Analysis Plan

Project Name: Measuring the Downstream Health Effects of SSI Take-up Among Older Adults

Project Code: 2205

Date Finalized: May 1, 2023

NOTE: This is an updated version of a prior pre-analysis plan which was posted on 09/09/2022. The original Analysis Plan and upload date can be found here. We have added one analysis to the Statistical Models & Hypothesis Tests section, and we have added several sensitivity analyses in the Additional Exploratory Analysis section. We have not changed or taken out any analyses that were proposed in the initial version of this plan. Because preliminary results from the initial analysis indicate the effect of SSI receipt on mortality is larger than anticipated and because we gained access to data for a longer observation period at the individual-day level, we are likely sufficiently powered to detect an effect with alternative approaches like the basic 2SLS and ITT that we did not previously think we were sufficiently powered for. Therefore, we are updating this analysis plan to include one additional primary outcome and several additional exploratory analyses.

This document serves as a basis for distinguishing between planned confirmatory analyses and any exploratory analyses that might be conducted on project data. This is crucial to ensuring that results of statistical tests are properly interpreted and reported. For the Analysis Plan to fulfill this purpose, it is essential that it be finalized and date-stamped before we complete our proposed analyses. This analysis is being conducted on data from a prior randomized control trial that has already been analyzed to test the effect of the treatment, mailed letters, on SSI enrollment. We are now using this data to test the effect of letter–induced SSI benefit receipt on mortality, leveraging the fact that the mailed letters exogenously increased SSI benefit receipt.

Project Summary

While there is robust evidence that social welfare policies improve health, we know less about how the administration of those policies affects health. In a collaboration from the Georgetown Better Government Lab and OES, funded by SSA's Retirement and Disability Research Consortium at the University of Wisconsin-Madison Center for Financial Security and by Arnold Ventures, we propose to examine the mortality consequences of receiving Supplemental Security Income (SSI). Relatedly, we will also examine whether decreasing administrative burden has mortality consequences through higher take-up rates. We leverage random assignment of a prior Social Security Administration demonstration field experiment, which exogenously increased take-up among eligible beneficiaries (Hemmeter et al. 2020), to causally identify the effect of SSI on mortality using mailed-letter assignment as an instrumental variable. The initial study identified over 4 million adults 65 or older who were potentially-eligible for SSI and randomly sent letters to

10 percent of them in September 2017, informing them of their potential eligibility. It found that these letters, a form of low-cost nudge that reduced learning costs, increased SSI applications by 600 percent (or 5.0 percentage points) and led to an increase in SSI awards of 340 percent (or 1.8 percentage points) over the study period. In this study, we take advantage of the random assignment to letters and use it as an exogenous leverage to estimate the effects of SSI award on recipient's mortality in a two-stage estimation procedure.

Preregistration Details

This Analysis Plan will be posted on the OES website at [oes.gsa.gov](oes.gsa.gov) before we conduct the proposed analysis.

Hypothesis

H1: Receipt of letter-induced SSI benefits will decrease mortality.

We acknowledge that SSI receipt is indicative of low income and resources—factors that increase mortality (Backlund et al. 1999). Therefore, our hypothesis refers to a marginal effect of receiving SSI benefits on mortality.

We also acknowledge that because of program linkages and limitations of our data, we will not know whether the effect of SSI benefit receipt on mortality is a pure influence of SSI participation, or whether it is also influenced by the subsequent access people may gain to other programs (e.g. Medicaid/SNAP). Enrolling in SSI benefits may lead to automatic enrollment or enable access to an easier enrollment process for other benefit programs, such as Medicaid and SNAP, but we do not have access to enrollment data for these other programs to parse out these effects. For our broader question, which is understanding the effects of administration on mortality, both potential effects are relevant. Because there's this easy administrative linkage from SSI to Medicaid, increased access to SSI means—due to administrative reasons—access to Medicaid as well. So we're getting the total effect of this administrative easing by reducing learning costs. Another way to think about this is that one possible mechanism for the influence of SSI participation on mortality is via its impact on increasing access to other welfare support programs given these programmatic linkages. This is one of many potential mechanisms that may explain that relationship.

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):
Data comes from SSA administrative records. Some observations in the data are at the individual-monthly level, and other observations are at an individual-day level. The population was drawn from Old-Age, Survivors, and Disability Insurance benefit records. The follow-up data

identifies any individuals who applied for SSI benefits, as well as a small number of SSI-program-related covariates (e.g., age, sex, date of death (if applicable), state of residence, whether a SNAP application was filed at the same time as the SSI application). The data comes from the following sources:

- *Master Beneficiary Record (MBR):* The MBR contains records for all individuals receiving OASDI benefits.
- *Supplemental Security Records (SSR):* The SSR contains records for all individuals who receive a federal SSI benefit, as well as those who applied for the benefit. This data source also contains information on whether SSA supported an application for SNAP at the same time they applied for SSI benefits.
- *Numerical Identification System (Numident):* The Numident contains records for all individuals who filled out an application for a Social Security Number (Form SS-5) and death records, among other records.
- *Disability Insurance and Supplemental Security Income Demonstration Projects and Experiments System:* This is the SSA demonstration system of record, which includes all non-SSA data for all demonstration projects. This provides the random assignment identifiers of the SSI take-up demonstration field experiment.

Sample Selection for Randomized Control Trial:

- Individual random-assignment was used.
- Inclusion Criteria: The MBR was used to identify individuals potentially-eligible for SSI. Individuals must be aged 65-80, receiving OASDI payments below the eligibility threshold for SSI, and not enrolled in SSI at the time of randomization
- Expected study enrollment timeline: Our data comes from a randomized controlled trial conducted as a demonstration project by SSA. Assignment to the mailed letter group was based on administrative data in the MBR. As such, enrollment and consent were not relevant for this study.
- Balance Checks: The treatment and control groups were determined to be balanced across relevant variables: age, sex, prior SSI receipt, potential SSI payment amount, Medicaid-SSI linkage type for individual's state of residence, beneficiary type, and WEP/GPO status.

Table 1. Summary Statistics and Balance Checks for Random Assignment

	Overall		Control		Letter		
	Mean	**Sd**	**Mean**	**Sd**	**Mean**	**Sd**	**p-value**
Age (overall)	71.33	4.45	71.34	4.45	71.32	4.45	0.15
65 (%)	7.68	26.63	7.68	26.63	7.69	26.65	0.73
66-70 (%)	41.66	49.30	41.65	49.30	41.70	49.31	
71-75 (%)	28.98	45.37	28.98	45.37	29.00	45.38	

76-80 (%)	21.68	41.21	21.69	41.21	21.61	41.16	
Male (%)	31.71	46.54	31.72	46.54	31.65	46.51	0.34
Prior SSI Receipt (%)	12.50	33.07	12.50	33.07	12.50	33.08	0.98
Potential SSI Amount	218.92	176.60	218.96	176.61	218.58	176.46	0.19
1st Quintile (%)	20.03	40.02	20.03	40.01	20.11	40.08	0.26
2nd Quintile (%)	19.98	39.99	19.98	39.98	19.98	40.02	
3rd Quintile (%)	20.00	40.00	20.01	40.01	19.91	39.93	
4th Quintile (%)	20.00	40.00	19.99	39.99	20.03	40.03	
5th Quintile (%)	19.99	40.00	20.00	40.00	19.92	39.94	
Medicaid-SSI Type							
209(b) (%)	13.10	33.74	13.10	33.75	13.05	33.69	0.56
SSI Criteria (%)	5.36	22.53	5.36	22.52	5.38	22.57	
1634 (%)	81.54	38.80	81.54	38.80	81.57	38.77	
Beneficiary Type	81.14	39.12	81.14	39.12	81.17	39.10	0.24
Other	0.18	4.28	0.18	4.27	0.19	4.40	
Worker	81.14	39.12	81.14	39.12	81.17	39.10	
Spouse	16.28	36.92	16.28	36.92	16.27	36.91	
WEP/GPO Case (%)	24.86	43.22	24.87	43.22	24.85	43.21	0.77

Outcomes to Be Analyzed:

The primary outcome of interest is mortality. Our primary analyses examine change in mortality two different ways: using a covariate-adjusted Cox Proportional Hazards Model with a control function ("survival approach"), and a more standard covariate-adjusted Two-Stage Least Squares (OLS LATE) Model. We explain the estimation and interpretation of these estimates in more detail below.

Imported Variables:

The MBR contains information about each individual's address including zip code and state of residence. We will merge in several aggregate-level controls for the context in which a person lives from the American Communities' Survey public use files. These variables will be measured at the zip code or state level as follows:

Zip Code Level Variables:

Economic characteristics

- Median Household Income
- Percent of people living below the poverty threshold
- Unemployment rate

Racial Demographics

- Percent of population who are Black
- Percent of population who are Hispanic

Resources

- Measure of education - Percent of people with more than HS

Data Source for all zip code-level variables: American Community Survey, 2019 5-year estimates.

The following state-level variable is already in the SSA data:

State-Level Variables:

- Whether the state has automatic Medicaid enrollment given SSI benefit receipt, whether the state has the same eligibility threshold for Medicaid and SSI but no automatic enrollment, or whether the state has a more restrictive Medicaid eligibility threshold than the SSI eligibility threshold. (Social Security Administration's Program Operations Manual System)

Transformations of Variables:

For subgroup analyses, we will transform the following variables into categorical variables as follows:

Table 2. Variable Transformation Descriptions

Subgroup (At the time of mailing/random assignment, unless otherwise noted)	**Categories**	**Variable name(s) in dataset**
Age	Continuous	age
Sex (as indicated at time of SSN application)	Male/Female	male
Projected SSI payment amount	Monthly maximum federal benefit (for a non-blind individual) + 20 - OASDI benefit	potentialssi
Average monthly benefit received	Calculated as the sum of an individual's monthly, federal SSI benefit totals divided by the total number of months during our period of observation they received SSI benefits).	avessi28

	Here, we use the average monthly benefit received rather than the total amount or the number of months receiving benefits, because these will likely be highly correlated with mortality. People who die early in the observation period will necessarily receive fewer total months of benefits and a lesser total amount of benefit. We use the average value rather than the proportion of increase because the literature on income and mortality shows that every additional dollar matters.	
Windfall Elimination Provision/Government Pension Offset	Yes/No	wepgpo created using wep_ind, gpo_ind. If wep_ind = 1 or gpo_ind = 1, new variable wepgpo = 0
OASDI Beneficiary Type	Other, Worker, Spouse, Widow (categories used in Hemmeter et al. (2020)).	bentype2
SNAP application taken, if applied for SSI	This variable will be constructed using the two variables from the MBR. 1) Whether the person was already enrolled at SNAP at the time of SSI application and 2) whether a SNAP application was also taken at the time of SNAP application, we will construct a categorical variable where 0 = no application taken and not already enrolled in SNAP, 1 = not already enrolled in SNAP but application taken, and 2 = already enrolled in SNAP. Note: This variable will not be used in treatment effect estimation because of the potential for post-treatment bias. We will only use it to report descriptive statistics about the people who fall into these three groups.	fs_recipient (This data element indicates whether the SSI recipient currently receives food stamps as part of a food stamp family or had applied for food stamps within the 30-day period prior to filing an SSI application.) The categories are: **N** (Not receiving food stamps), **Y** (Currently receiving food stamps or has applied within 30 days prior to filing for SSI benefits), **Z** (Invalid character(s) transmitted), and **Blank** (no input made). fs_request (This data element indicates whether SSA field office personnel took a food stamp application for the

		claimant at the time of application for SSI benefits). The categories are: **N** (SSA did not take a food stamps application), **Y** (SSA took the food stamps application), **Z** (Invalid character(s) transmitted), and **Blank** (no input made). For this subgroup analysis, blank and "Z" responses will be coded as 0 indicating that no application was taken and the person was not already enrolled in SNAP (as described in the previous column). We will also report descriptive statistics detailing how many observations have Blank and "Z" values to convey how much of a problem missing data may be here.
Medicaid State Type	Categorical variable with three possible values: Automatic Medicaid enrollment state (209b), nonrestrictive Medicaid eligibility state without automatic enrollment (SSI Criteria), or restrictive Medicaid eligibility state (1634)	state_group
Race	Black, White, Other, Missing (Note: The Numident does not include the full set of OMB approved race and ethnicities.)	race
Random assignment to mailed-letter group	Yes/No	treatment
Lives at the same address as someone who received a letter	Yes/No	New variables: addr_duplicate = 1 if there is anyone in the SSA data (i.e. only people who are potentially eligible for SSI as of Sept

		2017) who has the same address as the individual (i.e. addr_zip, addr_city, addr_state, addr_scc), else addr_duplicate = 0; count_addr_duplicates = number of people in the data who have the same address; spillover_possible = 1 if there is anyone in the data who lives at the same address and at least one of those people was assigned to the letter group, else spillover_possible = 0
Date that individual received additional SSA mailing, if applicable	Date	

Transformations of Data Structure:
We will transform the data from cross-sectional into time-to-event data for the survival approach.

Data Exclusion:
Since we are leveraging the random assignment from Hemmeter et al. (2020) for our analysis, we will use the 4,016,461 OASDI beneficiaries who are potentially eligible for SSI in their sample. We will not exclude any of these observations.

Treatment of Missing Data and Outliers:
As this is administrative data, any missing SSA-related data can be directly assigned a null value. There should not be any null values for the variables in our primary analyses. We will report descriptive statistics, so that it is clear how much missing data there is here and how the scope of the missingness affects any of our inferences. There will also likely be missing data for whether a SNAP application is taken. In this case, missing observations will be assigned a zero-value meaning that no SNAP application was taken (see the Variable Transformation chart above for more information about how this variable is constructed).

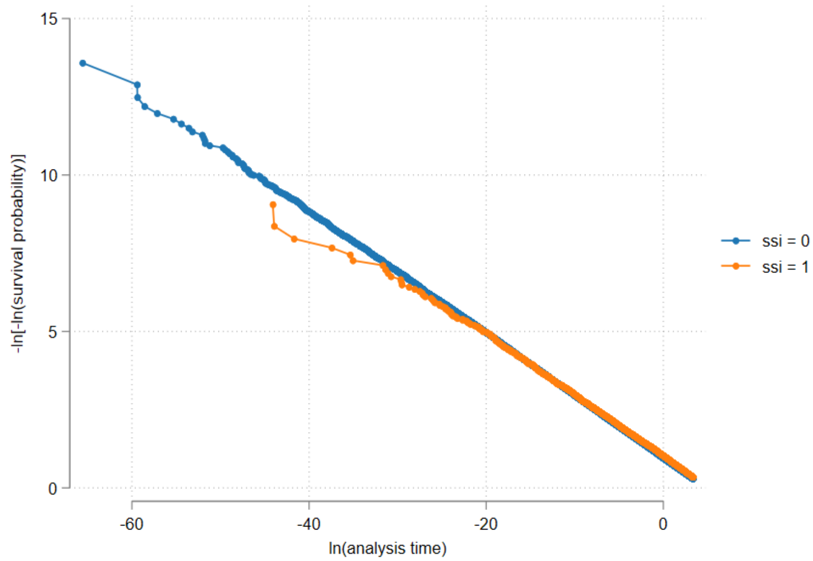
Given our outcome of interest is mortality or time-to-death and our data is right-censored at the end of our observation period, we do not expect any outliers.

Descriptive Statistics, Tables, & Graphs
Note: All of the graphs below are based on simulated data. They show an example of what each of the graphs may look like, but they are not based on data from our study.

We will use two methods to demonstrate the proportional hazards assumption, important to our survival approach, is met: 1) a log-log plot and 2) the Kaplan-Meier curve compared with predicted cox model curve:

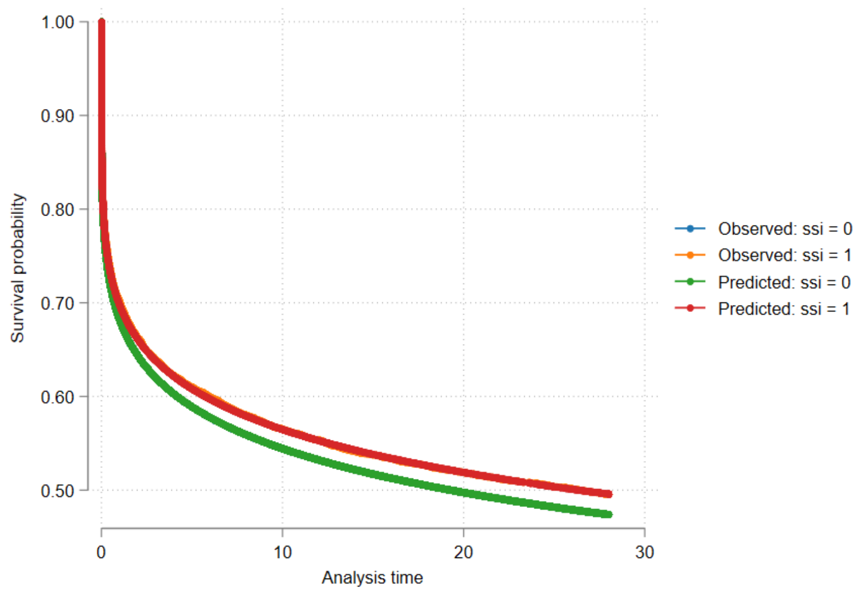
1) **Log-Log plot:** The two lines in this plot should be roughly parallel if the proportional hazards assumption is not violated.

Figure 1. Example Log - Log plot using Simulated Data



2) **Kaplan-Meier curve compared with predicted cox model curve:** The predicted and observed curves should be close together if the proportional hazards assumption is met.

Figure 2. Example K-M vs Predicted Cox Model Curve using Simulated Data



We will also include the following plots:

Figure 3. Hazard Curve for SSI Recipients vs Non-Recipients

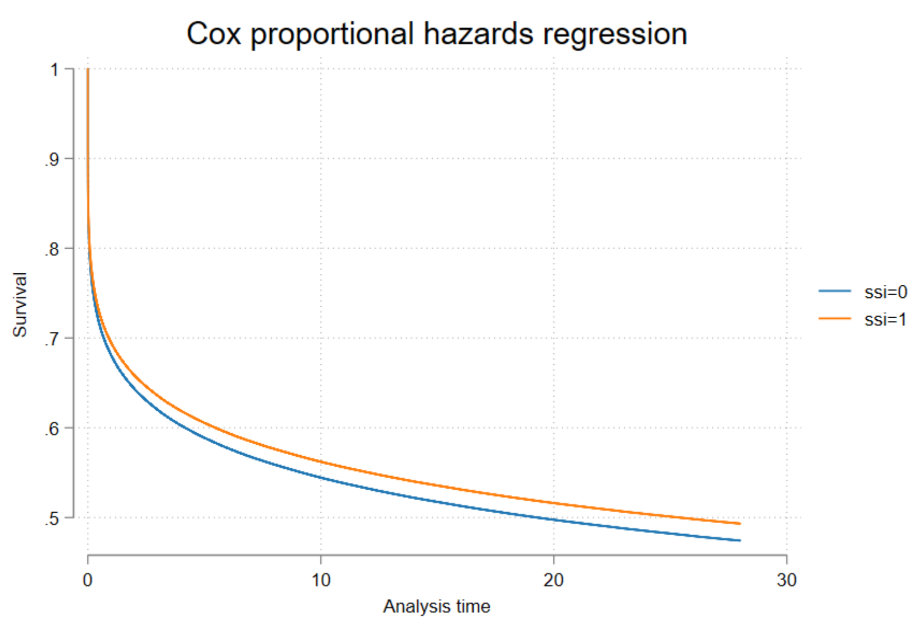
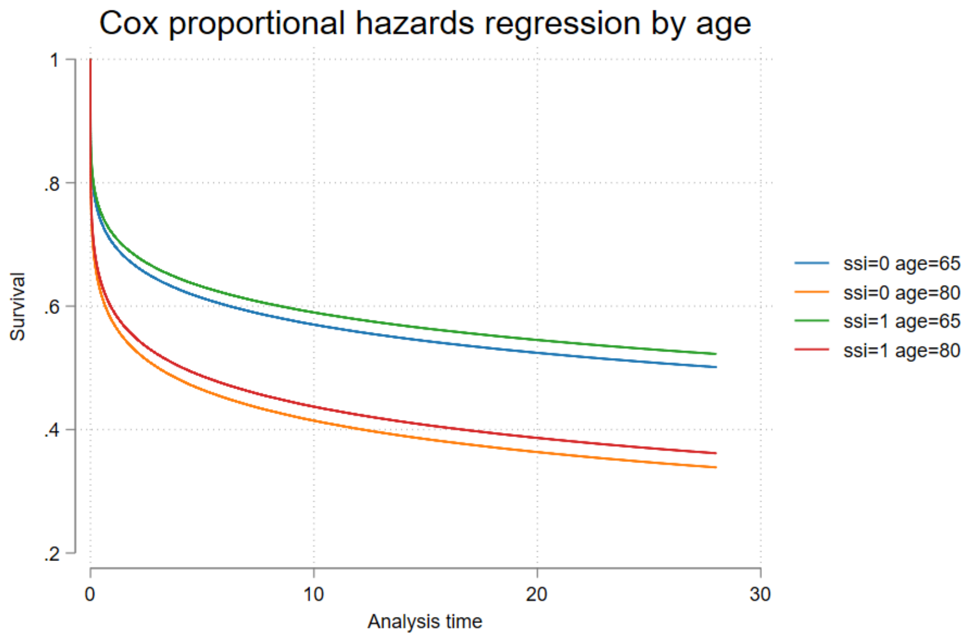


Figure 4. Hazard Curve By Age Group, By Race, and By Projected Benefit Amount for SSI Recipients vs Non-Recipients

As an example, the graph below shows hazard curves by age using simulated data.



Statistical Models & Hypothesis Tests

This section describes the statistical models and hypothesis tests that will make up the analysis — including any follow-up tests or any exploratory analyses that can be anticipated prior to analysis.

Confirmatory Analyses:
H_0: The receipt of letter-induced SSI has no effect on mortality.

That is, in the models below, H_0: $\beta_1(t) = 0$ for all *t*.

Statistical Models:
We use the random assignment to the mailed letter group as an instrumental variable. Recent studies have used similar designs to study the effect of other social and health insurance programs on mortality and health, such as TANF (Courtin et al. 2020), EITC (Dow et al. 2020), SNAP (Heflin, Ingram & Ziliak (2019), Medicaid (Baicker et al. 2013) and general marketplace-based health insurance coverage (Goldin, Lurie, and McCubbin 2021). Goldin, Lurie, and McCubbin (2021)'s and Baicker et al. (2013)'s approaches are most similar to ours. Goldin, Lurie, and McCubbin (2021) use random assignment to a mailed-letter group as an instrumental variable and show letter-induced enrollment in ACA marketplace health insurance coverage decreases mortality. Baicker et al.

(2013) use random selection into Oregon's Medicaid lottery as an instrumental variable approach to assess how Medicaid coverage affects health.

Given this instrument, we use two primary approaches to test the effect of SSI on mortality.

1) Covariate-Adjusted Cox Proportional Hazards Model with a Control Function
2) Covariate-Adjusted Two-Stage Least Squares (OLS LATE) Model

Covariate-Adjusted Cox Proportional Hazards Model with a Control Function:
We will use a Cox Proportional Hazards Model estimated through a control function approach (Tchetgen Tchetgen et al. 2015). Initially, we pre-registered the survival approach after *ex ante* power analyses indicated that a time to event approach gives us greater statistical power to detect an effect on mortality. Also, we are interested in time to death as an outcome, in addition to the probability of death over our study time frame. The 2SLS approach described below only provides the latter.

As our primary survival analysis, we will use covariate adjustment at both stages: the control function and the Cox Proportional Hazards model (Equations ~~3~~ 1 and ~~4~~ 2 below). We think it is important to include state fixed effects in both the control function and the hazard function because there are factors that we cannot measure that vary at the state level and affect both mortality and the likelihood of enrollment in SSI benefits given someone received a letter. These factors include: environmental quality, quality of local healthcare systems, the number and accessibility of SSA field offices, the amount the state offers as an SSI supplement (if any), and other program linkages to SSI benefits (such as SNAP and Medicaid).

The covariate-adjusted survival model for our confirmatory and exploratory analyses is:

Equation 1: Adjusted Control Function:

$$SSI\ Enrollment\ =\ c_0\ +\ c_1 Letter\ +\ c_2 Sex\ +\ c_3 Age\ +\ c_4 Potential\ SSI\ Amt\ +$$
$$c_5 WEP/GPO\ +\ state\ fixed\ effects\ \ +\ \Delta$$
$$where\ \Delta\ =\ SSI\ Enrollment\ -\ Pr(SSI\ Enrollment\ =\ 1|\ Letter)$$
$$(\Delta\ is\ mean\ zero\ residual\ error\ independent\ of\ assignment\ to\ mailed\ letter\ group)$$

Equation 2: Adjusted Hazard Function:

$$\overline{h}(t|SSI\ Enrollment,\ Letter)\ =\beta_o(t)\ +\ \beta_1(t)\ SSI\ Enrollment\ +\ \beta_2(t)\ Age\ +\beta_3(t)\ Sex\ +$$
$$\beta_5(t)\ WEP/GPO\ +\ state\ fixed\ effects\ \ +$$
$$\{\rho_o(t)\ +\ \rho_1(t)\ Letter\}\ \Delta$$

Equation 1 includes the covariates we expect to affect enrollment in SSI, while Equation 2 includes the covariates we expect to affect mortality. While there are additional covariates we would ideally include in Equation 2 because of their established effect on mortality, we are limited by what variables are included in the MBR.

As a sensitivity analysis, we will also test the most parsimonious model, without any adjustment for fixed-effects or covariates following the approaches of Baicker et al. (2013) and Goldin, Lurie, and McCubbin (2021).

We use discrete time; therefore, the hazard rate for an individual is the probability that the individual will experience an event at time *t* while the individual is at risk for having the event.

Covariate-Adjusted Two-Stage Least Squares (OLS LATE) Model:
At the time of pre-registering our initial analysis plan, we did not believe we were sufficiently powered to estimate an effect with this approach. However, we now have evidence that we are sufficiently powered to estimate a 2-Stage Least Squares (2SLS) regression model. The 2SLS approach is commonly used for instrumental variable analyses like ours (e.g. Gerber, Green, and Sachar 2003; Baicker et al. 2013; Goldin, Lurie, and McCubbin 2021). Goldin, Lurie, and McCubbin (2021) and Baicker et al. (2013) use the 2SLS approach to determine whether or not enrollment in insurance makes a person less likely to die or experience a health problem, respectively, over the time period under analysis. Similarly, Gerber, Green, and Sachar (2003) use the 2SLS approach to determine the probability a person votes.

2SLS uses Ordinary Least Squares (OLS) regression to estimate a Local Average Treatment Effect (LATE); therefore, from here, we refer to the approach as the OLS LATE approach. In this case, the LATE is the effect of letter-induced enrollment in the SSI program on mortality. This is given by β_1 in the Stage 2 equation below.

Stage 1:

$$SSI\ Receipt \ = \ \gamma_o + \gamma_1\, Letter\ Assignment \ + \ \gamma_2\, Age \ + \gamma_3\, Sex \ + \gamma_4\, WEP/GPO \ + \gamma_5\, Potential\ SSI\ Amt \ + \ state\ fixed\ effects \ + \ \upsilon$$

Stage 2:

$$Mortality \ = \beta_o + \ \beta_1 \widehat{SSI\ Receipt} \ + \ \beta_2 Age \ + \beta_3 Sex \ + \ \beta_4 WEP/GPO \ + \ state\ fixed\ effects \ + \ \epsilon$$

We covariate adjust Stage 1 using the same covariates used to adjust the Control Function in the survival approach, and we use the same covariates to adjust Stage 2 that we used to adjust the hazard model.

Power Analysis:
Simulation-Based Approach:
To select our methods for this study, we conducted a simulation-based power analysis for the survival and the OLS LATE approaches. Simulation-based power analysis generally assumes a close match between the simulated data generating process and the statistical model in question. However, because we wanted to compare power for the two approaches, we use one data

generating process to simulate data and calculate the minimum detectable effect (MDE) for each approach on this central dataset. In practice, this required several nested loops in Stata.

Because the data is right-censored survival data, we use "survsim," a Stata package for simulating survival data according to a Weibull distribution. This distribution requires two parameters: *lambda*, which affects the scale of the distribution, and *gamma*, which affects the shape. As shown in the code in the Appendix, we test the robustness of our power calculations for a range of each of these variables. The shape and scale of survival data are indicative of the average survival rates of the population and how those survival rates change over time. To determine reasonable ranges for these, we use estimates from the Social Security Administration's Life Expectancy tables and take into account that our population is very poor, older adults who will have lower life expectancy than an average, higher-income older adult. We also consider that COVID-19 pandemic, which occurred during our observation period, accelerated the mortality rate beyond usual SSA estimated averages.

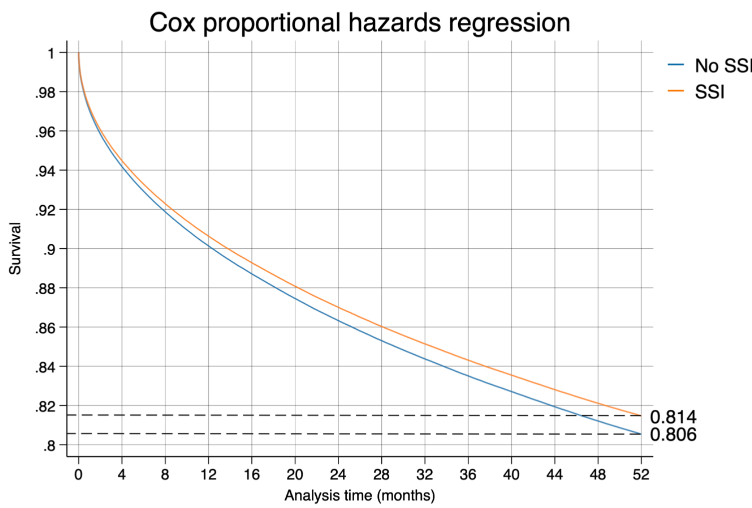
We also use the following parameters from the SSA data:

- Length of observation period: 52 months
- Number of observations: 4,016,461 with 400,000 randomly assigned to letter group
- Age and sex distributions from the SSA data
- Effect of letter on SSI enrollment: 0.005 for non-letter recipients and 0.0166 for letter recipients. For both groups, the standard deviation of the letter's effect on enrollment is 0.001.

Power Analysis Results:
At 80 percent power for the OLS LATE approach[1], the MDE is an approximately 42 *percent* reduction in mortality. This translates to a roughly **7 percentage points** reduction in mortality if the mortality rate for the control group is assumed to be 16-20 percent, which we think to be a reasonable range. At 80 percent power, the MDE for the survival approach is a 7 *percent* reduction in mortality, which translates to a less than **1 percentage point** reduction. While the OLS LATE model gives us a single estimate of the effect of SSI on mortality, the survival approach also gives us how the effect changes over time. Because of this, the MDE for this model is best understood through a graph.

[1] In the OLS LATE model, we include sex and age as covariates because these are included in our data from SSA, and they have a well-established and known effect on mortality; therefore, including them significantly increases our power to detect an effect. We do not include these in the power calculation for the survival approach simply because the model is well-powered without including them.

Figure 5. Simulated survival curve showing an average, over-time mortality reduction of a 7 percent for SSI recipients vs non-recipients



In Figure 5 above, the values are based on gamma of 0.05 and lambda of 0.3. While the percent still alive at the end of the time period for each group changes when we change the gamma and lambda values, the difference that we are able to detect does not change significantly within reasonable ranges of lambda and gamma. The y-axis scale on this plot is limited in range from 0.8-1.00, so that we can emphasize the magnitude of the difference between the two curves at 24 months: 81.4 percent survival rate for people who received SSI benefits compared to a survival rate of 80.6 percent for those who did not receive SSI benefits.

For reference, the U.S. Social Security Administration's (2019) Actuarial Life Tables estimates people aged 65 to 80 to live, on average, to 85.5 (Males) or 87.5 (Females) years of age, with the yearly probability of death for people aged 65-80 being 0.03 (Males) and 0.02 (Females).

Another reason we initially pre-registered the survival approach is that we have a longer time period of analysis than do prior studies like Goldin, Lurie, and McCubbin (2021). To this point, the survival curves in Figure 5 demonstrate that the longer the time period under analysis, the more likely we are to be able to detect an effect, as the survival curves get further apart.

Exploratory Analyses:

First, we will test that the proportional hazards assumption is not violated using the methods described in the Descriptive Statistics, Tables, and Graphs section.

We will report descriptive statistics for the average elapsed time between letters being mailed and people enrolling in SSI benefit. We will also test for local level effects of where a person lives by specifying a model with additional control variables for the zip code contexts in which people live. The services available to people from eligible populations are often determined not by what their problems are or what services they need, but by where they live (Sandefur and Smyth 2011). The variables we will include are listed in the Imported Variables section.

Finally, we will conduct subgroup analyses as sensitivity tests using the same modeling and estimation approach as our confirmatory model. The following chart summarizes the subgroups we will test.

Subgroup Analyses:

Subgroup	**Defense for exploring heterogeneity**
Age (Quintiles)	Hemmeter et al. (2020) show that there are heterogeneous effects of mailed letters on enrollment in SSI by age group. It is also plausible that there are heterogeneous effects of SSI benefits on mortality by age group because as people age, their health becomes more susceptible to intervention.
Sex	Females tend to have longer life expectancy than do males.
Potential SSI payment amount (Quintiles)	Letters had a larger effect on enrollment for people who had a smaller projected payment amount; therefore, we include this as a covariate in the control function. With respect to the effect on mortality, we predict there may be a larger reduction in mortality for those with a larger potential payment amount (and thus fewer resources at the baseline) who did manage to enroll because research has shown that even small absolute amounts that account for relatively large percentage increases in a person's income can have significant health effects. I. In fact, the literature would suggest that every additional dollar a person receives would decrease their mortality. However, in this case, we may not find any significant differences between income groups because Medicaid receipt swamps any of those differences because it not only increases financial resources by increasing income, but also by decreasing out-of-pocket (OOP) healthcare costs that remain with Medicare alone. This decrease in OOP healthcare costs in turn increases health care access, which also negatively affects mortality. There may be a correlation between the additional income gained from SSI eligibility and the likelihood that they become new Medicaid beneficiaries.Since we do not have Medicaid enrollment data, we cannot parse out the difference in these two effects in this study.
Medicaid State Type	While there is heterogeneity in states' linkages between Medicaid and SSI, Hemmeter and Bailey (2015) show that 97.4 percent of people aged 65 and older who receive SSI benefits are also enrolled in Medicaid, so despite heterogeneity in state's SSI-Medicaid linkages, most SSI beneficiaries over 65 are enrolled in Medicaid. Further, while we would love to know the role of Medicaid in all of this, for our broader question, which is understanding effects of administration on mortality, it really doesn't matter. Because there's this easy administrative linkage from SSI to Medicaid, increased access to SSI means—due to administrative reasons—access to Medicaid as well. So we're getting the total effect of this administrative easing by reducing learning costs.
Race	There is evidence that even at similar income levels, racial disparities in mortality remain among older adults (Yao & Robert 2008). We will conduct exploratory analysis to see whether the effect of SSI benefits on mortality varies by racial group, particularly whether it differs between Black, White, and Latino/a individuals. We acknowledge limitations in the racial data contained in the Social

	Security Record we are using (see further discussion of this in the Limitations section below).
COVID-19 effects	Social support programs like Supplemental Security Income are in part intended to support people in times of emergency. As such, we argue our inferences remain externally valid even though one of these emergencies, the COVID-19 pandemic, occurred during our period of observation and affected the mortality of subgroups within our population differently. For example, research shows low-income and Black individuals are more likely to die from COVID-19 than higher-income and white individuals, respectively (Lopez, Hart, and Katz 2021; Millet et al. 2020; Price-Haywood et al. 2020). As a sensitivity analysis, we will test whether there were significant changes in the effect of SSI benefits on mortality potentially induced by the COVID-19 pandemic by testing whether there are significant changes in the relationship between SSI benefits and mortality between the two following time periods: 1) our entire period of observation Sept 2017 to January 2021 and 2) Sept 2017 until March 2020 (when cases of COVID-19 began to spike in the U.S). But it should be noted that since we cannot identify cause of death, any differences we find between the two time periods may not be due to COVID-19 and may be affected by the fact that the longer a person receives SSI benefits the greater the potential for those benefits to positively affect their health and extend their life.
Spillover Effects	In our data, we have addresses for all individuals. We will match these and construct an indicator for whether a person lives at the same address as someone who received a letter, but were not assigned to the mailed-letter group themselves. It is possible that someone who was not assigned to the mailed-letter group could have been exposed to a letter if they lived at the same address as someone who did receive a letter. In the original study, Hemmeter et al. (2020) did not test for this possible spillover effect.

California's 2019 SNAP Benefit Changes:

It should be noted that during our period of analysis, California expanded eligibility for SNAP benefits to SSI beneficiaries. Previously, recipients of SSI benefits in CA were not eligible for SNAP because the state offered an SSI supplement. However, in 2019, CA removed this restriction and SSI beneficiaries became eligible for SNAP benefits. There was widespread take-up of this new benefit. Approximately 75 percent of those eligible enrolled; therefore, given CA's large population size and that they make up a significant portion of our population and that SNAP benefits have been independently linked to decreased mortality (Heflin et al. 2019), we will test to make sure that this increase in SNAP benefits is not driving our findings. State fixed effects should capture this, but as an additional sensitivity analysis, we will also duplicate the analysis leaving CA out of the sample.

Cause of Death:

It should also be noted, given the mechanisms through which increases in income and access to Medicaid decrease mortality, we do not expect this effect to be constant across all causes of death. We expect these mechanisms to primarily operate on preventable causes of death (as categorized by the OECD), such as:

- Endocrine and metabolic diseases (e.g. diabetes, high blood pressure, obesity)
- Cardiovascular and Circulatory System Disease

- Respiratory Diseases
- Neoplasms (cancer)

If we are able to obtain a data use agreement with the Center for Disease Control to use National Death Index data, we will restrict our analysis to only mortality by preventable causes.

Possibility of Contamination of the Control with Second Letter Mailing:
Throughout 2020, after the onset of the COVID-19 pandemic and given the enrollment induced by the letters from the prior RCT, SSA sent out additional letters to people who were potentially eligible for SSI informing them about the benefits. We are working on getting more information about this second mailing: When did the mailing occur? Who was the population targeted? We will use this information to 1) conclude if it is possible our control was contaminated, and 2) if the control was potentially contaminated, we will conduct sensitivity analyses where we restrict the time period of observation to only that prior to the second mailing. It is possible, but not guaranteed, that we will be able to access the assignment to the second-mailed letter group. Although the second mailing was not random, we will use the assignment to the second mailed letter group for additional sensitivity analyses.

Additional Exploratory Analyses (Updated 5/1/23):
Alternative Specifications of the Primary Models:
Intent-to-Treat Model:
This model is a conservative test of the effect of reducing administrative burden on mortality.

$$Mortality = \beta_o + \beta_1 Letter\ Assignment + \beta_2 Age + \beta_3 Sex + \beta_4 WEP/GPO + state\ fixed\ effects + \epsilon_{it}$$

Alternative Measurement Strategies:
Instrument for SSI Application rather than for SSI Enrollment:
We will also estimate our two primary models instrumenting for SSI application instead of successful SSI application.[2] We expect that in these models, we should find that the effect of SSI application on mortality is slightly smaller than the effect of SSI enrollment on mortality. This is a more conservative approach than our primary models because not everyone who applies for SSI is approved for the benefits.

Fixed Effect for SSA office location:
Using this public database of SSA office locations published by SSA, we aim to calculate and include a SSA field office-fixed effect in our primary models as a sensitivity test if feasible.

Measuring Age by Month:

[2] We have already conducted this exploratory analysis for the survival mode as of 02/10/23. We include this here, as we also plan to conduct this analysis for the new, additional primary analysis: OLS LATE.

In our primary models, we include age as a covariate. The primary specifications include the age, measured in years, of an individual at the time the demonstration letters were mailed. Age is calculated from the individual's date of birth, which is included in the SSA data. As an exploratory analysis, we will also calculate age by year if feasible and re-estimate our primary models. In late life, life expectancy changes significantly even within a year.

Inference Criteria, Including Any Adjustments for Multiple Comparisons:
For all tests, we will use two-tailed tests, and we will use an alpha value of 0.05 to assess statistical significance. However, we will report precise p-values and we will contextualize interpretations for results where $0.05<p<0.10$. We will not correct for multiple tests because we are only conducting two confirmatory tests.

Limitations:
Limitations for exploratory analyses of heterogeneous effect of SSI benefits on mortality by race:
There is significant missing data for the race variable we have, as well as it is difficult to compare the data that is there. Even when individuals had to manually file Social Security card applications at an agency field office, race and ethnicity information was voluntary because it is not necessary for SSA to administer the program. Since the information is voluntary, it is a self-selecting sample. The agency has changed the number and definition of race categories over the years, so long-term comparisons are difficult (Martin 2016).

Limitations to subgroup analyses using SNAP data (that is collected as part of SSR):
We think it plausible that people who receive both SNAP and SSI benefits may have greater reductions in mortality than people who only receive SSI benefits. However, the data we have only records whether SSA supported a SNAP application alongside an individual's SSI application (see Transformations of Variables section for further discussion of this). The number of cases where a SNAP application was taken are relatively few, and taking an application does not indicate the person actually received the benefits. Therefore, we cannot analyze the subgroups of people who were enrolled in SNAP prior to being grouped directly because this would be subject to post-treatment bias. We will report descriptive statistics about the subgroups of people who applied for SSI benefits that were already receiving SNAP, who applied for SNAP at the same time they applied for SSI, and who neither submitted a new SNAP application nor were already receiving benefits. Descriptive statistics will allow us to get a better understanding of who belongs to these groups.[3]

Limitation of letters:
One limitation of the original RCT trial study design (and with all other conventional mail-based RCT trials) is that we cannot confirm whether recipients opened and read the letters. Hence, our analysis is based on Intention-to-Treat. Another limitation is that SSA does not know SSI eligibility with certainty until after an individual applies, since they do not have information on individuals'

[3] Relatively few OASI recipients receive other public assistance (2.1%) Characteristics of Noninstitutionalized DI, SSI, and OASI Program Participants, 2016 Update (ssa.gov) (Table 11).

assets or full income. Therefore, letters are sent to “likely eligible beneficiaries” following a preliminary assessment of people’s eligibility made before sending out letters.

Link to an Analysis Code/Script:
The code for this project will be housed on our github page.

References

Backlund, Eric, Paul D. Sorlie, and Norman J. Johnson. 1999. "A comparison of the relationships of education and income with mortality: the National Longitudinal Mortality Study." *Social Science & Medicine* 49(10): 1373-1384.

Baicker, K., Taubman, S. L., Allen, H. L., Bernstein, M., Gruber, J. H., Newhouse, J. P., ... & Finkelstein, A. N. (2013). The Oregon experiment—effects of Medicaid on clinical outcomes. *New England Journal of Medicine*, *368*(18), 1713-1722.

Courtin, E., Aloisi, K., Miller, C., Allen, H. L., Katz, L. F., & Muennig, P. (2020). The Health Effects Of Expanding The Earned Income Tax Credit: Results From New York City: Study examines the health effects of the New York City Paycheck Plus program that increases the Earned Income Tax Credit for low-income Americans without dependent children. *Health Affairs*, *39*(7), 1149-1156.

Dow, W. H., Godøy, A., Lowenstein, C., & Reich, M. (2020). Can labor market policies reduce deaths of despair?. *Journal of health economics*, *74*, 102372.

Gerber, A. S., Green, D. P., & Shachar, R. (2003). Voting may be habit forming: evidence from a randomized field experiment. *American journal of political science*, *47*(3), 540-550.

Goldin, J., Lurie, I. Z., & McCubbin, J. (2021). Health insurance and mortality: Experimental evidence from taxpayer outreach. *The Quarterly Journal of Economics*, *136*(1), 1-49.

Heflin, C. M., Ingram, S. J., & Ziliak, J. P. (2019). The effect of the supplemental nutrition assistance program on mortality. *Health Affairs*, *38*(11), 1807-1815.

Hemmeter, J., J. Phillips, E. Safran, and N. Wilson. (2020). Communicating Program Eligibility: A Supplemental Security Income Field Experiment. https://bit.ly/3qOgPhT

Hemmeter, J. and P. S. Davies. 2019. “Infant Mortality Among Supplemental Security Income Applicants”. *Social Security Bulletin* 79(1).

Herd, P., Schoeni, R. F., & House, J. S. (2008). Upstream solutions: does the supplemental security income program reduce disability in the elderly?. *The Milbank Quarterly*, *86*(1), 5-45.

Lopez, L., Hart, L. H., & Katz, M. H. (2021). Racial and ethnic health disparities related to COVID-19. *Jama*, *325*(8), 719-720.

Millett, G. A., Jones, A. T., Benkeser, D., Baral, S., Mercer, L., Beyrer, C., ... & Sullivan, P. S. (2020). Assessing differential impacts of COVID-19 on black communities. *Annals of epidemiology*, *47*, 37-44.

Price-Haywood, E. G., Burton, J., Fort, D., & Seoane, L. (2020). Hospitalization and mortality among black patients and white patients with Covid-19. *New England Journal of Medicine*, *382*(26), 2534-2543.

Sandefur, R. L., & Smyth, A. (2011). Access across America: First report of the civil justice infrastructure mapping project. *Available at SSRN 1962790*.

Tchetgen, E. J. T., Walter, S., Vansteelandt, S., Martinussen, T., & Glymour, M. (2015). Instrumental variable estimation in a survival context. *Epidemiology (Cambridge, Mass.)*, *26*(3), 402.

U.S. Social Security Administration. (2019). *Actuarial Life Table*. https://www.ssa.gov/oact/STATS/table4c6.html

Yao, Li, and Stephanie A. Robert. 2008. "The contributions of race, individual socioeconomic status, and neighborhood socioeconomic context on the self-rated health trajectories and mortality of older adults." *Research on Aging* 30(2): 251-273.