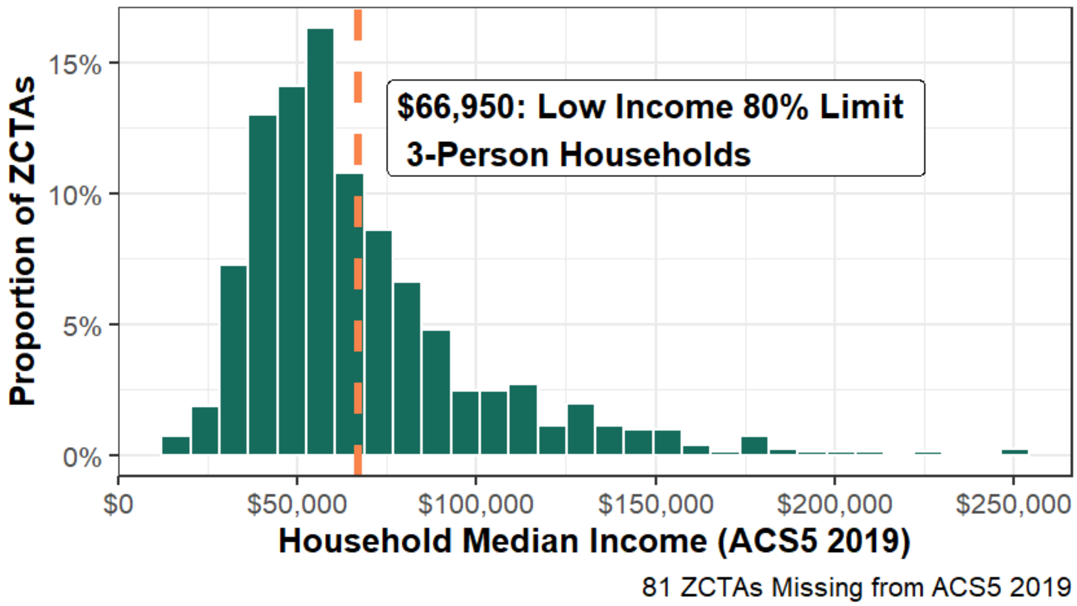
A critical piece of the quasi-experimental design is ensuring we understand how ZIP codes were assigned to FSP. OES was able to independently replicate this designation of the 500 ZIP codes using ZCTA 5 year estimates data from the 2019 American Community Survey and the statewide three-person low-income household limit from HUD. We replicated the list using the following steps:

1. Using the American Community Survey 2019 5 Year Estimates (table S1901), identify the household median income at the ZIP code level.
2. Take the HUD-defined statewide low-income limit for a household of three in 2021. This household size was used because most of the tenants served by VA DHCD to date are households of three or less.
3. Designate ZIP codes in which the median income is less than the statewide low-income limit as "FSP-eligible". This produces the 500 ZIP codes designated by VA DHCD in their public materials.

This replication exercise clarified that Census Bureau ZCTAs were used, rather than US Postal Service (USPS) ZIPs. Importantly, 81 ZCTAs were missing from the ACS5 2019 due to suppression of areas with low cell sizes, and those ZCTAs were thus excluded from the FSP list. Figure 4 shows the frequency distribution of ZCTAs by median income. The 2021 statewide low-income limit for three-person households is depicted using the vertical dashed line.[6]

[5] This was reported as $66,950 (see HUD website here for details).
[6] Throughout the remainder of this plan, we use the terms ZIP code to refer to these ZCTAs for readability. Although the phrase "ZIP codes" generally refers to USPS delivery routes (which are not geographic units), ZCTAs are areal representations of USPS ZIP codes created by the Census Bureau. As described in Project Description, FSP was implemented with the use of ZCTA-level data from the ACS.

Figure 4. Frequency distribution of ZIP codes by median income



Importantly, the implementation of the FSP did not simplify the income eligibility documentation requirements to the same degree for all potential applicants in areas where it was implemented. Only households with three or fewer members were allowed to forgo uploading additional income eligibility verification. In addition, not all of those potential applicants living in non-FSP ZIP codes faced the same requirement for documenting their income eligibility: applicants who already proved their income to qualify for a different means-tested benefit did not need to re-prove it for ERA, but were instead required to show proof that they received the other benefit (by, for example, uploading a certificate of TANF receipt).[7] Households with $0 in annual income or no bank account were required to upload a legal attestation to these facts, because they did not have income-related paperwork. We conceive of any tenant or landlord applicant who was required to upload a tenant income document, benefits certificate, or document attesting to zero income/no bank account as subject to providing full (and not simplified) income eligibility documentation . In our quasi-experimental analogy, every potential tenant or landlord applicant whose potential application would pertain to an otherwise-eligible household of size three or fewer in an FSP ZIP code was thus "treated." Our "control" potential applicants comprise the otherwise-eligible households of size three or fewer residing in non-FSP ZIP codes. Figure 2 shows the different kinds of documentation requirements faced by differently-situated applicants.

The above details on the program are presented as justification for the main assumptions justifying causal inference in the two quasi-experimental analytical approaches we use (we explain

[7] These included the Special Supplemental Nutrition Program for Women, Infants, and Children (WIC), Supplemental Nutrition Assistance Program (SNAP), Low Income Home Energy Assistance Program (LIHEAP), Temporary Assistance for Needy Families (TANF).

our reasoning for choosing these and not other similar approaches in detail in the appendix, see: Simulation study details).

The first is a regression discontinuity, in which the ZIP code's median income is the running variable and the statewide limit used to assign ZIP codes to FSP is the threshold. The key assumptions underlying this approach are that a) the number of applications in a given ZIP code can be expressed as a continuous function of the ZIP code's estimated median income, which is smooth at the point the FSP threshold is applied and b) ZIP codes were not sorted onto either side of the threshold using some characteristic other than their estimated median income. Based on the above, we have no reason to believe either assumption does not hold.

Hypotheses

The second analytical approach uses simple linear regression, adjusting for the estimated median income and pre-FSP number of applications to remove confounding. The core assumption underlying this approach, represented on a diagram in the appendix (Causal relationships between key variables), is that the only determinant of whether a ZIP code got FSP is where its estimated median income fell with respect to the HUD-defined statewide threshold. Again, we have no reason to believe this assumption does not hold.Our main research question is: to what extent does simplifying the requirement to individually document income eligibility from otherwise-eligible renter households of size three or fewer broaden and streamline access to assistance?

The time and administrative burdens involved in documenting income eligibility likely vary among applicants subject to this requirement: for example, finding the most up-to-date W2 or paystub, which would be the case of a potential tenant applicant falling into the lower-right of Figure 2, might be more time-consuming than calling the local SNAP office to obtain a certificate of benefits, or signing and then uploading a sworn statement that the applicant is paid in cash. In all cases, however, providing this documentation requires time and effort. Moreover, program staff must verify the recency, accuracy, and validity of the documentation provided. Therefore, we hypothesize that simplifying the requirement to document income eligibility for applications pertaining to three-person households in FSP ZIP codes will lead to more households of size three or fewer from those areas submitting an application to the program. We also hypothesize that the FSP led to faster application processing times for applications still pending when the administrator-facing changes were introduced. We describe how we construct these outcomes and how we identify the causal effects in which we are interested below.

While it seems possible that simplification could have little to no impact on the outcomes described, we think it is implausible that the intervention we study would have *decreased* applications or *increased* processing times. This belief is reflected in the rules we have for interpreting different patterns of results, described in the appendix (Interpretation of different patterns of main results).

Data and Data Structure

This section describes the data sources we plan to use to construct variables to be analyzed, as well as changes to be made to the raw data with respect to data structure and variables.

Data Sources and Transformations:

As described above, multiple vendors processed applications to VA's ERA program. Consequently, there are multiple data sources that need to be transformed and combined to create our final analytic sample. The primary data sources are:

1. **"Tenant Applications" data**: Dataset *(n* = 98,699) of application IDs and submission dates for one vendor, who processed tenant-initiated applications . These applications are the only ones for which we have submission dates, as the submission date is unavailable for all other applications. This dataset covers applications submitted from 9/29/20 to 12/1/2021, the last date at which an application was submitted to this vendor.
2. **"Application Status" data:** Dataset (*n* = 212,268) of most recent application status. This contains data on all applications, irrespective of who initiated them.
3. **Payments Data:** Dataset (*n* = 189,414) of payment results for applications that were fulfilled. Includes data on all applications, with payments made from 06/2020 to 09/2022 (the earlier payments represent the prior program in place funded by the State Housing Trust Fund).

The datasets include data on city, county, ZIP code, race, ethnicity, gender, disability status, veteran status, area median income buckets, household size, most recent application status (i.e. submitted/approved/paid), Fact Specific Proxy status, Categorical Eligibility status, payment amount (only in "**Payments Data"**) and submission dates for applications from a single vendor (only in **"Tenant Applications"** data). We note that these data only include submitted applications, and not applications that were initiated but never submitted. We undertake several steps to transform the various datasets and variables, described below.

First, we transform the raw application data and covariates into an application "universe" relevant for our analyses. Then, we collapse these applications into a ZIP-code level dataset to use in our main analyses.

The joining process proceeds in the following steps:

1. First, we join the **"Tenant Applications"** data (#1 in Data Sources) to **"Application Status"** data (#2 in Data Sources). All applications are found with statuses (though we find 278 duplicated statuses), so this procedure maintains a row count of 98,699. For our confirmatory analysis, we focus on these tenant-initiated applications due to the availability of submission dates. We only use data from the remaining applications (for which status data is available, but not submission date) for additional exploratory and descriptive analysis.

   a. At this step, we remove 278 applications with duplicate statuses, as described in more detail in **Data Exclusion** Below.
      i. Duplicates are identified by having identical values in the application_id column.
   b. All submission application IDs are found in the status data, so there is no missingness. This creates the **"Merged application date-status" data**.
2. Next, we add the **"Payments data"** (#3 in Data Sources**)** to the output of the previous step ("**Merged application date-status" data**). We do this in order to construct the processing time variable described below.
   a. Not all applications will be paid, so **"Payments data"** is only merged for relevant applications.
      i. We find 59,711 of our 98,699 applications are in the **"Payments data"**, and 38,988 are not.
   b. **"Payments data"** also includes covariates (though with different column names and coding styles – i.e. capitalization), so first we confirmed covariate values were the same between datasets.
   c. We merge on Application ID, Payment Status, Payment Amount, Payment Date Bucket, and Payment Date.
   d. Some applications appear multiple times in the **"Payments data"**, but none of these are applications tied to the vendor processing tenant applications. We take no further action on these duplicates, since all non-tenant initiated applications are dropped from our analyses.
   e. Finally, we join the "**Merged application date-status"** dataset to the **"Payments data"**. These 98,699 applications with status and payment data comprise our analytic sample **before** being collapsed to the ZIP-code level, as described in the next step.

Second, we collapse our application-level dataset into a ZIP code level ("wide format") dataset. For each ZIP code in the data, we create an observation in our dataset with the following column values:

- **Count of Applications Pre-FSP (`N_app_pre`):** Here, we count all applications from Step 1 of our data transformation which: 1) came from households of size three or fewer; 2) were submitted before the start of FSP on 6/10/2021 (non-inclusive), the date at which vendors received the list of eligible ZIP codes from VA DHCD and received clearance to communicate about the program to applicants. VA DHCD confirmed that the vendor communicated the change via email to applicants soon after this date Any application that was initiated but not completed is not included in the data available from VA DHCD.
- **Count of Applications Post-FSP (`N_app_post`)**: Similarly, we count all applications from households of size three or fewer submitted on or after 6/10/2021 . Throughout the remainder of this Analysis plan, we refer to these time periods as "Pre-FSP" and "Post-FSP."

- **Median Income (`med_inc`):** First, we merge the 2019 ACS 5-year estimates for median household income at the ZCTA level as described in [How the fact-specific proxy (FSP) worked]().
- **Running Variable (`running_variable`):** We transform the median income variable to be used as our running variable in the regression discontinuity design (described in more detail in [Statistical Models & Hypothesis Tests]()) in the following way:
  1. First, we take the difference between the cutoff threshold of $66,950 and the Zip-level median income estimate: `med_inc - 66950`. This centers the cutoff at 0, as is conventional in regression discontinuity designs to facilitate easier extrapolation to the boundary using polynomial regression.
  2. Second, we multiply this difference by -1 to account for the treatment being assigned to units **below**, rather than above, the threshold.
- **FSP (`fsp`)**: We code ZIP codes as FSP-eligible if the estimated median income is at or below than $66,950 and FSP-ineligible if it is greater than that threshold: `ifelse(med_inc <= 66950, 1, 0)`.
- **Demographic Counts**: Additionally, we sum the number of applicants with households of size three or fewer in each pre-post period that self-report belonging to specific demographic groups. We include all demographic measures in the data, using the same categories used in the datasets. Gender is included as measure in the Application Status data, but is notably not measured by the vendor for whom we have application submission dates. These variables are used in our Exploratory Analyses as described below in [Statistical Models & Hypothesis Tests](). A limitation on the construction and interpretation of this variable is that people who do not self-report their demographic information are not included in these count outcomes. We sum the number of
applicants who self-reported being in the following categories in both the pre- and post-FSP periods.
  1. **AMI Category** (`N_app_post_ami_0_30, N_app_post_ami_31_50, N_app_post_ami_51_80, N_app_post_ami_81_over`): At or below 30%, 31-50%, 51-80%, Over income threshold (80%)
  2. **Race** (`N_app_post_white, N_app_post_black_or_aa, N_app_post_asian, N_app_post_NHPI, N_app_post_AIAN, N_app_post_multi`): White, Black or African American,Asian, Native Hawaian or Other Pacific Islander, American-Indian or Alaskan-Native, or Multi-racial
  3. **Ethnicity** (`N_app_post_latinx, N_app_post_not_latinx`): Hispanic or Latino, Non-Hispanic or Latino
  4. **Gender** (`N_app_male, N_app_female, N_app_non_binary, N_app_trans, N_app_no_identify`) : Male, Female, Non-binary, Transgender, Do not identify as female, male, or transgender, Other
  5. **Veteran** (`N_app_post_veteran`): Applicant self-reports being a veteran
  6. **Disability** (`N_app_post_disability`): Applicant self-reports having a disability

- **Processing Time** (processing_time)**:** We also calculate average processing times per ZIP code for use in an exploratory analysis below. Note that this analysis is only run among applications that were submitted after 12/22/20 (when VA moved to the two-door model), and had not yet been processed by June 10, 2021 (the date at which vendors received the FSP ZIP code lists).
    1. To construct the ZIP-code level outcome, we take the difference (in days) between the processing date (from data source #2 above) and the Submission Date (from data source #1 above) at the individual level.
    2. Then, we winsorize the outcome using DescTools::Winsorize(), which replaces any values below the 5%-quantile or above the 95%-quantile of the distribution with those respective quantiles. We do this because there are errors in the data that produce negative processing times or processing times greater than a year, which create outliers that could skew the regression results. However, we winsorize the outcome only after dropping the small number of negative processing times identified earlier. We do this to avoid assigning quick processing times to invalid observations.
    3. Finally, we subset to households of size three or fewer and average these winsorized times to the ZIP code level, weighting all applications equally.
- **Total Payment Amount (total_paid_pre, total_paid_post):** For our exploratory analysis, we sum the total dollar amount paid to all applicants with households of size three or fewer in the pre- and post-FSP periods, summing all payments across payees.

Outcomes to Be Analyzed:

The primary outcome is the number of applicants with households of size three or fewer who submitted an application within a given ZIP code following the implementation of FSP (N_app_post).

Exploratory analyses focus on:

- Zip-code level number of applicants from underserved communities, constructed as described above under the bullet “Demographic Counts”
- Processing time for applications, among those who applied but whose applications were processed after the introduction of FSP by vendors on June 10, 2021, constructed as described above (processing_time)
- Total amount paid to renters in a given ZIP code (total_paid_post)

Data Exclusion:

Our main, confirmatory analyses focus on tenant-initiated applications processed by a single vendor, because these are the only ones for which we have application dates. We address this limitation in the robustness analyses by pooling all applications, and pooling the pre- and post-FSP outcomes (this does not pose a threat to our identification strategy, which is premised on the cross-sectional variation in Zip-code level median income).

Further, we drop any applications from Chesterfield or Fairfax counties, as these counties implemented their own separate ERA programs. Program administration communicated with these counties to ensure applicants were served by the appropriate ERA program based on their address, and that duplicate benefits were not received.

We identified duplicate applications in the "**Application Status**" dataset. Since these appear to be mistakes, we identify applications by `application_id` that appear more than once in the data. We verified that every instance of a duplicate application has one status of `SUBMITTED` in the `status` column and a second status with another value. We remove these duplicate rows by dropping rows in the dataset with a duplicated `application_id` that have a status other than `SUBMITTED`.

Finally, the ACS data used for FSP assignment does not report a median income estimate for 81 ZIP codes in Virginia. This likely occurs due to ACS data suppression, when the ACS declines to report certain estimates for privacy reasons, often in cases where the population counts are small. We do not treat these ZIP codes as part of our study, as they had zero probability of ever receiving FSP treatment. In our quasi-experimental analogy, it is as though these units were never part of the "experiment," since they were not assigned to treatment conditions with a probability between 0 and 1. This poses an external validity limitation: our results do not pertain to ZIP codes so small as to be subject to data suppression by the Census Bureau.

We do not plan to make further exclusions.

Treatment of Missing Data:

Our confirmatory analyses drop tenant-initiated applications from ZIP codes with missing ACS data as described above, because we do not consider them part of the quasi-experimental sample. By this definition of the sample, we have no missing data for our confirmatory analyses since we can observe: *all tenant-submitted applications between the period at which the two-door (one tenant and one landlord vendor) model was implemented and the day before the implementation of the tenant-facing FSP changes* (running from 12/22/2020 to 6/10/2021), *all tenant-submitted applications after FSP was implemented* (our primary outcome, running from 6/10/2021 until 12/1/2021), and *the median income used as our running variable* (from ACS).

The primary remaining sources of missing data come from our demographic variables used in our exploratory analyses, as some tenants declined to report their demographic information. Our approach is to restrict our analysis to self-reporters. A limitation to this approach is that it requires the assumption that FSP does not affect self-reporting of demographics. We have no reason to believe that it should. Furthermore, alternative approaches based on bounding and imputation would not necessarily improve the quality of our results.

Statistical Models & Hypothesis Tests

Our analyses fall into four categories: *confirmatory analysis*, our main results that will be the headline results in the abstract; *exploratory analysis*, which look at different outcomes that are policy relevant but not the central focus of the study; *descriptive analysis*, which attempts to describe trends in the program as a whole without inferring causality; and *robustness checks*, which are mainly intended to contextualize the confirmatory analyses by showing how the results change under different analytical choices.

Confirmatory Analyses:

We are interested in estimating *the causal effect of simplifying income eligibility documentation requirements on the number of otherwise eligible 3-or-fewer-person households who applied to Virginia's ERA program.*

Naive comparison of the number of applicants in ZIP codes that did and did not get FSP would likely be biased by an obvious confounding variable: for a given population size, as the proportion of low income households in a ZIP code increases, it has a higher likelihood of benefiting from FSP (because the median income will decrease) and, in all likelihood, it will have a higher number of applicants to the ERA program (because the size of the eligible population there increases).

We take two approaches to causal identification to address this threat:

1) a regression discontinuity design and,

2) linear regression with adjustment for confounders.

In both cases, we split the applications data into a pre-FSP period and post-FSP period (based on the June 10, 2021 date), and leverage our understanding of the FSP assignment mechanism described above (see How the fact-specific proxy (FSP) worked). Both confirmatory analyses also focus on applications from three-person or fewer households, as households of larger sizes were not eligible for FSP. We look at the overall effect on applications from all households in a robustness analysis.

Both approaches exhibited low bias in a simulation study described in the Appendix (see Simulation study details), where we allowed for multiple forms of confounding. However, the regression discontinuity approach (#1 above) performed slightly better in terms of bias but worse in terms of power. Moreover, each approach targets a different underlying causal effect (or estimand). Whereas the linear regression with adjustment for confounders (#2 above) attempts to understand the causal effect described in the underlined text above for every ZIP code in the sample, the regression discontinuity design (#1 above) limits itself to understanding this effect right at the threshold where a ZIP code would pass from eligible to ineligible for FSP.[8] We therefore include both approaches as main analyses, and describe in the appendix our decision rules for interpreting the (potentially divergent) results the two approaches may provide (see Interpretation of different patterns of main results). We expect positive results, so treat negative results as non informative unless they are very strong.

Regression Discontinuity Design: First, we use a regression discontinuity design (RDD), in which the estimated median income in each ZIP code is the running variable and the HUD-defined statewide low-income limit for three-person households is the threshold. The threshold used is a common, publicly available threshold used to determine eligibility for programs and we were able to fully replicate the assignment to FSP using publicly available data. We have no reason to believe that sorting to either side of the threshold occurred due to the design of the FSP. To implement the RDD, we use the rdrobust package for R, which uses local polynomial regression and an automated bandwidth selection algorithm to select data to be used in the estimation procedure. Our outcome is the post-FSP number of applicants in a given ZIP code. We base our inferences on the p-values estimated via the bias-corrected, robust standard errors that the rdrobust package estimates, as these provided the best coverage properties in the simulation study. We use the automated bandwidth selection approach implemented in the rdrobust package.[9] The code we plan to use is as follows:

```
with(data, rdrobust(y = N_app_post, x = running_variable, c = 0,
                    covs = N_app_pre, bwselect = "mserd"))
```

Adjustment for Confounders: Second, we use the fact that we can measure the precise variable (i.e. (the Census-estimated median income) that determined which ZIP codes did and did not get FSP to conduct a linear regression analysis in which we adjust our estimates to account for confounding. The directed acyclic graph (DAG) in the Appendix (see Causal relationships between key variables) shows that, according to our understanding of the causal relationships, we can derive unconfounded estimates of the causal effect of the FSP variable on the number of applications by conditioning on the estimate of median income and pre-FSP application numbers. Because the only parent of the FSP variable is the median income estimate, conditioning on the median income estimate is sufficient to ensure that there are no unblocked backdoor paths running from FSP to the post-FSP number of applications. We also condition on pre-FSP application counts, as these provide a pre-treatment estimate of the size of the program. The code we plan to use is as follows:

```
lm_robust(formula = N_app_post ~ fsp + N_app_pre + med_inc_est, data = data)
```

[8] In quasi-experimental approaches, these two different causal effects are often referred to as the "Average Treatment Effect" (ATE) and the "Local ATE" (LATE), respectively.
[9] More details on this procedure are available in Cattaneo, Matias D., Nicolás Idrobo, and Rocío Titiunik. *A practical introduction to regression discontinuity designs: Foundations.* Cambridge University Press, 2019. https://arxiv.org/pdf/1911.09511.pdf

Exploratory Analysis:

We conduct two main kinds of exploratory analyses. While they both answer important questions, we treat them as exploratory due to concerns about their reduced statistical power.

Processing times: The first exploratory analysis seeks to understand whether simplifying income eligibility documentation through the FSP reduced the processing time of FSP-eligible applications (from households in FSP ZIP codes and with three or fewer members). The rationale for this analysis is that the simplification of these documentation requirements sent to vendors in June 2021 (see Figure 1) reduces an administrative burden on DHCD VA administrators who no longer have to follow up with applicants to get correct / legible versions of bank statements, W2s, SNAP certificates, etc. Importantly, the outcome of interest (processing time) is only observed if a person applies. As such, we cannot simply compare the processing times of all applicants in FSP and non-FSP ZIP codes, as this would be subject to what is sometimes called "post-treatment bias." For example, suppose that FSP does reduce processing times for anyone who applies, but it also causes those with much more complicated cases (and longer processing times) to apply for the program. In that case, it is possible the processing time would be longer in the FSP group than in the non-FSP group, and the analysis would lead to the incorrect inference that the FSP increased processing time.

To avoid this bias, we must find a group whose processing time was feasibly affected by FSP, but whose probability of applying was not. Since, in conventional understandings of causality, an effect cannot precede its cause, and the plan to implement FSP (along with benefiting ZIP codes) was not publicized prior to its implementation, those who applied to the VA ERA program before FSP was implemented constitute such a group.[10] We construct the processing time outcome as described above (see Data Sources and Transformations). We analyze it using the same procedures and regression models described in the confirmatory analyses, with two modifications:

1. We subset to applicants with three or fewer household members whose applications were yet to be processed by the June 10, 2021 implementation of FSP
2. We do not include a control for "baseline" processing times, since no such control is available

Applications from Underserved Groups: The second exploratory analysis seeks to understand whether the FSP induced applicants belonging to historically underserved populations to apply to the program. Specifically, we focus on those with low incomes (between 0 and 30%, 31-50%, and 51-80% of AMI), those who identify as Black, Asian, Hawaiian/Pacific Islander, Native American, or multi-racial, those who identify as Hispanic or Latino, those who identify as women, transgender, or non-binary, veterans, and those who are disabled, as these are measured categories designated as "underserved groups" in the Executive Order on Advancing Racial Equity and Support for Underserved Communities Through the Federal Government and identified as underserved in

[10] See timeline in Figure 1 for further details

other OES work on ERA. When conducted in terms of relative proportions of the applicant pool, this analysis is subject to a similar challenge to the one described above: it is possible to estimate increases in the proportion of women applicants, for example, if FSP causes men not to apply. Thus, we focus principally on whether FSP had an impact on these groups, without trying to estimate, for example, whether this impact was statistically significantly larger than that for non-underserved groups.

We conduct analyses using the outcome variables described above (see Data Sources and Transformations), which count the number of applications in a given ZIP code that originate from a given group of interest. All demographic categories except for gender can be measured in the dataset for which we have submission dates, so for those categories we conduct the two confirmatory analyses as above, controlling for baseline application numbers. When analyzing applications by gender, we use landlord-initiated applications, which do not allow for the pre-post distinction due to the lack of application submission dates. In this analysis, we do not control for the baseline, and simply analyze the FSP vs. no-FSP difference.

Total amount paid. If the implementation of FSP worked to broaden access to VA's ERA program, it follows that it should have also increased the total amount paid in a given ZIP code. We suspect this outcome may exhibit a great deal of noise, given the many sources of variation from one ZIP code to another. However, we emphasize that it is not subject to the post-treatment bias issue which would arise if we sought to estimate the impact of FSP on the *average* payment amount to individuals (for example, if the program induced people with lower rental arrears to apply, the average payment amount in the treatment group would be lower than in the control group, leading to an incorrect inference that the FSP reduces the amount that people received). Here, the ZIP code is the unit of analysis, and we are simply interested in whether more money was spent overall in a ZIP code due to the broadening of program access brought about by FSP. We employ the same two analyses described in our confirmatory analysis.

As in our processing times analysis, we adopt June 10, 2021 (the date at which vendors received the FSP ZIP code lists) as our pre-FSP date. This is because it is plausible that vendors began processing applications and distributed money more quickly right after FSP ZIPs were shared.

Descriptive Analysis:

We are interested in describing program demographics across the entire sample of applications in the dataset constructed by using all the application status and payment data. This will enable us to contextualize the broader program but also the analytic sample used in our confirmatory analyses, Specifically, we plan to look at raw counts (and averages, where appropriate) and the distribution of the following variables - this will be helpful to describe what key attributes looked like across the sample in addition to key patterns that emerged across the different sets of applications and also to aid in inference across categories or applicants. We also intend to include statewide program averages for the measures that were reported directly to Treasury, as part of grantee reporting requirements.[11]

- Income levels (%)
  - <30% AMI
  - 31-50% AMI
  - 51-80% AMI
  - Above 81% AMI [Over-Income]
- Payment amount
- Household size
- Region type
  - City
  - County
- Gender
  - Male
  - Female
  - Non-binary
  - Transgender
  - Do not identify as female, male, or transgender
  - Other
- Race
  - White
  - Black or African-American
  - Asian
  - American-Indian or Alaska Native
  - Native Hawaiian or Other Pacific Islander
  - Multi-racial
  - Don't know/declined
- Ethnicity (%)
  - Hispanic or Latino
  - Non-hispanic or Latino
- Veteran
- Disabled

A limitation of this analysis is that it depends on self-reported demographics, which are not observed among people who decline to report. Above in the exploratory analysis of effects on demographic-specific counts, we imposed the assumption that simplifying income eligibility verification through FSP does not affect the probability of self-reporting one's demographics because if it does not hold then we cannot distinguish the effect on demographic-application counts from the effect on self-reporting. By contrast, our descriptive analysis is interested in estimating applicant demographics across our sample. These estimates rely on a different assumption than our causal exploratory analysis, namely that rates of self-reporting are the same among different demographic groups. If one demographic group reports at a much lower rate than the others, then when we estimate their relative proportions among reporters, that group will

[11] We report differences between eligible vs. recipient demographic estimates from the OES ERA descriptive study

appear less well-represented than it is in fact. We will note this limitation in the presentation of results.

Robustness Checks:

We plan to conduct four robustness checks of the two main confirmatory analyses.

Temporal Bandwidth. We will demonstrate how the point estimates and statistical significance changes as the temporal bandwidth around the threshold shrinks (e.g. three months following FSP implementation etc.).

Pooling Tenant Applications: We will perform an additional analysis (following the same procedure as our confirmatory analysis) comparing application counts between FSP and non-FSP ZIP codes overall between 2/22/2021 (when the Treasury guidance allowing the use of program flexibilities like FSP was released) and 12/1/2021 (when the two-door model was changed to a single vendor began processing all applications). We note that this is a conservative estimate, as it includes data on applications that were submitted before FSP was fully implemented.

Ignoring Household Size: We will conduct the confirmatory analyses using an outcome created by counting all applications, irrespective of household size.

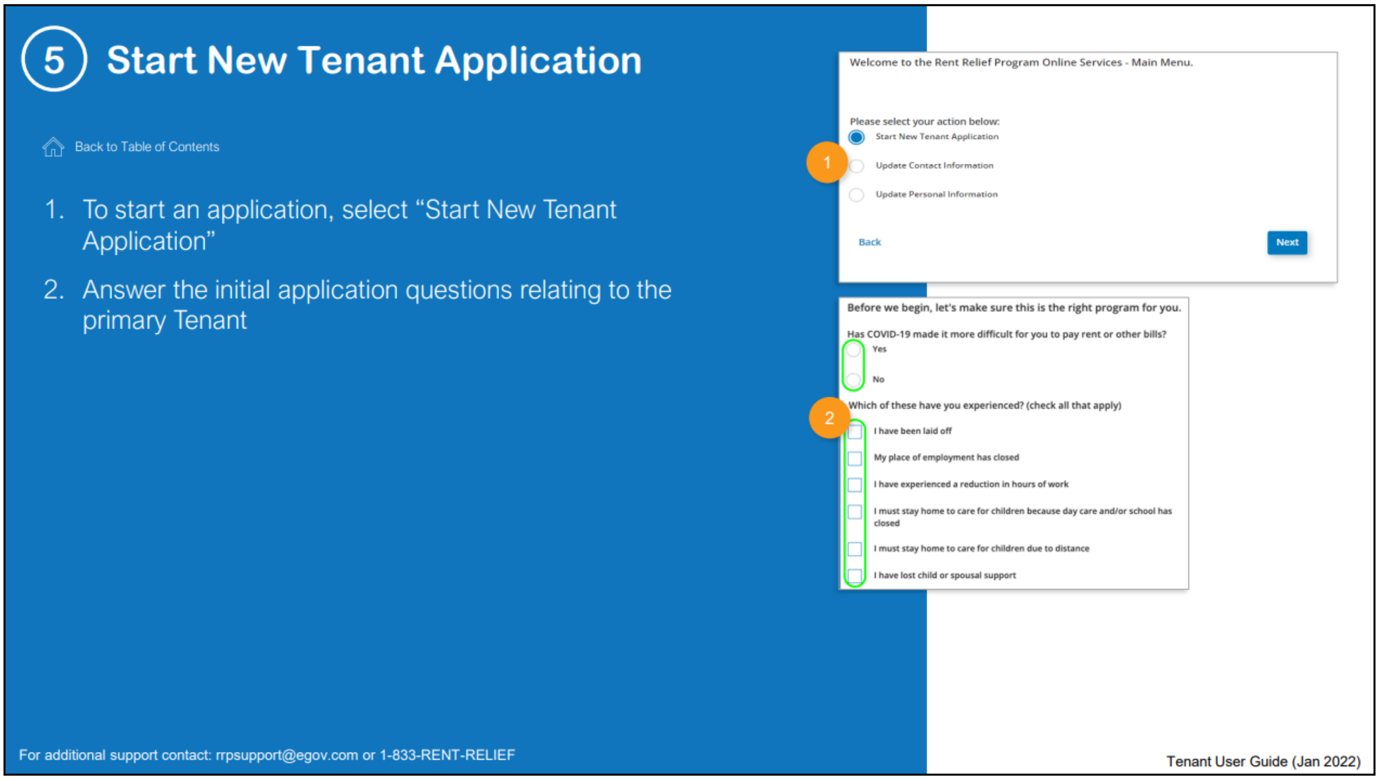
Full Data Inclusion. We will conduct the two exploratory analyses on a dataset constructed by using all the application status and payment data. Given that much of this data does not have application submission dates, we will not be able to distinguish between pre- and post-FSP applications. As a result of this data limitation, our primary outcome will be the total application count in each ZIP code and we will not be able to control for pre-FSP applications as in the main confirmatory analyses.

Inference Criteria, Including Any Adjustments for Multiple Comparisons:

We rely on standard errors estimated as described above in order to form p-values used in statistical significance tests. In all analyses, the null hypothesis is that the average effect of the treatment is zero and the test is two-tailed. We will use an alpha of 0.05 to determine statistical significance. To minimize the risk of false positives posed by multiple comparisons, we attempt to limit the number of tests in our confirmatory analyses to the bare minimum, selecting the two approaches to estimation that seem most appropriate (see Simulation study details and Supplementary Power Analysis in the appendix). We make no adjustments to the p-values or alpha level to account for multiple comparisons.

Appendix

Figure A1. Screenshot of Tenant Application[12]



[12] This screenshot shows how tenants simply self-reported experiencing housing instability and financial hardship on their application (available here)

Table A1. Eligibility criteria for ERA1 and ERA2

Type of criteria	ERA1	ERA2
Risk of housing instability or homelessness	One or more individuals within the household can demonstrate a risk of experiencing homelessness or housing instability	One or more individuals within the household can demonstrate a risk of experiencing homelessness or housing instability
Financial hardship related to COVID-19	One or more individuals within the household has qualified for unemployment benefits or experienced a reduction in household income, incurred significant costs, or experienced other financial hardship due directly or indirectly to the COVID-19 outbreak	One or more individuals within the household has qualified for unemployment benefits or has experienced a reduction in household income, incurred significant costs, or experienced other financial hardship during or due directly or indirectly to the coronavirus outbreak; and
Income below a threshold	The household has a household income at or below 80 percent of area median income	The household is a low-income family (as such term is defined in section 3(b) of the United States Housing Act of 1937 (42 U.S.C. 1437a(b)).[13]

Source: Department of the Treasury FAQs, Last updated on July 27, 2022

[13] The term "low-income families" means those families whose incomes do not exceed 80 per centum of the median income for the area, as determined by the Secretary with adjustments for smaller and larger families, except that the Secretary may establish income ceilings higher or lower than 80 per centum of the median for the area on the basis of the Secretary's findings that such variations are necessary because of prevailing levels of construction costs or unusually high or low family incomes.

Simulation study details

Given our understanding of the "treatment assignment" mechanism and the fact that we have both temporal and spatial variation in how the treatment was rolled out, there are many different approaches one could take to estimating the average treatment effect on the number of applications. However, we have a preference for choosing a minimal number of approaches, to reduce the risk of false positives due to multiple comparisons, and our confirmatory analyses above settle on using two approaches. We used simulation studies to decide which approaches to take, based on the principles of minimizing bias and maximizing statistical power.

Our primary simulation study was conducted using DeclareDesign in R with simulated data . We used a sample size of 800 ZIP codes split into two periods (pre- and post-treatment). The parameters and their statistical distributions build on the assumption of confounding, encoded in the directed acyclic graph below (see Causal relationships between key variables). The parameters were specified as follows:

- `N_linc` = the true number of low income households in a ZIP code, unobserved by the researcher. Drawn from a uniform distribution between 100 and 1000.
- `med_inc` = the true median income in a ZIP code, unobserved by the researcher. Drawn from a negative binomial distribution with mu = 150,000 and dispersion = 15. This variable is negatively correlated with N_linc at rho = -.5, to represent the idea that, all else being equal, a higher number of low-income households will lead to a lower median income.
- `N_linc_est` = the Census-estimated number of low-income households in a ZIP code, observed by the researcher. This is constructed by adding normally-distributed noise to N_linc, with mean 0 and standard deviation 100, to represent the idea of a noisy but unbiased estimation strategy.
- `med_inc_est` = the Census-estimated median income in a ZIP code, observed by the researcher. This is constructed by adding normally-distributed noise to med_inc, with mean 0 and standard deviation 100, to represent the idea of a noisy but unbiased estimation strategy.
- `N_app_post_fsp_0` = the control potential outcome, defined for all ZIP codes but only partially observable to the researcher (observed for ZIP codes that are not FSP-eligible, unobserved for ZIP codes that are FSP-eligible). This is calculated as a fixed proportion (10%) of N_linc. In other words, exactly 10% of all income-eligible people in a ZIP code apply at baseline. `N_app_post_fsp_0` can be read as "`N_app_post` when fsp is set to 0."
- `N_app_post_fsp_1` = the treatment potential outcome, defined for all ZIP codes but only partially observable to the researcher (observed for ZIP codes that are FSP-eligible, unobserved for ZIP codes that are *not* FSP-eligible). This is calculated as a fixed proportion (10.5%) of N_linc. In other words, FSP increases the proportion of low income people who apply to the program by half a percentage point. Note that this does not mean the treatment effect is homogeneous: if there are 200 low-income people in a ZIP code, then one additional person is induced to apply by implementing the FSP there, but if there are

1000 low income people, then FSP induces five additional people to apply. `N_app_post_fsp_1` can be read as "`N_app_post` when fsp is set to 1."

- `prop_app_post_fsp_0` and `prop_app_post_fsp_1` = these are alternative codings of the outcome we considered, whereby the proportion of low income people who apply to the program is the outcome variable, rather than the raw count. We construct this by dividing the two potential outcomes above by the observed, estimated number of low income people (since this is what the researcher has access to): `N_linc_est`.
- `N_app_pre` = the pre-FSP outcome. We construct this by subtracting normal noise with mean 5 and standard deviation 5 from the control potential outcome, N_app_post_fsp_0, and restrict the minimum to 0. In this context, the "parallel trends" assumption would imply that the pre-post change in the control potential outcome is the same for ZIPs that do and do not get FSP. The construction of the pre-treatment outcome by subtracting normal noise with mean 5 essentially means the trends are parallel for the *number* of applicants, since all ZIP codes increase by a constant amount of 5, but not for the *proportion of low income residents who apply*: an average increase of 5 will be smaller, proportionally, in places with more low income residents, who are also more likely to be in the treatment group because the median income there is lower.
- `prop_app_pre` = constructed by dividing N_app_pre by N_linc_est.
- `fsp` = the treatment indicator, 1 if med_inc_est is below the median of the estimated ZIP-code median incomes, 0 otherwise. This is not exactly how FSP was implemented, but it is a convenient shortcut, in that it ensures exactly half of the ZIP codes will be assigned to treatment on each simulation, and we do not have reason to believe this shortcut results in a loss of generality.
- `running_variable` = the transformation of `med_inc_est` we use as the running variable in the regression discontinuity. Calculated by subtracting the median of the estimated ZIP-code median incomes from med_inc_est, then multiplying by negative 1. This results in a variable that sorts all ZIP codes that do not benefit from FSP because their median income is too high to the left of 0, and ZIP codes that do receive FSP because their median income is sufficiently low to the right of 0.
- `N_app_post` and `prop_app_post` = the outcome observed by the researcher, which is a combination of the treatment and control potential outcomes of the ZIP codes, revealed according to the distribution of `fsp`.

We simulate the study 2000 times, randomly generating the variables above and re-estimating the coefficients, standard errors, and p-values from the different analytic approaches we consider (described in more detail below). We also calculate the true value of the estimand on each simulation (the true average treatment effect, or ATE), which we can use to compare to the estimated ATE to assess the potential bias, $E[\widehat{ATE} - ATE]$, for example.

Our simulation study yielded a number of helpful insights. In particular, it became clear that we should use the raw number of applications in a ZIP code as the outcome, rather than the estimated proportion of low income people who applied. Using the proportion was much less-powered as an approach, due to the noise in its construction. After some iteration, we narrowed down our candidate approaches to five different alternatives. These are listed in the rows of Table A3 below. In the columns, we list the performance of these different approaches. The approaches are implemented as follows:

1. **RDD** = a regression discontinuity using the rdrobust package for R, with the running variable constructed as above, and controlling for baseline (pre-FSP) number of applications in a ZIP code. We use the p-values and standard errors from the robust, bias-corrected variance estimators.
2. **Adjust for confounders** = this is a linear regression of the number of applications on an indicator for FSP, the baseline (pre-FSP) number of applications, and the estimated median income. See Causal relationships between key variables for a justification of this model. We use the heteroskedasticity robust standard error estimators implemented in estimatr:lm_robust().
3. **Adjust for baseline** = this is the same as the confounder-adjusted regression, but it does not include the control for the estimate of median income.
4. **Two-period DiD** = this is a two-period difference-in-differences regression, in which we stack the pre and post periods into a long-format dataset, and focus on the interaction between an indicator for a ZIP code ever getting FSP and an indicator for the post-treatment period. Standard errors are clustered at the ZIP-code level to account for autocorrelation between the two periods.
5. **Two-period DiD, adjusting for confounders** = this is the same as model 4, supplemented with an additional control for estimated median income.

Table A3. Results of the simulation study

	A	B	C	D	E	F	G
Approach	**True** ATE	**Expected** $\widehat{ATE}$	**Bias**	$SD(\widehat{ATE})$	**Expected** $\widehat{SD}(\widehat{ATE})$	**Coverage**	**Power**
1. RDD	2.75	2.78	0.03	1.34	1.28	0.94	0.59
2. Adjust for confounders	2.75	2.84	0.09	0.60	0.59	0.95	1.00
3. Adjust for baseline	2.75	3.24	0.49	0.37	0.37	0.74	1.00
4. Two-period DiD	2.75	3.07	0.32	0.36	0.36	0.84	1.00
5. Two-period DiD, adjusting for confounders	2.75	3.07	0.32	0.36	0.36	0.84	1.00

We compare the results column by column:

A. **True** ATE: This tells us the average value of the true, underlying average treatment effect, across simulations. All estimators target this same estimand, including the RDD. Usually the RDD only identifies the ATE at a hypothetical point located exactly at the threshold, but here, to facilitate comparison among estimation approaches, the local ATE and the ATE are the same by construction (see note on N_app_post_fsp_1 above). The true effect of FSP is specified to be an increase of 2.75 applications, averaging across ZIP codes.
B. **Expected** $\widehat{ATE}$: This is the estimate of the ATE that the respective approaches generate, typically estimated as a regression coefficient.
C. **Bias**: The average difference between A and B, across simulations, tells us the bias. The RDD and the regression adjusted for confounders are the least-biased approaches. There is some over-estimation of the true impact, on the order of about .03-.09 of an application, likely due to the fact that measurement error prevents us from perfectly removing the confounding (the positive association between the probability of getting FSP and the number of low-income applicants that is caused by the unobserved number of low income applicants). Adjusting for the baseline only (approach 3) is insufficient to reduce bias substantially. The difference-in-differences approaches are both biased, likely due to the higher degree of measurement error introduced by the estimation of an interaction coefficient (recall that parallel trends are satisfied here).
D. $SD(\widehat{ATE})$: This is the standard deviation of the estimated ATE, across simulations. In other words, it is akin to the true standard error that the standard error estimator is seeking to estimate. By this measure, the results provided by the RDD are twice as noisy as those provided by the adjust-for-confounders approach (1.34 vs. 0.60).
E. **Expected** $\widehat{SD}(\widehat{ATE})$: This is the expected estimate of the standard error, which is usually displayed in the regression table. We calculate this by taking the square root of the average of the squared standard error estimates across simulations. This can help to reveal bias in the estimates of the *variance*. Comparing to the "true standard error" in column D, all approaches get the standard error more or less correct.
F. **Coverage**: The coverage corresponds to the probability that the true estimate is contained inside the confidence interval, and it therefore reflects bias both in the estimate of the ATE (B) *and* in the estimate of the standard error (E). The nominal coverage probability for a 95% confidence interval is 95%. Only the first two approaches have coverage equal to 95%: the bias in the point estimates for approaches 3-5 leads to a coverage probability that is too small.
G. **Power**: Power is the true positive rate – the probability that a p-value will fall below the alpha confidence level when the null hypothesis is untrue. The RDD has the lowest power, which makes sense given its true standard error is also greater. There is simply more noise in this approach, due to the number of parameters estimated and the need to extrapolate to a threshold using a subset of the data close to that threshold. This is in part driven by the statistical distribution used to generate median incomes in the simulations – however we

used a negative binomial distribution in part because its skew and long right tail mirrors the empirical frequency distribution, as depicted on Figure 4.

In brief, the RDD provides low bias when either the ATE is constant or the estimand is the LATE, but it has low power. Adjusting for confounders is slightly more biased, but neither approach is strongly biased, and both should be consistent.

Supplementary Power Analysis

Since the main drawback to the RDD is its apparent lack of power, we supplemented the simulation study with additional power analyses that used the actual baseline data in combination with the *rdpower* package. We reasoned that if the minimum detectable effect, as best we could guess it, was below what we thought was plausible, then we would use an additional, better-powered estimation approach.

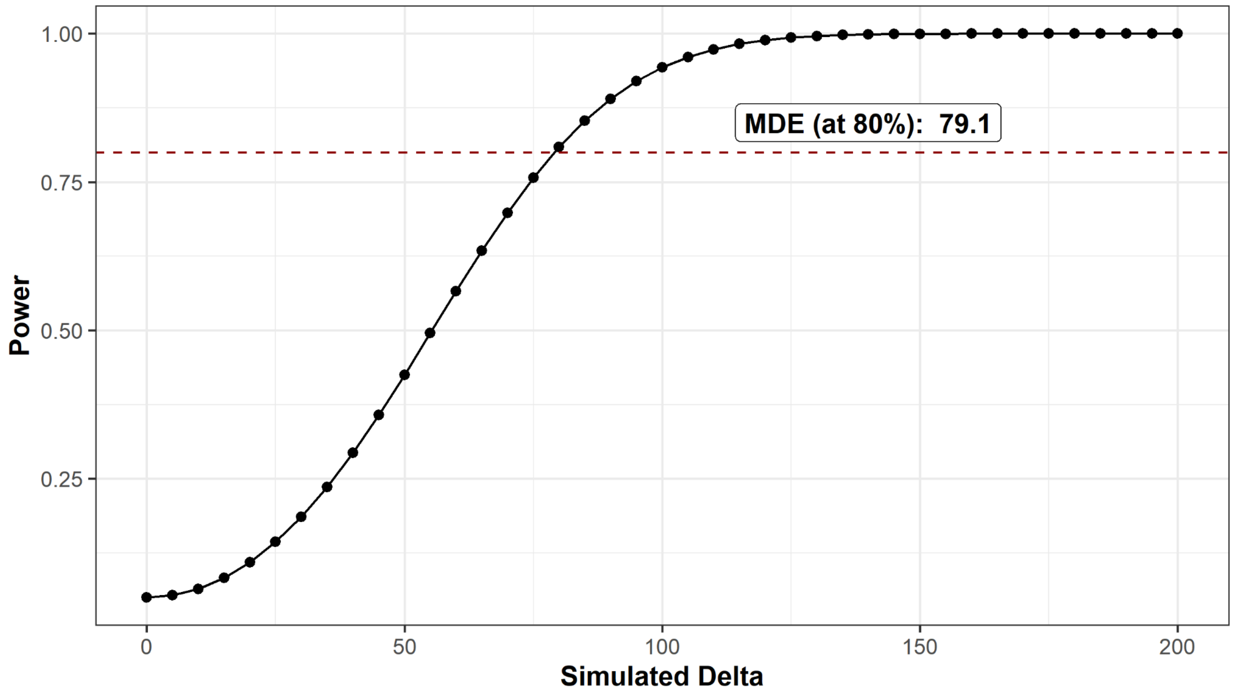
Using actual baseline data, we implemented the following procedure:

1. Calculate the number of application counts pre and post FSP implementation (using 5/28/2021 as the transition date) in our analytic sample.
2. Download ACS 2019 5 year estimate for the median income to use as a running variable.
3. Using the *rdpower* R package[14] , we calculate power for a series of potential effects between 0 and 200 applications, in intervals of 5. We use median income as the running variable, post-FSP applications as the outcome, and pre-FSP applications as a model control.
4. Finally, we separately calculate the minimum detectable effect (MDE) using the same variables as in step 3.

Results are visualized below in Figure A2. Our results suggest that an RDD alone would only be well-powered enough to detect effects as large as 70-80 additional applications during our sample time period to achieve 80% power. This power is primarily driven by the number of ZIP codes, since our unit of analysis is at the ZIP code level.

[14] Cattaneo, Matias D., Rocio Titiunik, and Gonzalo Vazquez-Bare. "Power calculations for regression-discontinuity designs." *The Stata Journal* 19, no. 1 (2019): 210-245.

Figure A2. Regression Discontinuity Power Analysis



Many effect sizes below the MDE strike us as plausible. Thus, there is a high risk of a false negative if we rely on the RDD alone, and we therefore use the two approaches that performed best in terms of bias in our simulation study: RDD and adjustment-for-confounders. Our plan for interpreting potentially divergent results is described below, in Interpretation of different patterns of main results.

Interpretation of different patterns of main results

Given our confirmatory analysis employs two different approaches to answering the same question, there is a possibility that the two sets of results may diverge. We plan to interpret differing patterns of evidence using the following decision rules, referring to the regression discontinuity design as the RDD and the adjust-for-confounders approach as the AFC:

1. We treat RDD as the less biased estimator (for the LATE or a constant ATE), and therefore weigh its evidence more highly than the AFC.
2. We treat results as convincing when they are statistically significant.
3. We treat results as more convincing when they are consistently signed across RDD and AFC.
4. We expect positive results, so treat negative results as non informative unless they are very strong.

Proposed interpretations:

Red indicates the signs or statistical significance do not agree across RDD and AFC, green indicates they agree.

	RDD		AFC		How we interpret this pattern of evidence	Description / justification of interpretation
	Sign	**Significance**	**Sign**	**Significance**		
1	-	Null	-	Null	We find no evidence of impact	RDD and AFC are both null so the interpretation is straightforward: We find no evidence of impact.
2	-	Null	+	Null	We find no evidence of impact	
3	+	Null	-	Null	We find no evidence of impact	
4	+	Null	+	Null	We find no evidence of impact	
5	-	Significant	-	Null	We find evidence of negative impact	We did not expect a negative result but we place more weight on the RDD, which was significant and negative. The signs agree, so we interpret this as evidence of negative impact.
6	-	Significant	+	Null	This evidence is suggestive of negative impact but inconclusive.	We did not expect a negative result but we place more weight on the RDD, which was significant and negative. The signs disagree however and our simulation study demonstrates AFC as likely better powered. So this is weak evidence of negative impact.
7	+	Significant	-	Null	This is suggestive evidence of positive impact.	We place more weight on the RDD, but the better-powered AFC was inconsistent. This is weak evidence of positive impact.
8	+	Significant	+	Null	This is evidence of positive impact.	We place more weight on the RDD, which is positive as expected and significant. The AFC is not significant but its sign is consistent.
9	-	Null	-	Significant	This is suggestive evidence of negative impact.	We did not expect a negative result, but the estimators are consistently signed and the better-powered estimate is significant.
10	-	Null	+	Significant	This is suggestive evidence of positive impact.	We place more weight on the RDD, which is negative and statistically insignificant. However, the better-powered estimator is in the expected direction and significant. Due to increased risk of bias and inconsistency in signs, we consider this suggestive evidence.
11	+	Null	-	Significant	The results are inconclusive.	The least biased estimator points in the expected direction but is not statistically significant, while the better-powered and potentially more biased estimator points in the unexpected direction. It seems unwise to draw any inferences from this pattern of results.
12	+	Null	+	Significant	This is evidence of positive impact.	The estimates both point in the expected direction, but only the better-powered estimator is statistically significant. This is evidence of positive impact.
13	-	Significant	-	Significant	This is strong evidence of negative impact.	Both estimates are negative and statistically significant.
14	-	Significant	+	Significant	The results are inconclusive.	These two patterns of results – with statistically significant estimates pointed in opposite

	RDD		AFC		How we interpret this pattern of evidence	Description / justification of interpretation
	Sign	Significance	Sign	Significance		
15	+	Significant	-	Significant	The results are inconclusive.	directions – suggest the inferences we draw about the program are highly sensitive to the estimator we choose. So much so that we should consider the findings inconclusive on the basis of these two results alone, and would have to do more analysis. This pattern of results seems very unlikely.
16	+	Significant	+	Significant	This is strong evidence of positive impact.	This is straightforward to interpret and justify.

Causal relationships between key variables

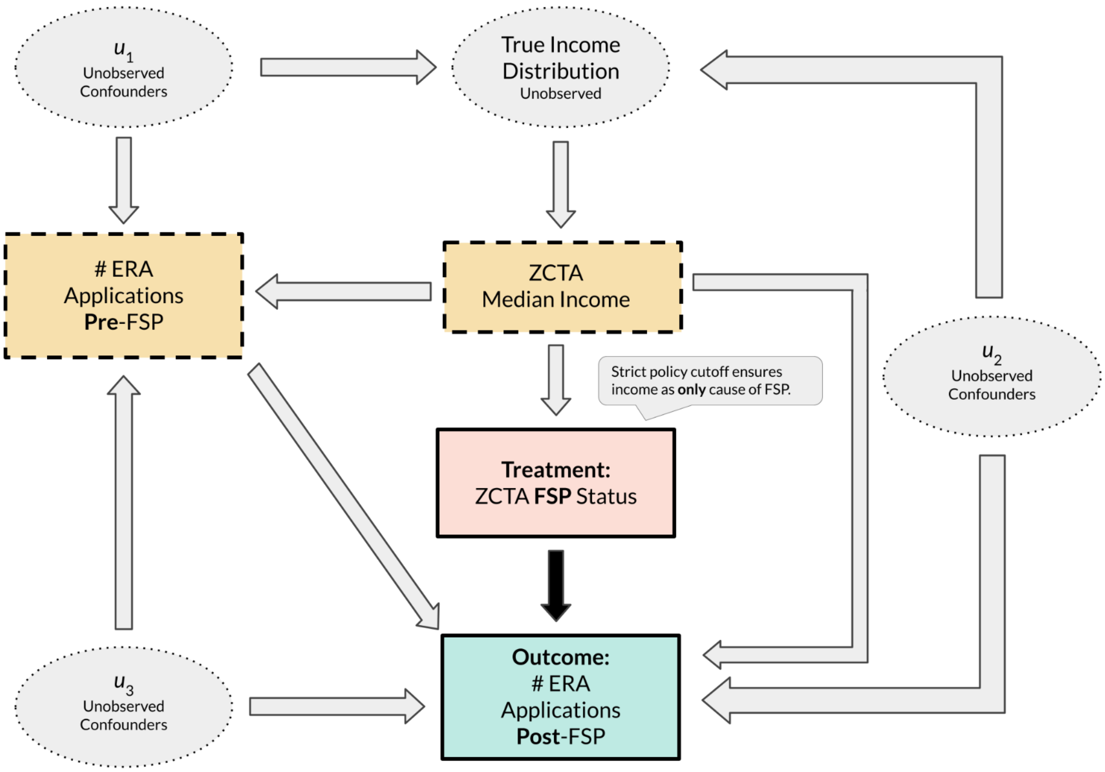
Our primary inferential goal is to estimate the causal effect of the simplification of income eligibility documentation requirements (FSP) on the number of program applicants (i.e. the Treatment → Outcome pathway). As discussed further in Statistical Models & Hypothesis Tests, we gain causal identification by leveraging our understanding of the strict policy cutoff that assigned FSP status **only** on the basis of median income.

This relationship can also be visualized in a Directed Acyclic Graph (DAG) as shown in Figure A3. The strict cutoff assigning FSP status ensures that, even in a situation with unobserved confounders that may influence our variables of interest (e.g. unobserved confounders that vary with both income and application counts), controlling for estimated ZIP-level median income allows us to estimate the causal effect of interest by satisfying the backdoor criterion.[15] Namely, we have a pre-treatment set of covariates (median income) which block every path (also called "d-separates", see Footnote 14 for more details) between our treatment (FSP) and outcome (application count) of interest that passes through the treatment.

The DAG demonstrates that, by conditioning on the median income estimate, we leave no unblocked backdoor paths between FSP assignment and the post-FSP number of applications. This known treatment assignment process also ensures there are no "bad controls" in this model (i.e. pre-treatment variables that induce a correlation between treatment and outcome through their inclusion as a control), because we can be confident that income is the only determinant of treatment status.

[15] Pearl, J. (2009). *Causality* (2nd ed.). Cambridge: Cambridge University Press. doi:10.1017/CBO9780511803161

Figure A3. Directed Acyclic Graph (DAG) of Causal Relationships



Note: Dotted circles indicate unobserved data, and dashed lines indicate model controls. Our primary causal effect of interest is visualized by the solid lines between Treatment and Outcome, and is identified by satisfying the backdoor criterion as discussed below.

A version of Figure A3 can also be visualized on DAGitty[16] with the following code, which can be pasted into the "Model Code" box on the DAGitty website to inspect the model:

```
dag {
bb="0,0,1,1"
"Income distribution" [latent,pos="0.504,0.086"]
"Median income" [adjusted,pos="0.514,0.176"]
"N Applications (Post)" [outcome,pos="0.556,0.327"]
"N Applications (Pre)" [adjusted,pos="0.269,0.332"]
FSP [exposure,pos="0.412,0.252"]
U1 [latent,pos="0.256,0.151"]
U2 [latent,pos="0.768,0.184"]
U3 [latent,pos="0.412,0.431"]
"Income distribution" -> "Median income"
"Median income" -> "N Applications (Post)"
"Median income" -> FSP
"N Applications (Pre)" -> "N Applications (Post)"
FSP -> "N Applications (Post)"
U1 -> "Income distribution"
```

[16] Johannes Textor, Benito van der Zander, Mark K. Gilthorpe, Maciej Liskiewicz, George T.H. Ellison. Robust causal inference using directed acyclic graphs: the R package 'dagitty'. *International Journal of Epidemiology* 45(6):1887-1894, 2016.

```
U1 -> "N Applications (Pre)"
U2 -> "Income distribution"
U2 -> "N Applications (Post)"
U3 -> "N Applications (Post)"
U3 -> "N Applications (Pre)"
}
```