

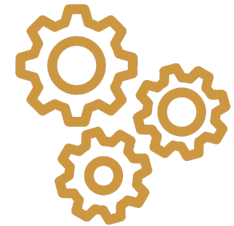
## Analysis Plan

Project Name: Decreasing homelessness and financial instability with unconditional cash transfers

Project Code: 2306

Date Finalized: 12/20/2024

Date Revised: 2/25/2025



Note: This is an updated version of a prior Analysis Plan which was posted on 12/20/24. The original Analysis Plan is available upon request by emailing [oes@gsa.gov](mailto:oes@gsa.gov).

---

## Project description

The Office of Evaluation Sciences (OES) at the U.S. General Services Administration is partnering with a county to understand the impact of its unconditional cash transfer program on financial wellbeing and housing stability, as measured through county-administered benefits records. The county used a weighted lottery to allocate cash transfers to vulnerable individuals and will follow those who entered the lottery to determine effects on their use of county-administered benefits after six months and one year.

The county designed and managed the cash payment evaluation. A non-profit human services agency worked with the county to administer the cash transfer program. OES provided technical assistance with the randomization protocol, and will conduct the analysis of program impact using administrative data.

## Randomization procedure

### *Randomization implementation*

The county categorized all 9,546 cash transfer applicants into one of three priority groups: individuals at-risk of homelessness ("homeless risk", Group 1), individuals with children or other dependents who are not at risk of homelessness ("families", Group 2), and individuals aged 55 or over who are not at risk of homelessness and do not have dependents ("seniors", Group 3). According to the program's priorities, the lottery was designed such that Group 1 had a 50% higher probability of receiving cash than Group 2, who themselves had a 50% higher probability of receiving cash than Group 3. An additional program requirement was that no fewer than 2,250 individuals receive cash transfers. Therefore, OES and the county designed a randomization procedure to ensure a specific number of applicants would receive cash transfers, subject to the constraint that the proportion of Group 1 who received an initial offer of cash transfers would be higher than Group 2, and that the proportion in Group 2 would be higher than in Group 3.

Applicants were randomized as follows:

- (1) Calculate how many people should be selected from each of the three priority groups for cash transfers such that Group 1 has 50% higher probability of being selected than Group 2, and Group 2 has a 50% higher probability of being selected than Group 3, as follows:

$$\circ \frac{m_1}{n_1} = p * \frac{m_2}{n_2}$$

$$\circ \frac{m_2}{n_2} = p * \frac{m_3}{n_3}$$

$$\circ m_3 = M - m_1 - m_2$$

Where  $M$  is the total number of people who are going to be selected for cash and  $m_1$ , for example, refers to the number of people in Group 1 selected for cash, so  $M = m_1 + m_2 + m_3$ .  $N$  is the total sample size and  $n_1$ , for example, is the number of people in the sample in Group 1, such that  $N = n_1 + n_2 + n_3$ .

- (2) Set the prioritization factor ( $p$ ) to 1.5.

- (3) Calculate the number of people who should be selected from each priority group:

$$\circ \text{Group 3: } m_3 = M / \left( \frac{1 + p * p * n_1 + p * n_2}{n_3} \right)$$

$$\circ \text{Group 2: } m_2 = \left( p * \frac{m_3}{n_3} \right) * n_2$$

$$\circ \text{Group 1: } m_1 = M - m_2 - m_3$$

The county created a waitlist condition in addition to the offered cash condition in case individuals who were offered the cash transfer were found to be ineligible, unreachable, or disinterested in receiving the cash. In principle, by creating such a waitlist, the county avoids running the lottery multiple times. The county created a waitlist whose size was equal to 30% of those offered cash. Thus, the randomization procedure assigned individual applicants within each randomization priority group (Group 1 - 3) to each of three conditions: cash (2,250), waitlist (675), and no cash (6,621).

In practice, too few individuals in both the cash and waitlist conditions were delivered the cash transfer (due to ineligibility, unreachability, and refusal), and it was not possible to reach the requisite number of people for cash transfers (2,250) using the original cash and waitlist conditions. The county therefore conducted a second lottery among the 6,621 individuals who were originally assigned to not receive cash transfers. We refer to the two lotteries as “wave 1” and “wave 2”. In wave 2, 570 individuals were initially offered cash transfers, as 570 individuals remained to reach the target allocation of 2,250. The waitlist condition comprised 331 individuals

in the second wave. The county determined the wave 2 waitlist condition size by taking the wave 1 lowest take-up rate by priority group, assuming that this rate will hold for all priority groups, and factoring in the potential for the take-up rate to be 5% below the first wave take-up rate.

Table 1 below shows the distribution of individuals under the randomization procedure in waves 1 and 2.

**Table 1.** Cash transfer program allocated lottery priority group and condition sizes

| Wave 1           | Cash  | Waitlist | No Cash |          |         |
|------------------|-------|----------|---------|----------|---------|
| 1) Homeless risk | 185   | 56       | 266     |          |         |
| 2) Families      | 1,798 | 539      | 5,056   |          |         |
| 3) Seniors       | 267   | 80       | 1,299   |          |         |
| Total:           | 2,250 | 675      | 6,621   |          |         |
| Wave 2           |       |          | Cash    | Waitlist | No Cash |
| 1) Homeless risk |       |          | 46      | 27       | 193     |
| 2) Families      |       |          | 456     | 265      | 4,335   |
| 3) Seniors       |       |          | 68      | 39       | 1,192   |
| Total allocated: |       |          | 570     | 331      | 5,720   |
| Total offered:   |       |          | 570     | 273      | 5,778   |

#### *Implications for cash and no cash condition*

Individuals in the waitlist condition were offered cash transfers in a randomly determined order. Since all individuals in the first waitlist condition were eventually offered cash transfers or determined to be ineligible, in practice there is no analytic distinction between the cash and waitlist condition in the first randomization wave. In the second randomization wave, 273 of the 331 individuals in the waitlist condition were offered cash transfers. The remaining 58 individuals in the waitlist condition were never offered the cash transfers, and they effectively remained in a condition comparable to the no cash condition.

For the purposes of the analysis, we refer to those that were offered cash, whether or not they received it, as the “offered cash” or “cash” condition. We refer to those that actually received the cash as the “received cash” condition. The received cash condition is a subset of the offered cash condition. Those that were not offered cash are referred to as the “no cash” condition. Taken together across the two waves, there are 3,768 individuals in the offered cash condition in the whole study ( $2,250 + 675 + 570 + 273 = 3,768$ ) and 5,778 in the no cash condition.

We will account for the differential probabilities of assignment to the offered cash condition by inversely weighting individuals according to the actual proportion of individuals in each priority group who ended up being offered cash. In addition to accounting for condition-level differences, the inverse propensity weights (IPWs) will account for the re-randomization of individuals in the wave 1 no cash condition in wave 2 (see [transformation of variables](#)).

#### *Implications for cash transfer timing*

Eligibility verification of individuals was conducted in cohorts on a rolling basis, meaning that cash condition individuals did not receive their cash offers or transfers all at the same time. The list of individuals assigned to get cash transfers and waiting to be processed for eligibility verification was randomly ordered. Thus, whether individuals were in the first or second randomization wave will only be meaningful insofar as it affects when their eligibility verification was conducted, and so when they received cash. As the timing of offers of cash transfers was randomized across individuals, we use difference-in-differences estimation to understand the impact of the cash over a period following its offer and receipt (see [statistical models](#)).

#### *Implications for compliance*

In the analysis of randomized studies, when individuals are selected at random to receive a given intervention but do not receive it, this is referred to as “noncompliance.”<sup>1</sup> In this study, noncompliance arises because individuals selected to receive cash transfers were later found to be ineligible, did not reply to the communications from the county, or refused to receive the cash. The rates of noncompliance differ for the three priority groups across the two waves: individuals at risk of homelessness had a noncompliance rate of  $103 / 308 = 33\%$ , families of  $1202 / 3012 = 40\%$ , and seniors of  $220 / 448 = 49\%$ . We will estimate the effect of receipt of cash among compliers using the methodology described in the [statistical models](#) section.

## **Estimands and hypotheses**

#### *Background context*

To better understand the population targeted by the cash intervention and their potential impacts, OES and the county interviewed program staff at several organizations in the county: two non-profits, one which implements the cash transfer program and another with expertise in homeless assistance, and two county government entities, one with expertise in foster care and another that operates multiple county-administered benefits programs.

The general outcome that interests us and our agency partner is county-administered benefits use. Specifically, in our scoping work and the tests we detail below, we aim to understand how a one-time influx of cash might affect the uptake and ongoing use of different county-administered benefits. From the perspective of our agency partner, it is useful to understand whether the cash

---

<sup>1</sup> OES is using the term “noncompliance” in accordance with its meaning in the statistical literature. In this context, noncompliance refers to an instance where a participant selected for a trial or experiment does not ultimately satisfy the elements of the intervention they are selected or randomized for. This is distinct from the meaning of the term “noncompliance” in the context of federal grant management, in which a recipient of federal funds does not satisfy the requirements of federal rules, guidance, or award terms and conditions. In the case discussed here, OES is not aware of the County making any payments or taking other actions that fail to comply with Treasury rules, guidance, or award terms and conditions.

*reduces* benefits use by increasing an individuals' self-sufficiency.

Therefore, our scoping work focused on understanding whether or not this substitution effect would occur, and if so, for whom. Based on these conversations, we believe that the cash payments are most likely to reduce benefits use among individuals who are on the margin of stability. Individuals on the margin of stability may need to use county-administered benefits in some months but do not use county-administered benefits in all months.

Those we spoke to expressed the expectation that, for individuals experiencing entrenched forms of disadvantage and instability, and for whom various benefits constitute an important and ongoing source of support, a one-time payment of \$4000 is unlikely to remove the need for such benefits. Moreover, county officials took steps to ensure that the cash payment would not be included in the income calculations used to determine benefit eligibility (i.e., for the four benefits that makeup the outcome in this study), so there is no reason to expect that individuals would no longer be eligible for benefits due to the temporary increase in their monthly income from the transfer.

However, for individuals on the margin of stabilization this payment could make a difference that is measurable in county-administered benefits use. For such individuals, a large one-time cash transfer could mean that they have available cash that they can use when they experience challenges. Based on our scoping work, we propose a causal mechanism through which a one-time cash transfer could decrease benefits use for those on the margin of stability that is specific to the benefit in question:

- Supplemental Nutrition Assistance Program (SNAP): For families on the margin of stability, a one-time influx of cash would enable them to pay for larger expenses (e.g., health care or childcare) that would otherwise be paid for out of their regular income, freeing up regular income to pay for food.
- Temporary Assistance for Needy Families (TANF): For families on the margin of stability, a one-time influx of cash would enable them to cover expenses that are a primary barrier to employment (e.g., a car repair), enabling them to retain, obtain, or increase employment when they otherwise would not be able to.
- Foster Care Placement: For families on the margin of stability, a one-time influx of cash would allow them to cover expenses that would otherwise destabilize and stress their household and prevent them from caring for their children (e.g., childcare, children's clothing or out-of-pocket healthcare expenses).
- Homelessness Assistance: For families on the margin of stability, a one-time influx of cash would allow them to cover expenses that would otherwise jeopardize the stability of their housing (e.g., back-owed rent, expenses threatening continuous employment).

We define marginal individuals as those who either stopped or started using a given benefit in one or more of the six months prior to the first cash offer. For example, if 10% of the sample either started or stopped using foster care assistance within six months prior to the first cash offer, we consider these 955 individuals ( $0.10 \times 9,546$ ) to be marginal for foster care assistance. We define marginality for each benefit and as such individuals may be marginal for more than one benefit if

they stopped or started using more than one benefit within six months before the study. We will aggregate up the benefit-level marginal samples to get the total marginal sample. The total marginal sample will thus be the number of individuals who are marginal for any benefit (i.e., at least one).

A concern with the analysis of program effects among marginal individuals is that we do not know in advance how many individuals switch benefits in a 6-month period. If few individuals switch benefits we may end up with a small sample and an underpowered analysis (i.e., one with a high risk of false negatives). To mitigate the risk that our procedure for defining marginality will lead to an underpowered analysis, we outline [below](#) a decision rule under which we will broaden our definition of marginality in order to ensure a large enough sample size for a sufficiently powered analysis.

To our knowledge, the evidence thus far on the effects of cash transfers on financial well-being in developed countries has been mixed, although none of it focused on the subpopulation of those on the margin of stabilization.<sup>2</sup> Below we identify the specific primary and exploratory hypotheses we formed based on these scoping conversations.

#### *Definition of estimands*

At the highest level, we wish to estimate the average causal effect of receiving the unconditional cash transfer on individuals' use of benefits during specific periods of time after the offer of cash is first made.

The unit of observation/analysis is a person-month. Importantly, we expect timing of cash transfers to matter: we do not expect that the outcomes of an individual will be the same in June if they received the cash in March versus in May of that same year (because individuals get the cash at different times). This implies that there are a larger number of potential outcomes to consider and thus to summarize over in order to define our estimand (Callaway and Sant'Anna, 2021). Practically, this will inform how we choose to average over that heterogeneity. As discussed below, traditional fixed effects regressions do not average the potential outcomes evenly and can result in bias because of how individuals are weighted, and we will aim to address this problem using alternative specifications.

To understand the effect of receiving the cash transfer, we focus on individuals in our sample whose cash transfer outcomes we do observe or could have observed. Those individuals are the "compliers": people who would receive the cash if offered it.

Our main estimand is defined as the average difference between the observed outcome of those who received cash and the outcome that would have been observed had they never been assigned to receive cash, over a 6-month period following their receipt of an offer of cash. We refer to this as the "6-month complier average treatment effect on the treated", or "6-month CATT". As explained below ([transformations of variables](#)), this 6-month period does not necessarily follow

---

<sup>2</sup> See for example Liebman et al. (2022) who find no significant positive effects of cash transfers on food insecurity after two years; Jaroszewicz et al. (2024) who find no significant positive on effects of a cash transfer on subjective financial wellbeing; Bartik et al. (2024) who find no effect on net worth, credit limits, delinquencies, utilization, bankruptcies, or foreclosures; and Dwyer et al. (2023) who do find significant positive effects of a cash transfer on homelessness.

immediately after the offer of cash, because individuals did not receive cash immediately after the offer. In fact, we expect there to be a delay in the receipt of cash; we therefore estimate impacts of the transfer after the modal delay between the offer and receipt. For example, if most people who received cash got it two months after the offer, then the modal delay is two months. Our estimand is defined over the 6-month period following this modal delay.

Through simulation studies, we established that the CATT is estimable using the methods we describe below. We are also interested in the effect of receiving an *offer* of cash. This effect is more straightforward to estimate because it is the *offer* and not the *receipt* of cash that is the main source of random variation we can causally identify via the lottery. Further, the estimate of the offer of cash is likely to be more precise. We refer to this effect as the “6-month intent-to-treat effect”, or “6-month ITT”.

For both estimands — the 6-month CATT and ITT — we are also interested in understanding their magnitude among the sub-population of individuals defined as on the margin of stability (using procedures described below). We refer to those sub-population-specific effects as the conditional CATT and ITT.

#### *Primary hypotheses*

We estimate county-administered benefits use six months (plus modal delay) after the offer of cash among compliers and all individuals offered cash (the CATT and ITT). We also test these two hypotheses for those on the margin of stability (the conditional CATT and ITT).

- **Hypothesis 1 (6-Month CATT)**

Compared to never receiving cash, the **receipt** of cash will cause a decrease in the amount an individual uses county-administered benefits in the 6-month period following a modal delay after an offer of cash.

- **Hypothesis 2 (6-Month ITT)**

Compared to never being **offered** cash, the offer of cash will cause a decrease in the amount an individual uses county-administered benefits in the 6-month period following a modal delay after an offer of cash.

- **Hypothesis 3 (Conditional 6-Month CATT)**

Among individuals defined as at the margins of stability, compared to never **receiving** cash, the receipt of cash will cause a decrease in the amount an individual uses county-administered benefits in the 6-month period following a modal delay after an offer of cash.

- **Hypothesis 4 (Conditional 6-Month ITT)**

Among individuals defined as at the margins of stability, compared to never being **offered** cash, the offer of cash will cause a decrease in the amount an individual uses county-administered benefits in the 6-month period following a modal delay after an offer of cash.

## 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.

### Data source(s):

#### *Administrative data*

We will focus on individuals' use of the major benefits program in this analysis: state-level Supplemental Nutrition Assistance Program (SNAP), state-level Temporary Assistance for Needy Families (TANF) program, foster care placement, and homelessness assistance. Our primary data source is a state-level public benefits database. In addition, we will obtain data on homelessness assistance from the [Homeless Management Information System](#) (HMIS) database. For legal reasons, medical benefits will not be available in the data we access. We are requesting the below data for the period one year prior to any study individuals' offer of cash transfers until one year (plus the modal delay) after any study individuals' offer of cash transfers (i.e., December 2022 - December 2024).

- Demographics and other indicators, including:
  - Marital status
  - Household income
  - Hourly wage
  - Employment status
  - Whether they have a registered car
  - Household size
  - Number of children in the home
- Monthly county-administered benefits, including:
  - State-level SNAP
    - Application (submitted/approved/denied)
    - Household allocation (in dollars)
    - Household use (in dollars)
  - State-level TANF
    - Application (submitted/approved/denied)
    - Household allocation (in dollars)
    - Household use (in dollars)
  - Foster care placement by child in the household
    - Date of removal (from household)
    - Reason for removal (from household)
    - Date petition was filed that led to removal
    - Date of detention hearing
    - Date child was declared a dependent
    - Funding prior to detention hearing
    - Date of change in placement
    - Reason for change of placement



- Date of termination of placement
  - Reason for termination of placement
- Homelessness assistance
  - Number of days per month using temporary residential services (e.g., homeless shelter)
  - Number of days per month using temporary non-residential services (e.g., shower)
  - Number of days per month using permanent residential services (e.g., rent assistance)

#### *Program data*

- Date that study recipients were offered cash
- Date of distribution of cash transfers

In addition, applicants to the cash transfer program completed a comprehensive application which included the following variables:

- Demographics including variables such as:
  - Education level
  - Household composition
- Financial wellbeing including such variables as:
  - Household income
  - Paycheck frequency
  - Risk of loss of home
  - Use of county-administered benefits
- Self-stated impression/satisfaction scores, including:
  - Impression from their first interaction with the cash transfer program
  - Satisfaction with life
  - General physical wellbeing

Note that the county does not have a unique ID that identifies individuals in both the application data and the county-administered benefits data. The county's data officers matched the data on cash transfer applicants to its historical data on county-administered benefit usage. They conducted probabilistic matches based on a combination of name, age, address, email, and phone number. Matches were given a score. For example, a match based on full name / age range (+- 3 years) / full address / phone / email got a score of 100. A match based on full name / age range (+- 3 years) / partial address / phone got a score of 90, and a match based on phonetic sounding of name / age range (+- 3 years) / partial address got a score of 75. Based on the county's analysis of the accuracy of the match scores, we consider all matches with a score of 75 or more to be a true match. To the extent that there are errors in the matching process, we do not expect this to bias our results since they are not expected to be correlated with treatment status, but will decrease our precision.

## Outcomes to be analyzed:

Our outcome is individuals' total monthly county-administered benefits use, measured using data on county-administered benefits used by the individuals in the lottery. We construct the outcome from the following underlying data: for homeless assistance benefits, this will be the number of days per month that services were used; for state-level TANF and SNAP, this will be the dollar amount used per month; and for foster care, this will be a count for the number of children in the household of a study individual that are in foster care in that month. From these data, we will construct binary indicators of benefit use for each of the four benefits. Our outcome measure is individuals' total monthly benefits, constructed by summing across the four binary variables.

For the conditional CATT and ITT hypotheses, the outcome is the count of benefits used in a given month that an individual is defined as marginal for. For example, if a person is marginal for food and homelessness assistance and they get both in February, their outcome would be two in that month. If they only receive food assistance in that month their outcome would be one, and if they got food assistance and foster care assistance – an outcome they are not marginal for – their outcome would still be one.

Since our primary specifications use difference-in-differences estimation, the outcome is interpreted as the change in total benefits use for an individual in the months after the offer (or receipt) of cash transfers compared to before the offer (or receipt) of cash, compared to this difference for those not offered cash.

## Transformations of variables:

### *Construction of panel data*

We will merge the program data on offer and receipt of the cash transfer and application data, with the monthly administrative data. The resulting dataset will be structured as a panel, with one observation for every person included in the lottery for every month in the year preceding and the year following the lottery.

As individuals are likely to vary in how long they take to respond to the offer of cash (e.g., to locate the necessary document that proves eligibility), we will ascertain the modal time between offer and receipt and “start the clock” as of this date. For example, if the modal individual begins receiving cash in the second month after being offered cash, we will define the 6-month CATT and ITT over the period spanning from month two to month seven. We will drop from our panel the months of data after the offer and prior to the modal receipt time.

From the program data we will construct the following indicators for random assignment and receipt of cash:

- A dichotomous indicator equal to one if the individual was offered cash in that month or any preceding month and zero otherwise.
- A dichotomous indicator equal to one if the individual received cash in that month or in any preceding month and zero otherwise.
- An integer variable corresponding to the month in which an individual was offered cash, coded as zero for individuals who were never offered cash.

### *Construction of weights*

We will account for individuals having different probabilities of assignment to receive cash by priority group. The probability of assignment to the offered cash condition (cash or waitlist) in either round is 39% (3,768 / 9,546). This probability differed for individuals based on what priority group they were in prior to the start of the evaluation. The overall probability of assignment to cash was:

- 61% of individuals at-risk of homelessness (Group 1) N = 308 / 507;
- 41% of individuals who are in families (Group 2) N = 3012 / 7,393; and
- 27% of individuals who are seniors (Group 3) N = 448 / 1,646.

Our analysis will use IPWs to account for the differential probability of assignment to the cash condition, which varies by priority group number.

### **Data exclusion:**

No data on individuals who were included in the lottery will be systematically excluded from the sample. Any benefits or other data that does not pertain to individuals included in the lottery will be excluded.

### **Treatment of missing data:**

While our analysis assumes that anyone who does not match between the randomization file and the county-administered benefits files does not use county-administered benefits, it is possible that a mismatch occurs because of differences in data entry procedures. Because there is no reason to expect that the match rate between baseline data and county-administered benefit files is correlated with random assignment to the cash condition, we take no particular steps to address missingness related to data entry discrepancies. We will, however, check for differences between self-stated benefits use in the application data and the county's match rate to gauge the potential accuracy of the matches.

We cannot rule out the possibility of a correlation between random assignment and missingness from the outcomes data for other reasons. In particular, if an individual moves out of the county and uses public benefits in another county, they will be erroneously coded as not using benefits. We refer to this as a "false negative" for benefits use. To the extent that random assignment to the no cash condition makes moving more (or less) likely, we may thus under- (over-) estimate the effect of cash transfers on outcomes. From the Current Population Survey (CPS), we estimate that of county residents who are below the poverty line, on average for years 2020-2022, 2.5% moved at least between counties within the past year (includes being abroad, moving between states, or moving between counties within the state). Differential censoring of our outcome is thus possible and is made additionally problematic because we cannot observe it, posing challenges for interpretation. We describe robustness analysis we conduct to account for attrition in the robustness analysis description below.

## Statistical models and hypothesis tests

This section describes the statistical models and hypothesis tests that will make up the analysis — including any follow-ups on effects in the main statistical model and any exploratory analyses that can be anticipated prior to analysis.

### Confirmatory analyses:

#### CATT Analysis

To estimate the 6-month CATTs, we follow the approach to data-stacking laid out by Freedman et al. (2023), and incorporate IV regression using two-stage least squares. This approach performed best in an extensive simulation study conducted in preparation for this plan.

As noted above, individuals were offered cash in “cohorts” — for example, some percentage of people were offered the cash in July, and others in August. Moreover, there was typically a delay between receiving an offer and actually receiving cash. We defined the so-called “modal delay” above as the modal number of months between when individuals receive an offer and when they actually receive the cash. To implement the stacked IV regression, we build a dataset composed of many datasets, each of which corresponds to a different cohort. We drop individuals’ observations that fall within the modal delay, as these might otherwise dilute our estimates of the effect of the cash, which we causally identify based on *offers* rather than on *receipt* of cash.

Each dataset or “layer” in the stack corresponds to an event,  $e$ , in which a given cohort was offered cash. If individuals were offered cash from May to August, for example, then  $e$  can take on the values of 5, 6, 7, and 8, and the stack will comprise four layers. We refer to the modal delay as  $k$ . Thus, if the modal delay is 2, we expect individuals offered cash in the month of July ( $e = 7$ ) to typically receive the cash in the month of September ( $e + k = 9$ ).

When building a given layer, our goal is to construct a simplified version of the data that essentially resembles a 2x2 difference-in-differences, with six months of “pre-” and six months of “post-” offer data for the cohort in question, and the same pre- and post- months of data for individuals who never received an offer of cash during that period. To take account of the modal delay, we specifically include as “pre-” months, those from  $e-6$  to  $e-1$ , and as “post-” months those from  $e+k$  to  $e+k+5$ . Table 3 below provides an example of which months would be included in the layer corresponding to the July cohort. Note that, for the purposes of the regression, we use the recoded variables in the final two columns — thus, every layer is essentially centered in time at the moment the cohort is expected to receive cash (based on when they received the offer and the modal delay).

**Table 3.** Example of which months of data are included in a layer of the stacked data for individuals who receive an offer in July. Shaded months indicate inclusion. The modal delay,  $k$ , is set at 2.

| Month: | Notation:   | Recoded event variable: | Pre-post variable: |
|--------|-------------|-------------------------|--------------------|
| 1      | $e - 6$     | -6                      | 0                  |
| 2      |             | -5                      | 0                  |
| 3      |             | -4                      | 0                  |
| 4      |             | -3                      | 0                  |
| 5      |             | -2                      | 0                  |
| 6      | $e - 1$     | -1                      | 0                  |
| 7      | $e$         | NA (dropped)            | NA (dropped)       |
| 8      |             | NA (dropped)            | NA (dropped)       |
| 9      | $e + k$     | 0                       | 1                  |
| 10     |             | 1                       | 1                  |
| 11     |             | 2                       | 1                  |
| 12     |             | 3                       | 1                  |
| 1      |             | 4                       | 1                  |
| 2      | $e + k + 5$ | 5                       | 1                  |

Each layer only contains the observations from three kinds of individuals:

1. Individuals in cohort  $e$ ,
2. Individuals who were never offered cash,
3. Individuals who were not offered cash in any month prior to or including  $e + k + 5$ .

Each  $e$  dataset therefore comprises a year of data, including only the “clean” comparisons between “offered” and “yet-to-be” or “never” offered individuals. In each dataset, we create the following variables, which are included in addition to those described above:

- Offered – this variable is one for all periods for any individual who was offered to receive the cash in month  $e$ , and zero otherwise;
- Received – this variable is one for all periods for any individual who received the cash in any month, and zero otherwise;

- Post – this variable is one for period  $e+k$  and all periods thereafter, zero otherwise (see final column of Table 3);
- Post\_Offered – this variable is coded by multiplying Offered and Post together (it is equivalent to their interaction);
- Post\_Received – this variable is coded by multiplying Received and Post together (it is equivalent to their interaction);
- Stack\_Counter - this variable is an index for the  $e$  different datasets (e.g., 1 for the first, 2 for the second...);
- Stack\_Weight - this variable is the “Q-weight” from Wing et al. (2024), which corrects for the overrepresentation of some individuals in the stack;<sup>3</sup> and
- Event\_Time - this variable is used in exploratory analyses to understand the dynamics of the effects. See the third column of Table 3: it runs from -6 for the first period in the dataset, to zero for the period in which the cohort would typically receive cash ( $e+k$ ), through to five for the final post-intervention period.

Having obtained  $E$  datasets (where  $E$  denotes the total number of  $e$  cohorts), each corresponding to a specific  $e$ , the final step is to stack the datasets together into one single dataset and compute the corrective Stack\_Weight variable.

To estimate the 6-month CATT using the stacked data, we fit a two-stage least-squares regression. The first stage regression is as follows:

$$PostReceived_{it} = \beta_0 + \beta_1 Post_t + \beta_2 Offered_i + \beta_3 PostOffered_{it} + X\omega + \phi_i + e_{it}$$

Where  $i$  indexes the individual and  $t$  months, and

- $\beta_0$  is an intercept
- $\beta_1$  represents the pre-post difference in receipt for those never offered cash
- $\beta_2$  represents the difference in receipt for those who were and were not offered cash, before anyone was offered cash
- $\beta_3$  represents the effect of the offer on the receipt of cash (this is the main instrumental effect we care about)
- $X$  is a matrix of predictors from the application data, including income and zip code, and  $\omega$  is a vector of coefficients corresponding to their correlation with receipt

---

<sup>3</sup> The approach Wing et al. (2024) adopt improves over previous approaches in that the corrective sample weights, “Q-weights,” adjust for bias that otherwise arises in stacked regressions due to the way in which trends for the intervention and non-intervention groups are weighted differently across layers of the stack. This bias arose due to the differential sample sizes generated by the stacking procedure, and can therefore be accounted for through the corrective sample weights they derive. Additionally, the weights make the typical inclusion of unit-time fixed effects redundant, marking an additional departure from previous stacked regression approaches. For observations with Offered = 1, the weight is 1. For other observations with Offered = 0, the weight is the following fraction:  $\text{prop\_offered\_by\_eventtime} / \text{prop\_not\_offered\_by\_eventtime}$ .  $\text{prop\_offered\_by\_event}$  represents the number of individuals in this observation’s same event time and sub-experiment with Offered = 1, divided by the number of individuals in this observation’s event time with Offered = 1 across sub-experiments.  $\text{prop\_not\_offered\_by\_eventtime}$  repeats that for individuals with Offered = 0.

- $\phi_i$  is a fixed effect for priority group membership, and
- $e_{it}$  an idiosyncratic error term.

We will estimate the second stage using the following model:

$$Y_{it} = \beta_0 + \beta_1 Post_t + \beta_2 Offered_i + \beta_3 \widehat{PostReceived}_{it} + X\omega + \phi_i + e_{it}$$

where  $\widehat{PostReceived}_{it}$  is the predicted probability of receiving cash, as estimated in the first stage.  $\beta_3$  from this second regression is our estimate of interest.

Heteroskedasticity-consistent standard errors (HC2) will be clustered at the individual level (to account for serial correlation within individuals over time and across repeated inclusion in the stack). The regression will be weighted by the Stack\_Weight variable. To account for the differential assignment probabilities, we will normalize the stack weight and multiply this weight by the normalized IPWs.

As we have a panel setup, we will use stepped-wedge IPWs rather than the IPW construction described above. The steps to construct the stepped-wedge IPWs are as follows:

- Define Q, the set of possible treatment periods.
- Define the number of units as N, and the number of units who are never treated at any point as N\_NA.
- Define some treatment probabilities in terms of exposure events.
  - $\Pr(E_i = NA) = N\_NA / N$ . The probability of never being assigned to treatment is  $N\_NA / N$ .
  - $\Pr(E_i \neq NA) = 1 - \Pr(E_i = NA)$ . The probability of being assigned to treatment at some point is just the reciprocal of never being assigned to treatment.
  - $\Pr(E_{it} > t \mid E_i \neq NA)$ . This probability is conditional on being treated at some point in the series. For a given unit, i, in period t, who is going to be treated at some point, what is the probability that the treatment will happen in a period later than t.
  - $\Pr(Z_{it} = 0) = \Pr(E_i = NA) + \Pr(E_i \neq NA) * \Pr(E_{it} > t \mid E_i \neq NA)$ . This is the probability that unit i is not exposed to treatment in period t. There are two ways this can happen: either the unit is assigned to never be treated, which happens with the probability defined above,  $\Pr(E_i = NA)$ . Or, the unit is assigned to be treated but has not yet been treated, which corresponds to the joint probability that the unit gets assigned to be treated at some point,  $\Pr(E_i \neq NA)$ , and the unit will be treated in a period after the present one,  $\Pr(E_{it} > t \mid E_i \neq NA)$ .

- $\Pr(Z_{it} = 1) = 1 - \Pr(Z_{it} = 0)$ . For a binary treatment, the probability of being treated in any period is just the reciprocal of the probability of not being treated.
- Given an actual random assignment, get the propensity weight for each unit-period, using:  

$$p_{it} = Z_{it} * \Pr(Z_{it} = 1) + (1 - Z_{it}) * \Pr(Z_{it} = 0).$$
- Assign each unit an IPW,  $w_{it} = 1/p_{it}$

We note that IV is a consistent estimator under two assumptions. First, the exclusion restriction stipulates that the offer of cash only affects county-administered benefits receipt through its effects on receipt of the cash. This would be violated if, for example, the county distributed promotional materials about county-administered benefits only to those who were not in the offered cash condition, and those promotional materials in turn affected individuals' decisions to apply for benefits. The county provided information to *all* applicants on benefits that they may be eligible for and, to our knowledge, did not treat those offered cash differently from those not offered cash.

Second, the monotonicity assumption stipulates that the effect of the offer of cash was zero or positive on the probability of receiving cash, for every individual in the study. This requires, for example, that there do not exist individuals who will receive cash transfers as a result of getting randomized to the no cash condition. As the county controlled the distribution of cash transfers, we are not concerned about the validity of this assumption.

To estimate the conditional CATT we will follow the same procedure as above, but conduct it again with the subsample of individuals who are marginal for benefits use.

#### *ITT analysis*

To estimate the 6-month ITTs, we will employ the Callaway and Sant'Anna (2021) estimator. The Callaway and Sant'Anna (2021) estimator provides a *cohort-time average treatment effect*, i.e., the average treatment effect for cohort  $g$  at time  $t$ , where a "cohort" is defined by the time period when units are first offered a program/intervention. They show that the cohort-time average treatment effect is identified through outcome regression (OR), inverse probability weighting (IPW), or doubly robust (DR) estimands. The OR approach relies on modeling the conditional expectation of the change in outcomes for the comparison cohorts, the IPW approach relies on modeling the conditional probability of being in cohort  $g$ , and the DR approach exploits both OR and IPW components. They provide ways to aggregate the cohort-time average treatment effects into a single overall treatment effect parameter that is similar to the difference-in-differences estimates in the two period and two group case. They also provide aggregations to estimate how average treatment effects vary with length of exposure to the program/intervention; how average treatment effects vary across the cohorts receiving the program/intervention; and how cumulative average treatment effects evolve over calendar time.



Since we are interested in the 6-Month ITT, we estimate the effect for the cohort  $e$  offered cash in period  $e$ , with a modal delay, to  $e + 6 \text{ months} + \text{modal delay}$  (total post-offer months), by comparing the cohort's pre-offer to total months post-offer change in outcomes, to that of the comparison cohorts (including both those never offered cash, and for those offered cash, their outcomes before they were offered cash) over the same period. We aggregate these effects across the cohort-comparisons to get the average effect over the total months post offer of cash, by taking the weighted average of all cohort-time average treatment effects with weights proportional to the cohort size.

To estimate the 6-Month ITT on benefits use, we will use predictors from the application data, including income and zip code, and an indicator for priority group membership. HC2 errors will be clustered at the individual level. We do not separately include IPWs in this specification as the procedure itself uses a doubly robust estimation method. After this estimation step, the Callaway and Sant'Anna procedure averages the cohort-time average treatment effects; individuals are not given greater weight depending on when they are treated.

We will implement this by using the “csdid” package and:

- creating a variable equivalent to the one described above, which tracks the  $e$  period in which the unit was offered cash and the modal delay (0 if never). This is 8 if the modal delay is 2
- setting the comparison cohorts to include both never treated cohorts as well as not yet treated cohorts
- including a set of covariates as specified in the stacked regression above, that includes predictors from the application data (including income and zip code, and a vector of coefficients corresponding to their correlation with receipt) and a fixed effect for priority group membership
- using the doubly robust estimation method
- using the ‘csdid’ function in Stata and the simple aggregation method of all post-treatment effects, with  $\text{max\_e} = 8$  if there is a 2 month modal delay and computing the event study/dynamic effects to retrieve the estimates for effects after 3 to 8 months post offer

This procedure will drop the effects for the months of the modal delay and aggregate effects for the remaining months (months 3-8 in our example). It will compute the average effect for months 3-8 by taking the weighted average of all cohort-time average treatment effects for months 3 - 8.

To estimate the conditional ITT we will follow the same procedure as above, but conduct it again with the subsample of individuals who are marginal for benefits use.

## **Robustness analysis**

### *CATT analysis*

As a robustness check on the CATT estimate, in addition to the two-stage least squares specification we plan to estimate the effect of receipt of cash transfers by predicting compliance.

Because we do not need the outcome data to predict compliance, and can do this using the baseline and implementation data alone, we were able to conduct this prediction exercise prior to finalizing this plan. To construct the sample of predicted compliers we used a methodology similar to that we will use to predict marginality. Specifically, we:

- (1) Took the application data of everyone who was ever offered cash (irrespective of whether they did or did not receive cash), since for these individuals we know who the compliers and non compliers are (we do not know this for people who were never offered cash).
- (2) Split this subsample into 3/4 training data and 1/4 testing data.
- (3) Fit four binary classification / regression models to the training data: ridge regression (using elasticnet linear in Stata); lasso regression (using lasso linear in Stata); decision tree (using rforest with type class in Stata), and random forest (using rforest with type reg in Stata). We predicted whether the individual actually received cash (i.e., was not ineligible or unreachable, and did not refuse the cash) using the following predictors: cash transfer condition, household income, household income squared, age, age squared, interaction of age with income, household size, household size interacted with age, marital status, education level, satisfaction with life, general physical health, benefits use, impression of the cash transfer program from their first interaction with the program, inability to pay for food, paycheck frequency, risk of losing their home, and number of benefit programs they use. All predictors are self-reported except for the cash transfer condition membership.
- (4) For each model's predictions in the training data, we obtained the threshold for converting the probabilities into classes that maximize the True Positive rate while minimizing the False Positive rate by selecting the highest Youden's J.<sup>4</sup>
- (5) In the testing data we:
  - Used each model to predict the continuous probability of compliance
  - Used the thresholds obtained from the training data above to label testing observations as a complier or noncomplier
  - Computed confusion matrices and obtained the recall and precision in the testing data for each model
  - Chose the model that maximizes recall – the ridge regression. We used recall because, when we simulated this procedure during power analyses for this study, recall was the most important metric to optimize over. This is because the costs to statistical power from false negatives (failing to identify individuals who would be compliers) are higher than those from false positives (erroneously predicting noncompliers to be compliers).

With the ridge regression we obtained 84% recall (i.e., we correctly predicted 84% of the true compliers to be compliers) in the test data. Moreover, in our simulation analysis we found that with 85% recall we can increase our power beyond the minimum 2-3 percentage points ITT effect we

---

<sup>4</sup> The Youden's J estimation that is used is modified to incorporate an estimate of the baseline prevalence as described in Perkins and Schisterman (2006).

could detect. With this subsample of predicted compliers, we will estimate the CATT using the Callaway and Sant’Anna (2021) methodology.

We will consider the results of our robustness analysis to be suggestive. Thus, if we find differences in statistical significance, magnitude, or even sign, between the main and robustness analysis we will consider this to be additional evidence for or against our hypothesis, but not dispositive. For example, if we find that there is a small positive but not-quite-significant effect in the main specification, and a result similar in magnitude that is significant in our robustness test we will consider this to be *suggestive* evidence overall for the hypothesis. Similarly, if we find that there is a significant effect in the main specification and a statistically significant effect in the opposite direction in the robustness analysis, we will consider the overall result to be significant but note the countervailing suggestive evidence.

### *ITT analysis*

We will conduct a robustness check on the ITT using a cross-sectional analysis. In order to estimate a much simpler, albeit less well-powered, version of the average effect of being offered and receiving the cash, we can collapse the panel into a cross-sectional analysis. We will do so by aggregating across months to the individual level, creating a dataset with one row per individual, with the following variables:

- *y\_post* - the average outcome in the 6 months plus the modal delay (month 8 in our running example) including and following the first month in which cash was offered to any individual
- *y\_pre* - the average outcome in the 6 months preceding the first ever offer of cash to any individual
- *ipw* - the IPWs estimated as described above in *Construction of Weights*, using the priority group-level proportions of individuals assigned to receive the cash
- *offered* - an indicator that is one if the individual was ever offered cash, and zero otherwise
- *received* - an indicator that is one if the individual ever received cash, and zero otherwise

We will estimate the overall, cross-sectional ITT by regressing *y\_post* on *y\_pre*, priority group indicators, and *offered*. We will include Lin (2013) adjusted covariates. And, we will estimate the overall, cross-sectional CATT using a two-stage least-squares regression, in which *received* is instrumented by *offered*. This is another robustness check on the CATT, in addition to the one we will conduct on the predicted compliers.

### *Attrition analysis*

To account for attrition we will construct a range of possible estimates we might have obtained had we been able to observe the outcomes of those who move. To do so, we require estimates of the potential numbers of “false negatives” that out-migration could have caused — that is, the number of individuals whose benefit use is incorrectly recorded as 0 because they moved out of the county. Table 3 designates four types of individuals that may be in our data, based on whether they used benefits and / or moved during the study period:

**Table 3.** Potential hidden attrition due to county out-migration

|                                   | Did not move during study                            | Moved during study   |
|-----------------------------------|--|--|
| Did not use benefits during study | A. No false negatives (did not move or use benefits) | B. No false negatives (moved but did not use benefits)             |
| Used benefits during study        | C. No false negatives (did not move)                 | D. Potential false negatives (moved and continued to use benefits) |

We construct our estimate ranges using estimates of the proportion of the sample in the offered cash and no cash conditions that may have fallen into cell D. To do so, we follow a three-step procedure:

1. Using [individual-level data](#) from the 2023 round of the CPS ASEC, which covers the pre-lottery period, we will fit a multinomial model predicting the proportion of individuals falling into cells A, B, C, and D. We will construct the outcome using a combination of the MIGRATE1 variable (moved out of county within state, between states, or abroad) and variables recording benefits use (FOODSTMP, INCWELFR, INCASIST, RENTSUB, GQ). The predictors will include household income (FTOTVAL), household size (FAMSIZE), age (AGE), employment status (EMPSTAT), and marital status (MARST).
2. We will use this model in the lottery data, coding equivalent predictor variables for the individuals in our sample using the pre-lottery intake survey, with one key difference: to the baseline household income of individuals who ever received cash, we will add \$4000. We will predict the proportions of the no cash and offered cash conditions falling into cells A, B, C, and D separately. The key estimates are D\_no\_cash and D\_offered\_cash, which represent the estimated proportions of individuals in the two samples who, based on their baseline covariates, are predicted to have moved out of the county and used benefits.
3. We will use D\_no\_cash and D\_offered\_cash to estimate N\_D\_no\_cash and N\_D\_offered\_cash – that is, the number of individuals in each condition – no cash and offered cash – who are predicted to fall into cell D. We will then compute 2,000 additional estimates of the main analyses, in which we hold all other elements of the analysis fixed, but randomly select N\_D\_no\_cash and N\_D\_offered\_cash individuals in the no cash and offered cash conditions to have their benefits outcome changed from a 0 to a 1 prior to running the analysis.

We will report the interquartile range of the estimates obtained in this manner, as well as the average estimate. We will not report this range as the best guess of the true underlying effect, however. We will use this only to assist with interpretation of the results and understanding their sensitivity to out-migration.

#### *Other analyses*

In addition to the above analyses, we will additionally test all hypotheses using different match rate cutoffs for the population of individuals who are matched to the outcome data (e.g., instead of a 75% match rate use up to a 85% match rate cutoff).

We will also test the hypotheses using a binary outcome for benefits use rather than the total monthly benefits used (i.e., coded as a one if they use at least one benefit in that month and zero otherwise).

#### *Broadening the definition of marginality*

As noted above, there is a risk that our definition of marginality may lead to a small sample and an underpowered analysis. At a minimum, we want to ensure that we are able to detect an effect of five percentage points with at least 60% power. Here, we outline the rule for deciding whether to broaden the definition and how we plan to broaden the definition of marginality in this case.

With a 95% confidence level, we can use this minimum detectable effect and specified power level to determine a maximum standard error beyond which we will switch to a broader definition of marginality.

A common approach to post-hoc power analysis uses the “2.8 rule of thumb,” according to which one can estimate the minimum effect that is detectable with 80% power by multiplying the estimated standard error by 2.8. The number 2.8 is obtained by adding 1.96, the z-score for a 95% confidence interval, to 0.84, the 80th percentile of a standard normal distribution (see [here](#) and [here](#)). We also have a 95% confidence interval, but the 60th percentile of the normal distribution, needed for 60% power, is located at .253. Accordingly, we can determine the maximal standard error that will provide us with an estimated post-hoc power of 60% for a .05 MDE by solving for the standard error (SE) in the following equation:  $(1.96 + .253) \times SE < .05$ .

Using this method, an estimated standard error of .023 or smaller provides us with an ex-post power of at least 60% to detect any effect of size .05 or greater. Thus, if our standard error in the marginality analysis is above .023, we will use an expanded definition.

In the case that our standard error falls above this threshold, we will define marginality for a given benefit using two criteria: either, the individual switched their benefit status for that benefit in the preceding 6 months; or, the individual is eligible for that benefit, did not switch benefit status in the previous 6 months, but has (among non-switchers) a predicted probability of switching that falls in the top 25% of predicted probabilities for that benefit. As above, we include in the “marginal” analysis any individual who is marginal for at least one benefit, and code as their outcome the count of benefits used in a given month for which they were defined as marginal.

We will determine whether an individual is eligible for a given benefit using basic eligibility criteria. For example, for the case of foster care assistance, we will determine eligibility based on whether the individuals have children in their household. In the case of state-level SNAP, we will determine eligibility based on individuals’ income and household size.

To predict switching, we will estimate one random forest model per benefit among all individuals in the sample. For a given benefit, the outcome is a binary indicator of whether that individual switched benefits status (on or off) at least once in the 6 months preceding the first lottery date. Predictors will include the following variables from the intake survey, also used to predict compliance: household income, household income squared, age, age squared, interaction of age with income, household size, household size interacted with age, marital status, education level,

satisfaction with life, general physical health, benefits use, impression of the cash transfer program from their first interaction with the program, inability to pay for food, paycheck frequency, risk of losing their home, and number of benefit programs they use. We will select the random forest model by selecting the variables that minimize the out-of-bag error and validation error. Then, we will use this model to generate predictions for the sample as a whole for that benefit.

### **Exploratory analysis:**

We are not pre-specifying any exploratory hypotheses.

### **Inference criteria, including any adjustments for multiple comparisons:**

We will use a t-test to create p-values for a two-sided test with an  $\alpha=0.05$ .

We have four main hypotheses, each with one associated test and p-value. For the purposes of accounting for the risks associated with multiple comparisons, we think of these four tests as belonging to two separate “families.” We base families on whether the tests correspond to joint hypotheses, whereby a significant effect in either test would be considered probative for the joint hypothesis.

Specifically, we consider the first two of these tests — Hypothesis 1 (6-Month CATT) and Hypothesis 2 (6-Month ITT) — as corresponding to one joint hypothesis that the cash is effective overall, and the second two tests — Hypothesis 3 (Conditional 6-Month CATT) and Hypothesis 4 (Conditional 6-Month ITT) — as corresponding to a second joint hypothesis that the cash is effective for individuals at the margins of stability. Importantly, we will interpret these families independently. A statistically significant effect for Hypothesis 1 or 2, for example, would not lead us to conclude that the second joint hypothesis has been validated, or vice versa for Hypotheses 3 and 4.

With two tests per family, we will test for statistical significance within families using the Holm-Bonferroni adjustment procedure, holding the family-wise error rate (FWER) at 5%. We will take the  $m=2$  p-values, test the lowest against  $\alpha/m$  (here  $0.05/2$ ), and the next lowest against  $\alpha/(m-1)$  (here  $.05/1$ ), or stop after the first test if the first test is not significant.

### **Limitations:**

This analysis plan describes solutions to address the implementation of the randomization and false negatives in our outcome due to out-migration from the county.

An additional limitation is our ability to capture the broader set of effects of cash transfers with the available administrative data. Cash transfers have been found to affect a myriad of outcomes, from feelings of empowerment, to child well being, to labor participation. The county is particularly interested in how cash transfers impact individuals’ financial wellbeing, which the state public benefits database and HMIS datasets cannot capture directly. This includes factors like expenditures, savings, investments, and assets, which some studies have found to be impacted by cash transfers. Debit card data are only available for individuals who received cash transfers.

### **Bibliography**

Bartik, Alexander W., et al. *The Impact of Unconditional Cash Transfers on Consumption and Household Balance Sheets: Experimental Evidence from Two US States*. No. w32784. National Bureau of Economic Research, 2024.

Callaway, Brantly, and Pedro HC Sant'Anna. "Difference-in-differences with multiple time periods." *Journal of econometrics* 225.2 (2021): 200-230.

Dwyer, Ryan, et al. "Unconditional cash transfers reduce homelessness." *Proceedings of the National Academy of Sciences* 120.36 (2023): e2222103120.

Freedman, Seth M., et al. *Designing difference in difference studies with staggered treatment adoption: Key concepts and practical guidelines*. No. w31842. National Bureau of Economic Research, 2023.

Jaroszewicz, Ania, et al. "How effective is (more) money? Randomizing unconditional cash transfer amounts in the US." *Randomizing Unconditional Cash Transfer Amounts in the US (July 5, 2022)* (2022).

Liebman, Jeffrey, et al. "The Chelsea eats program: Experimental impacts." *Rappaport Institute for Greater Boston* (2022).

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

Perkins, Neil J. and Enrique F. Schisterman, The Inconsistency of "Optimal" Cutpoints Obtained using Two Criteria based on the Receiver Operating Characteristic Curve, *American Journal of Epidemiology*, Volume 163, Issue 7, 1 April 2006, Pages 670–675, <https://doi.org/10.1093/aje/kwj063>.

Wing, Coady, Seth M. Freedman, and Alex Hollingsworth. *Stacked difference-in-differences*. No. w32054. National Bureau of Economic Research, 2024.