Analysis Plan

Project Name: Descriptive Study of Equity in the Distribution of the Emergency Rental Assistance Program

Date Finalized: 5/20/2022

Project Description

The first and second rounds of the Emergency Rental Assistance (ERA) program provide over $46 billion USD in funding to states, territories, and other entities ("grantees") to assist households that are unable to pay rent or utilities, with the goal of preventing eviction or housing instability in the wake of the pandemic. The program provides direct cash assistance to renters (and landlords) in order to assist with rent, utilities, and other housing related expenses. Renters must apply and be found eligible under three criteria to receive assistance:[1]

1. at risk of housing instability or homelessness;
2. experience of financial hardship related to COVID-19;
3. income that falls below an area-specific threshold.

We define the key question for equity in the distribution of this program as follows: to what extent did those who *received* ERA resemble those who were *eligible for* ERA?

High-Level Summary of Analysis

OES proposes to conduct a descriptive study of the extent to which those who received ERA resemble those who we predict to be eligible for ERA. We intend to study the two rounds of ERA as one program, since the eligibility requirements and implementation across rounds were very similar.

To estimate the degree to which the eligible population resembles the recipient population, we estimate and compare the "demographic profiles" of these two populations. For example, if we estimate that 21% of the eligible population are Hispanic or Latino and men and 21% of ERA recipients are Hispanic or Latino and men, then we would say that the two populations resemble each other in these dimensions.

Our estimation strategy thus focuses on estimating differences in the demographic profiles of **eligible** and **recipient** populations. We are interested in two kinds of estimates: what we refer to as the "outer cells" and the "inner cells". Table 1 below provides a hypothetical example of inner and outer cells. The outer cells present the demographic proportion of one dimension – e.g., 80% women and 20% men, 20% Hispanic or Latino and 80% not Hispanic or Latino. The inner cells represent intersections of these dimensions: cell A, for example, is the percentage of the population that identifies as women and Hispanic or Latina, and so forth. Understanding both the outer and the inner cells is important for understanding equity in the distribution of funding,

[1] See Table A1 in the appendix for more details on eligibility.

because heads of household who identify as Latina women, for example, may face challenges specific to that particular identity.

Table 1: An example of a demographic profile featuring inner and outer cells

	Women	Men	*Ethnicity outer cells:*
Hispanic or Latino/a	A	B	20%
Not Hispanic or Latino/a	C	D	80%
Gender outer cells:	80%	20%	100%

Table 2 below summarizes all of the demographic characteristics that we intend to estimate, the data sources we intend to use, and our intended methodologies for estimation. Broadly speaking. our estimation procedure involves three major stages described in this document:

1. **Estimate, at the state and national level, the demographic profile of the eligible population**:
   - This involves combining three Census datasets — the 2015-2019 American Community Survey (ACS), the 2020 Pulse Survey, and the 2020 Current Population Survey (CPS) — using a variety of reweighting and extrapolation techniques to leverage different strengths. Almost all demographic cells can be directly estimated using these datasets. The CPS and ACS do not, however, measure gender identity, they only measure sex as a binary variable. We rely on the Pulse survey for estimates of non-binary gender identification.
2. **Estimate, at the state and national level, the demographic profile of the recipient population:**
   - This step involves a combination of approaches. We estimate outer cells and inner cells that grantees reported on using Treasury's quarterly compliance reports from ERA grantees, with some imputation for missingness. We estimate inner cells that were not reported using ecological inference techniques.
3. **Estimate, summarize, and test the statistical significance of differences in the demographic profiles of the recipient and eligible populations**
   - We conduct a limited number of statistical tests to determine whether any observed divergences in the demographic profiles can be distinguished from statistical noise arising from the data and estimation techniques.

Table 2: Demographic profiles of interest in this study and how we plan to estimate them

<table>
<tr><th colspan="2">Demographic characteristic</th><th>How we estimate among eligible population</th><th>Reported by grantees?</th><th>How we estimate among recipient population</th></tr>
<tr><td>Race</td><td>% American Indian or Alaska Native, Asian, Black or African American, Mixed, Native Hawaiian/Pacific Islander, White</td><td rowspan="9">EITHER

reweighting 2020 Pulse using 2015-2019 ACS

OR

using 2020 CPS</td><td rowspan="4">Almost always reported (requires minimal imputation)</td><td rowspan="4">Averaging and imputation using quarterly compliance data</td></tr>
<tr><td>Gender</td><td>% Women, Men, Non-Binary</td></tr>
<tr><td>Ethnicity</td><td>% Hispanic or Latino, not Hispanic or Latino</td></tr>
<tr><td>Income</td><td>% less than 30% of the Area Median Income (AMI), 30-50% of the AMI, and 50-80% of the AMI</td></tr>
<tr><td>Race intersected with gender</td><td>% American Indian or Alaska Native men, Asian men, Black or African American men, Mixed men, Native Hawaiian/Pacific Islander men, White men, American Indian or Alaska Native women, Asian women, Black or African American women, Mixed women, Native Hawaiian/Pacific Islander women, White women, American Indian or Alaska Native non-binary, Asian non-binary, Black or African American non-binary, Mixed non-binary, Native Hawaiian/Pacific Islander non-binary, White non-binary</td><td rowspan="2">Not reported (estimated)</td><td rowspan="2">Estimated using ecological inference on quarterly compliance data, ACS, Pulse, and geocoded transaction data</td></tr>
<tr><td>Ethnicity intersected with gender</td><td>% Hispanic or Latino men, not Hispanic or Latino men, Hispanic or Latino women, not Hispanic or Latino women, Hispanic or Latino non-binary, not Hispanic or Latino non-binary</td></tr>
<tr><td>Race intersected with AMI</td><td>% American Indian or Alaska Native <30% AMI, Asian <30% AMI , Black or African American <30% AMI, Mixed <30% AMI, Native Hawaiian/Pacific Islander <30% AMI, White <30% AMI, American Indian or Alaska Native 30-50% AMI, Asian 30-50% AMI, Black or African American 30-50% AMI, Mixed 30-50% AMI, Native Hawaiian/Pacific Islander 30-50% AMI, White 30-50% AMI, American Indian or Alaska Native 50-80% AMI, Asian 50-80% AMI, Black or African American 50-80% AMI, Mixed 50-80% AMI, Native Hawaiian/Pacific Islander 50-80% AMI, White 50-80% AMI</td><td rowspan="3">Mostly reported (requires substantial imputation)</td><td rowspan="3">Estimated using imputation with quarterly compliance data</td></tr>
<tr><td>Ethnicity intersected with AMI</td><td>% Hispanic or Latino <30% AMI, not Hispanic or Latino <30% AMI, Hispanic or Latino 30-50% AMI, not Hispanic or Latino 30-50% AMI, Hispanic or Latino 50-80% AMI, not Hispanic or Latino 50-80% AMI,</td></tr>
<tr><td>Gender intersected with AMI</td><td>% Women <30% AMI, Men <30% AMI, Non-Binary <30% AMI, Women 30-50% AMI, Men 30-50% AMI, Non-Binary 30-50% AMI, Women 50-80% AMI, Men 50-80% AMI, Non-Binary 50-80% AMI</td></tr>
</table>

Pre-Registration Details

A draft of this Analysis Plan was posted to the private OES Github repository before receipt of the full set of data on individual transactions from grantees to households for quarters 1-4 in 2021 ("quarterly transaction data"). This version of the analysis plan was written after having access to the data, so it should not be considered blind to outcomes. This Analysis Plan will be posted on the OES website at oes.gsa.gov.

Note that we include both confirmatory analyses and exploratory analyses in this analysis plan. However, only the confirmatory analyses will be re-analyzed by a second analyst who is blind to the results in the first analysis, in accordance with the OES project process.

Confirmatory Analysis

Below, we provide more detail on the methods and specific datasets we plan to use to estimate demographic profiles of recipient and eligible populations. We address here a few key overarching considerations:

- *At what geographic level do we assess differences in demographic profiles*? Our main analyses will focus on differences in demographic profiles at the state level.[2] Note that state-level estimates should not be thought of as specific to *states as grantees*. For example, the state of Texas administered an ERA program and so too did a number of sub-state grantees (e.g., cities and counties) located in Texas. The demographic profile of recipient populations for Texas would include all of the recipients who received funds from the state-level Texas program as well as county and city-level programs, including any households who might have received payments from both. These results will be summarized by computing a population-weighted average of the state level estimates.
- *How do we quantify receipt of ERA at the individual level?* For the purposes of our main analyses, we consider ERA receipt in binary terms: a household either received any payment at some point during the program or they did not. We do not consider equity in the number of payments or total dollar amount received because, although we can measure the total amount of funds a household received (e.g., by summing the total value of disbursements to that household across months), we are not able to measure the denominator for these households, or how much money these households needed to reduce their housing instability.
- *Does our analysis focus on all household members, the household as a unit, or heads of a household?* For the purposes of our main analyses, we consider heads of households only. In principle, the members of households who received ERA could also be considered part of the recipient population. Grantees who collected data on non-head of households did not

[2] While some data does exist at lower levels of aggregation, we are still investigating issues around whether we can derive quality estimates for such levels. These issues include differential sparsity (some data sources on eligible populations do not cover all counties equally well due to random sampling) and differential overlap (some grantees have geographies that do not overlap statistical geographies, which could produce differences in demographic profiles that are purely methodological artifacts). These issues are greatly mitigated / non-existent at the state-level.

report this to Treasury, however, and this analysis is based on grantee reports. Any main analyses will analyze recipient households in terms of the demographic characteristics of the head of household.

- *Bootstrapping to model variance from sampling, imputation, and estimation.* To estimate the variance of the various statistics we use bootstrap resampling. A complication in this context is that we do not derive a single point estimate, because we are also using stochastic multiple imputation for missing data in various parts of the analysis described below. We follow an approach outlined in Schomaker and Heumann (2018) shown to have good properties in terms of low bias and correct coverage, which they refer to as "Multiple Imputation (MI) Boot." This involves first generating M imputed datasets. For each of the M datasets, a point estimate of the statistic is calculated, and B bootstrap samples are drawn. These bootstrap samples are used to estimate the standard error of the point estimate in each of the M imputed datasets. This yields M point estimates and M standard errors. We use these to calculate standard multiple imputation standard errors and confidence intervals. If computationally feasible, our plan is to set M = 100 and B = 2000, which means estimating every statistic 200,000 times.

1. Estimate of the demographic profile of the *eligible* population

We aim to produce, for each state, an estimate of the proportion of people who were potentially eligible for ERA who fell into each unique intersection of demographic attributes reported by grantees, along with the associated standard error for this estimated proportion. To do so, we will combine Census datasets that allow us to derive state-level estimates of the demographics of those potentially eligible for ERA.

In Table 3 we describe the datasets we will utilize. All of the datasets contain information on criteria that went into ERA eligibility and on demographics and are designed to provide state-level estimates. However, they present different advantages and disadvantages in terms of their ability to provide representative, state-level estimates of the demographic profile of those who were potentially eligible for ERA.

For our purposes, the key distinctions are as follows:

- The 2015-2019 5YR American Community Survey data (ACS) are produced with sampling strategies that are well understood and generally considered of high quality but contain no direct measures of COVID-related hardship;[3]
- The 2020-2021 Census Pulse directly measures COVID-related financial hardship and risk of housing instability, but does not appear to provide representative estimates due to challenges with sampling during the pandemic; and
- The Current Population Survey (CPS) provides measures of financial hardship and housing instability during the pandemic, and employs a sampling strategy that is well understood, but has a small sample size relative to the other two surveys.

In other words, the 2015-2019 ACS IPUMS data is high quality, but lacks direct measures of

[3] The 2020 ACS data has an advantage over the 2019 ACS in that it measures eligibility more closely to the pandemic's onset and provides a more recent snapshot of the demographic composition of states at the onset of ERA disbursement. However, this comes at the disadvantage of greater nonresponse bias and difficulties knowing the population it generalizes to.

eligibility, whereas the Pulse survey contains such measures but is deemed of low sampling quality by the Census Bureau. The CPS, for its part, has neither of these deficiencies[4] but it is smaller.

To make the most of these datasets we propose to take **three** approaches described below. In each approach, we will produce standard errors using weighted bootstrap resampling and Census-produced survey weights. We will use the estimates from the approach that produces the smallest standard errors as our confirmatory analysis, and report the results of the other methods in our robustness analyses.

All approaches involve subsetting to heads of household who pay rent. They also require labeling respondents as "likely eligible" for ERA based on responses to questions about pandemic-related hardship, household income as a proportion of the Area Median Income (AMI), and risk of housing instability or homelesness.

Note on eligibility criteria differences between ERA1 and ERA2. For the purposes of the main analysis, we use the eligibility criteria of ERA2 for determining whether a given respondent is eligible. While the criteria differ in several ways (see Table A2 in the appendix), for the purposes of identifying likely-eligible respondents, the key difference resides in whether household-size adjustments are made for income eligibility. As described below, we will report a robustness test in which ERA1 eligibility criteria are employed at this step.

The first of the three approaches involves simply using CPS data and is the simplest methodologically. The second two involve using the high-quality sampling strategy of the ACS to obtain better state-level estimates from the Pulse.

A. **CPS-Only Estimation**: Every renter head of household in the survey is defined as likely eligible for ERA if their household income falls below ERA2 AMI household-size adjusted thresholds AND EITHER: they experienced some form of financial hardship in the form of unemployment over the preceding year OR they had to move to their present home due to eviction or financial hardship OR they were unable to work because their employer lost business due to the coronavirus pandemic OR the pandemic prevented them from looking for work.[5] Among that group, we estimate demographic statistics mirroring those reported in Treasury's quarterly compliance reports and described in Table 2 above. We use the pre-constructed CPS survey weights to derive these estimates.
B. **Post-Stratification of 2020 Pulse using 2015-2019 ACS as Baseline:** The first step is to define respondents as likely eligible for ERA if their household income falls below ERA2

[4] The CPS and ACS do not generalize to exactly the same population, however, we do not anticipate this to be of any consequence in the ERA context. Whereas the ACS is designed to generalize to all persons in the United States, the CPS is primarily designed to estimate features of adults not in the armed forces or in an "institution" (e.g., prison, hospital, elderly care). When using these datasets, however, we will be subsetting to renter heads-of-household who are eligible for ERA, who will almost never be youths, the institutionalized, or members of the armed forces.
[5] This "or" logic departs from the formal ERA2 eligibility criteria, which require that households satisfy all three characteristics—AMI below threshold; financial hardship; risks of housing instability—to qualify for assistance. However, because the financial hardship criteria in ERA are more expansive than unemployment—for instance, the hardship can encompass reduced working hours—we use the or logic to capture households who remain employed but experience housing instability.

AMI household-size adjusted thresholds and EITHER: the respondent applied for Unemployment Insurance or other social security benefits since March 2020 OR they or someone in the household lost or expect to lose employment since March 2020 OR they report struggling to make rent payments OR they believe it is likely they will be evicted within the next two months. The second step requires identifying all variables that are common to the Pulse and ACS, including but not limited to the demographic variables that interest us for the purposes of analyzing equity. We then estimate the proportion of respondents in the ACS and Pulse who occupy each possible permutation of the common variables. The ratio of the proportions between ACS and one of the other datasets provides a weight that can be used to adjust for non-representativeness.[6] The weights constructed in this manner can be used to produce estimates of the profiles in Table 2 from a Pulse dataset made to resemble the people in the ACS dataset.

C. **Extrapolating to the 2015-2019 ACS using a Model fit with the 2020 Pulse:** The second method fits a statistical model to the Pulse data that uses the common variables identified in approach B to predict the "likely eligible" or "likely not eligible" labels among renting heads of household. The model is then used to predict, using the common variables in the ACS data, whether a respondent in the ACS data, for whom we do not have these labels, is likely eligible or not. We will fit the model using randomly-sampled training data and a suite of flexible binary classifiers, including regularized regression approaches (e.g., LASSO; ridge) and tree-based classifiers (e.g., decision tree; random forest; AdaBoost). Performance will be evaluated using recall as our performance metric in the held-out testing set. We focus on recall because we care more about minimizing false negatives (people who are eligible for ERA but who are predicted as not eligible) than minimizing false positives (people who are ineligible for ERA but who are predicted as ineligible).[7] Once we have classified ACS respondents as eligible or not using the predictive model, we estimate the profiles in Table 2 by calculating the proportions of predicted-eligible respondents occupying each unique intersection of equity-relevant attributes. This step will employ ACS pre-built weights.

[6] For example, suppose that 18-34 year-old, white, male, employed, non-Latino head of household renters comprised 2% of the ACS and 4% of the Pulse survey. Such respondents would receive a weight proportional to 2%/4% = ½.

[7] Because the models generate continuous predicted probabilities that we then translate into a binary label of "likely eligible" or "not likely eligible," depending on the prevalence of the "likely eligible" label, we will use ROC curves to examine whether we will modify the threshold for converting continuous predictions into binary above the default value of 0.5.

Table 3: Microdata used to derive state- and national-level estimates of demographic profile of those eligible for ERA

Dataset name	Temporal coverage	Key variables
Integrated Public Use Microdata Series - American Community Survey (IPUMS ACS)	The most recent datasets are the 2019 and 2020 surveys.[8] The US Census recommends against using the 2020 1 YR ACS PUM data for evaluation purposes due to its reliance on experimental weights.[9]	Has measures for household income, rental and utility costs, experience of financial hardship over the previous year.
Integrated Public Use Microdata Series - Current Population Survey (IPUMS CPS) microdata	Data are released at a monthly time interval, and the most recent published estimates are from January 2022. Therefore, we will likely merge months that shortly precede the ERA disbursement window.	Has measures for household income, experiences of job loss that month, renter versus owner and type of unit[10], experience with housing instability over the previous year, and financial hardship in the form of unemployment.
Census Household Pulse	Covers period from April 2020 to present (early 2022). New data is released frequently.	Has measures for household income, rental burden, impact of COVID-19 on finances, perceived risk of eviction, and ERA receipt and application.

2. **Estimate the demographic profile of the *recipient* population**

We aim to produce, for each state, an estimate of the demographic attributes of ERA recipient heads of household, along with the associated standard error for this estimated proportion. To do so, we will apply ecological inference and imputation techniques to combine grantee reports on recipient demographics and transaction-level spending.

Phase one of our estimation procedure produces a set of estimates of the demographic profile of those who were eligible for ERA. The second phase of our estimation procedure seeks to provide corresponding estimates for the recipients of ERA. Estimating the demographic profile of the *eligible* population is relatively straightforward, as we possess microdata on demographics. When it comes to understanding the characteristics of *recipients*, however, we face two challenges:

[8] The 2021 ACS will not be released until March of 2023.
[9] See https://www.census.gov/programs-surveys/acs/library/flyers/flow-chart.html and https://www.census.gov/newsroom/blogs/random-samplings/2021/10/pandemic-impact-on-2020-acs-1-year-data.html
[10] Unlike other variables, housing data in the CPS is collected during the Annual Social and Economic Supplement (ASEC), which is only fielded yearly.

- First, for the most part, grantees were only required to report data on outer cells. That means we have to employ statistical techniques to infer proportions of inner cells.
- Second, grantee reports that were required are incomplete, as not all grantees reported in all quarters for all expenditure categories. In practice, this means we need to employ techniques to infer missing data points that might otherwise bias our inferences.

To address these challenges, we employ a combination of imputation and ecological inference techniques. We base our analysis on three datasets built from reports submitted by grantees to Treasury on a monthly and quarterly basis:

1. Quarterly compliance reports
    - **a.** Certain grantees were required to report,[11] for every quarter of ERA I and ERA II, the number of households receiving different kinds of assistance broken down by income level, race, ethnicity, and gender. In principle, this data should contain, for every grantee subject to reporting and in every quarter in which they spent, a breakdown of ERA I and ERA II funds spent across the following demographic categories:[12]
        - i. Ethnicity: data not collected, declined to answer, Hispanic or Latino/a, not Hispanic or Latino/a
        - ii. Gender: data not collected, declined to answer, female, male, non-binary
        - iii. Race: American Indian or Alaska Native, Asian, Black or African American, decline to answer, data not collected, Mixed, Native Hawaiian/Pacific Islander, White
        - iv. Race, gender, and ethnicity broken down by three tranches of area-specific income levels: less than 30% of the Area Median Income (AMI), 30-50% of the AMI, and 50-80% of the AMI
2. Transaction-level data
    - a. In principle, all grantees subject to a reporting requirement must provide data on individual transactions to recipients for every quarter in which funds were disbursed. This data does not include any demographic information. It includes, among other fields, the following:
        - i. Address of recipient household
        - ii. Amount received
        - iii. Date sent
        - iv. Type of assistance
3. Monthly compliance reports
    - a. In principle, all grantees subject to a reporting requirement must provide data on the count of unique recipients that received any assistance.

[11] States, counties, metropolitan cities, and territories were required to report quarterly demographic data, whereas - Tribes, Tribally Designated Entities, and the Department of Hawaiian Home Lands were not.
[12] When estimating demographic proportions, the proportions calculated in this report will be calculated using data for households with reported demographic information, and excluding data on cases of declined to answer or data not collected.

As Table 5 in the appendix outlines, only 342/418 = 82% of the grantees that were required to submit quarterly reports submitted at least one report on at least one demographic category in the quarterly compliance data. Some 48 grantees only submitted transaction-level reports and another 29 did not submit transaction-level data or quarterly compliance reports. We therefore face multiple forms of missingness and different challenges depending on which demographic cell we are trying to estimate. The types of missingness and strategies to address it are spelled out in detail in the section on missing data. We provide a high-level overview of those strategies here, insofar as they help us to estimate partially observed outer cells, unobserved inner cells, and partially observed inner cells of the demographic profile of recipients.

Estimation of Partially Observed Outer Cells

Recall that outer cells, such as the proportion of ERA recipients who identify as Black, are partially observed insofar as some grantees report data on these outer cells in some quarters. For estimating outer cells, we take the following approach:

- For grantees who have at least one quarterly compliance report detailing all demographic categories (referred to as Category 4 in Table 5), we simply take the average of all available quarterly reports
- For grantees who report some but not all demographics in at least one quarterly compliance report (Category 3), we impute missing categories based on a mix of relying on the mutually-exclusive nature of some demographic categories to back out missing counts from total recipients and observed counts and, for remaining missingness, multiple imputation
- For grantees who report no demographic data, either because they do not submit any quarterly compliance reports or because all compliance reports they submit are missing this data, but who do submit transaction-level data (Category 2), we plan to impute the demographics of recipients based on transaction data and ACS tract-level demographics
- For the 29 grantees about which we do not know anything, because they did not submit demographic data or transaction data (Category 1), we simply drop them from the analysis

The combination of these techniques gives us estimates of the outer cells of recipients. As with the eligible population demographic profile estimates derived in phase one, we will use sample weighted bootstrapping to estimate standard errors for these outer cell proportions.

Estimation of Unobserved Inner Cells

With outer cell estimates in hand, we can estimate inner cells that we cannot observe directly using ecological inference methods. To gain an intuition for how this works, consider Table 1 above, for example, which is based on publicly available aggregate data for Q1-Q4 2021.[13]

We know, for example, that because only 20% of the recipients are Hispanic or Latino/a, there are not more than 20% Hispanic or Latino men and not more than Hispanic or Latina women (or each inner cell—A and B—is constrained to sum to 20%). While we could then assume that the Hispanic or Latino/a recipients are evenly split between men and women, we also know that the total number of recipient men is only 20% across ethnicities, and so the split is likely *not* even within Hispanic or Latino/a recipients—it may skew more toward women. Knowing the outer cells of a table thus enables us to make an educated guess about the likely values of the inner cells, which forms the basis of the ecological inference framework.

[13] https://home.treasury.gov/system/files/136/Q1-Q4-2021-ERA-Demographic-Data.xlsx

While one form of ecological inference uses the constraints imposed by these outer cells to guess values for the inner cells, in the ERA context, we may improve our "educated guess" by bringing in auxiliary data to improve our prediction of the inner cells. That is, when we know that 20% of recipients in a grantee's demographic profile are Hispanic or Latino, and are trying to infer how many are Hispanic or Latino women, we can leverage two additional sources of data to improve our predictions.

First are contextual characteristics of where ERA recipients are located that we aggregate to the state level. The appendix describes how we geocode the ERA transaction data to create a dataset with (1) the number of households receiving funds in a tract and (2) the demographics of census tracts. We will then aggregate to the state-cell-level by estimating correlations between tract-level demographics and number of households assisted for each demographic inner cell. These state-cell-level correlations will become predictors in our ecological inference model.

Into this state-cell-level dataset, we will merge two additional sets of variables. First, the row and column outer cell estimates corresponding to that cell and state, as estimated using the procedures above. Second, state-level estimates of inner cells derived from responses to the Census pulse survey, which contains a question on ERA receipt.[14]

This combined data can be used to fit a multinomial dirichlet model for "ecological inference in Row x Column (R x C) tables," which is implemented in several R packages.[15] Essentially, the model works by modeling the counts in the column outer cells as the outcome of a random multinomial process governed by probabilities contained in the inner cells and counts in the rows. The probabilities correspond to the inner cell values we wish to guess but cannot observe, so these are modeled using a dirichlet prior (the dirichlet distribution takes the form of a vector of numbers between 0 and 1 that are constrained to sum to 1). Interestingly for our purposes, the model can include predictors of the probabilities, such as the Census pulse inner cell estimates and the correlations between transactions and demographics.

Provided this model does a reasonable job at prediction (i.e., better than random guesses constrained by the margins), we will use it to estimate the inner cells of recipient demographic profiles, even though we cannot observe these directly. This approach constitutes an improvement over noisier methods of demographic imputation that do not make use of all available information, such as random imputation based on known population quantities.

Estimation of Partially Observed Inner Cells

As noted above, some inner cells of demographic profiles are observed, such as the proportion of ERA recipient heads of household who identify as Black and have a household income below 30%

[14] HSE7 Have you or anyone in your household applied for emergency rental assistance through your state or local government to cover your unpaid rent or utility bills?

o My household applied and received assistance
o My household applied and is waiting for a response
o My household applied and the application was denied
o My household did not apply

[15] See here for one approach, for example:
https://search.r-project.org/CRAN/refmans/eiPack/html/ei.MD.bayes.html

of AMI. We plan to estimate these inner cells only for grantees who exhibit category 3 or 4 missingness: e.g., for grantees whose outer cells are observed. In those cases, we will use the same imputation methods used for all category 3 missingness in order to estimate those inner cells.

3. Estimate, summarize, and test the statistical significance of differences in the demographic profiles of the recipient and eligible population

We will estimate disparities by subtracting the "eligible" proportion of some intersection of attributes or some outer cell estimated in step 1 from the corresponding "recipient" proportion estimated in step 2. This part is straightforward once the demographic profiles have been estimated. However, complications arise when one tries to make an inference about whether these differences should be attributed any particular significance, or whether they are instead statistically insignificant in the sense that such differences were likely to arise simply due to noisiness in the data, even if the allocation procedure for allocating funds to recipients was perfectly demographically proportional to their size in the eligible population.

Of course, testing each individual difference would involve a very large number of tests and pose a high risk of false positives. If we were to focus on finding *any* statistically significant effect, this would pose a multiple comparisons problem. Instead, we propose to use a single test of the hypothesis that the process for allocating funds to different, intersecting demographic groups was perfectly proportional to their size in the eligible population. The test we propose to use provides us with a p-value that represents a useful probability, namely: the probability that the demographic profile of the recipient population would depart as much as it does from the demographic profile of the eligible population if recipients were drawn randomly with equal probability from the eligible population. This null hypothesis thus directly answers the research question of whether the recipients resemble the eligible.

To conduct the test, we assess the cumulative probability of observing the following test statistic if the null hypothesis is true:

$$\sum_{i=1}^{k} \frac{(m_i - np_i)^2}{np_i}.$$

This is a chi-square test statistic. Here, i indexes the cell (one cell for each intersection of attributes), k is the total number of cells (the total number of attributes), m is the "observed" count in a given cell (for us, the number of recipients in the i'th intersection of attributes), n is the total number of observed individuals (for us, the total number of recipients), and p is the proportion of individuals that would fall into that cell under the null hypothesis (for us, it is the proportion of *eligible* participants that have that intersection of attributes). In the appendix, we describe in more detail how we obtain the null distribution and observed value of this test statistic in order to obtain p-values. In simulations of this method, we noticed a tendency to over-reject in small samples. The appendix thus includes code that illustrates how we will adjust our alpha (false positive rejection rate) to account for any overrejection in this test, although we do not expect this to be an issue given the large sample size.

Having estimated the differences between the demographic profiles of recipient and eligible populations, then tested the statistical significance of the overall distance between the two profiles, we plan to use simulation studies in order to contextualize and interpret the findings.

An interesting implication of taking an intersectional approach to the analysis of inequity is the ability to see how hypothetical interventions that raise the receipt rate among one or more demographic groups, while holding the overall budget constant, affect the receipt rate of other groups. By simulating how interventions to redistribute to a given group impacts other groups, we hope to provide nuanced recommendations about the most effective strategies to address inequity in future rental assistance programs.

Exploratory Analysis

If it is possible to conduct the main confirmatory analyses, the exploratory analysis will simply be reported as supplementary analyses. However, if it is not possible to conduct the main analyses due to data issues, this analysis will function as fallback option.

Eligibility-Matched Area-Level Analysis
This analysis takes census tracts as the unit of analysis and cannot answer the main research question (the extent to which those who received ERA resemble those who were eligible). Instead, it answers the following question: among areas predicted to have similar proportions of people who were eligible for ERA, what are the correlations between demographic characteristics of the area and levels of receipt of ERA?

Note that, since this analysis makes no attempt at ecological inference, these analyses only provide valid inferences at the area-level. Attempts to infer individual-level processes from area-level correlations can be subject to bias from an ecological fallacy. For example, it is entirely possible that tracts with higher percentages of Black households receive more awards than tracts with lower percentages of Black households, but this does not mean Black households are more likely to receive ERA. It could be the case that within such tracts, Black households are *less likely* to get ERA than their White counterparts. Therefore, it is important to make inferences about areas only.

To investigate this question, our outcome will be the census-tract level count of unique households that received any kind of ERA assistance. We will use this outcome as the lefthand side in several multiple regression models. All regressions will condition on an estimate of the number of people in a tract who were eligible for ERA from the ACS tract-level estimates. For example, variable B25106 records, for 6 income categories, how many renters were paying less than 20%, 20-29%, or at least 30% of their income in housing costs. One good estimate of the eligibility for ERA in a given tract might be the minimum number of renters who fell below 80% of AMI and spent 30% or more on housing. Importantly, this estimate of eligibility for ERA, although *implicitly correlated* with demographic attributes (e.g., younger households may be more likely to be renters), does not explicitly take into account race, ethnicity, gender, age, or other equity-relevant attributes.

We then estimate partial correlation coefficients relating to the demographic composition of the tract. For example, we add the variable "% renter households that have women heads of household". If the partial correlation is positive and statistically significant, for example, we will infer that all households in tracts that also had more women heads of households received more funding overall. Note that we refrain both from assuming that this means women heads of household received more ERA funds and from inferring that the presence of more women households *caused* greater receipt at the tract level. The example below illustrates why one needs to be careful: whereas tract 1 had more women renters eligible for ERA (25% vs. 15%) and had a lower rate of funding (2.5% vs 7.5%), women in both areas were equally likely to receive funding conditional on the allocation to that area (5/25 = 15/75 = 20%).

Table 4: Example of a program in which women heads of household (HoH) were equally likely to receive ERA, conditional on tract, but tracts with more women HoH received less spending

Tract	Predicted number of renters that need ERA	Predicted number of renters in-need who are women HoH	Number of renters that got ERA (assumed to be perfectly observed)	Number of ERA recipients who are women HoH (assumed unobserved)
1	1000	250	25	5
2	1000	150	75	15

Missing Data

There are various forms of missing data that we address to minimize the chance of bias related to missingness. We first subset the data for each grantee to the quarters in which they report assisting one or more households.[16] Then, Table 5 below outlines four types of missing demographic data at the grantee level, how that form of missingness is detected, and strategies we plan to use to address in main or robustness checks.

[16] The reports of quarters that assistance was delivered derive from the quarterly reports, the monthly reports, or the transaction data that grantees submit to Treasury. Counts are based on grantee reports on demographics for the spending category "Number of unique households that received ERA assistance." The counts for reports on demographics for other categories of spending, such as the demographics of recipients below income thresholds (e.g., less than 30% of area median income), are roughly similar.
As there are numerous cases of multiple reports submitted by grantees for a given quarter, we select the report with the least number of missing attributes.

Table 5: Categories and Counts of Grantees by Missingness Status

Name of category	Grantee monthly report	Grantee quarterly report	Grantee transaction data	Number of grantees (out of 418)	How we will address
Missing all dem. records and not imputable (category 1)	Submitted so we know grantee exists/count of recipients	Either none submitted or submitted but w/out any demographics	None submitted so we cannot impute demographics based on area-level characteristics	29	Do not impute but report these grantees in an appendix
Missing all dem. records and imputable using area-level attributes (category 2)	Same as above or not needed since grantee existence shown in transaction data	Either none submitted or submitted but w/out any demographics	Submitted, defined as 1+ transaction recorded and attributed to that grantee	48	Discussed below in "Details on strategies for addressing category 2 missingness (missing all aggregated demographics but observe transactions)"
Observe some demographic attributes but not others (category 3)	Not needed since existence shown in quarterly data	Grantee has 1+ report with demographics but certain categories are blank	Not needed	45	Discussed below in "Details on strategies for addressing category 3 missingness (observe some demographics but not others)"
Observe all demographic attributes (category 4)	Not needed since existence shown in quarterly data	Grantee has 1+ report with demographics but certain categories are blank	Not needed	297	N/A (perfectly observed)

Details on strategies for addressing category 2 missingness (missing all aggregated demographics but observe transactions)

For some grantees, we are missing all *household-level* demographics in the quarterly compliance reports. We therefore use the address-level data to impute "possible demographics" of households based on the Census tracts where they are located. To make the example concrete, we discuss an example grantee—OES County—where the transaction data's "grantee" column indicates multiple transactions. Those transactions are geocoded to 10 different tracts. The Table below shows example rows and simplifies categories to Black or White for the purpose of illustration.

Table 6: Example of transaction-level data with tract-level demographic characteristics

Grantee	Address	Tract	Tract count Black / % Black	Tract count white / % white
OES county	100 main street unit 1	A	100 (25%)	300 (75%)
OES county	100 main street unit 10	A	100 (25%)	300 (75%)
OES county	200 main street	A	100 (25%)	300 (75%)
…				
OES county	1800 F street NW	B	50 (50%)	50 (50%)

To impute the % Black for the grantee OES county using these tract-level aggregates, as an example, we:

1. Sum the number of unique households present in each tract
    a. 3 households from tract A
    b. 1 household from tract B
2. The % Black at the grantee level is the recipient-weighted counts
    a. **Numerator:** 3 households * 100 Black in tract A + 1 household * 50 Black in tract B
        i. Result: 350
    b. **Denominator: (**100 Black in Tract A + 300 white in Tract A)*3 + 50 Black in tract B + 50 white in Tract B
        i. Result: 1300
    c. **Proportion:** 350 / 1300 ~ 27%

And so on for other grantees and other demographic categories.

Details on strategies for addressing category 3 missingness (observe some demographics but not others)

We will impute missing demographic data using two sequential strategies:

1. **Grouping demographic attributes into mutually exclusive categories and backfilling the counts for the missing category.** Some of the demographic categories reflect mutually exclusive attributes—e.g, grantees report that recipient households fall into one of three genders: man, woman, or non-bibary. In these cases, we will impute the missing category by (1) summing the observed counts for that general demographic attribute (e.g., count man + count non-binary) and (2) subtracting that sum from the total count of unique recipients

(e.g., count woman = total - (count man + count non-binary)).[17]

2. **Using multiple imputation to impute remaining missing values:** we will then use multiple imputation (implemented using Amelia in Stata or MICE in R) to impute the remaining missing counts. Multiple imputation will use two forms of observed data—demographic counts from grantees with *fully observed* demographic values; and demographic counts from a focal grantee for the observed categories—to impute the missing counts. However, because these counts are "guesses" rather than observed values, we need to ensure that our downstream variance estimation takes into account the greater uncertainty tied to these guesses. Our procedure for combining bootstraps with multiple imputation in order to derive point estimates and standard errors is described above.

Robustness Checks

In order to better understand and characterize the consequences of the analytic choices described above, we plan to conduct the following robustness checks.

Using ERA 1 eligibility criteria for AMI

The main analysis uses the ERA2 definition of eligibility based on household income relative to AMI, which contains adjustments for household size. We will conduct a robustness check where we calculate income eligibility without accounting for household size. This will likely result in some smaller households that would not qualify under the size-adjusted criteria being classified as eligible, as well as some larger households that qualify under the size-adjusted criteria being classified as *not* eligible. We will then investigate how our comparisons of eligible versus recipient demographics change with this alternate measurement of one input to eligibility.

Using alternative procedures for estimating eligible demographic profile

As described above we will employ one of three techniques to estimate the demographic profile of the population eligible for ERA. To understand how the selection of this technique over the others matters for inferences, we will also report results using the two other techniques.

Limitations & Challenges

Additional data quality issues with compliance reports

In addition to the main data quality issue discussed above—i.e., reports with counts of recipients but no demographic breakdowns; reports with some demographic categories reported but not others—there are duplicate reports with conflicting information and with no way to adjudicate which report is "final". There are numerous cases in which grantees submitted multiple reports for a given quarter. Since we cannot discern which report is correct, this analysis takes the submitted report with the least number of missing variables, by grantee quarter.

Additional limitations

---

[17] We will obtain the total count of recipients from demographic attributes where all categories are observed.

There are several limitations with the present analysis. We have discussed measurement limitations above. More fundamentally, perhaps, we will not know the reasons for any over or under-representation we find. Knowing *that* disparities in the demographics of those who were eligible and who received ERA exist does not tell us *what mechanisms* produced those disparities. Instead, we hope the present efforts complement the qualitative case studies of how grantees are approaching equity and serve as a complement to proposed evaluations that investigate the impact of specific application features (e.g., the adoption of self-attestation) on the composition of those receiving funds.

Appendix

Geocoding Transaction Data

Here, we describe the geocoding process in greater detail.

1. Clean the addresses to remove unit numbers and flagging non-geocodable addresses: the original transaction data contains unit information and other extraneous information that can prevent our ability to geocode the address. We also create indicators for addresses that definitively cannot be geocoded (e.g., grantees that put "HOMELESS", "REDACTED," and so on.
2. Use the Geocodio API (link here: https://www.geocod.io/features/api/) to geocode the unique cleaned addresses: we then use the Geocodio API, which accepts as its input an address string and batch geocodes up to 9999 addresses in a batch, to geo-locate each address to its point location (latitude and longitude).
3. Intersect the latitude and longitude with Census-provided, tract-level shapefiles: after we have each latitude and longitude, we use the Census Bureau API, and tidycensus wrapper in R, to pull tract-level shapefiles associated with the 2019 TIGER geographies. This tells us the tract in which the geocodable transactions are located.
4. Merge in tract-level characteristics from the 2019 ACS: as discussed above, we then merge in tract-level characteristics that correspond to the categories in the demographic compliance report (race/ethnicity; gender; different income bars) among renter-occupied units.[18]
5. Aggregate the counts from the tract to the grantee level for imputation of missing demographic compliance report information and to the state-level for ecological inference

[18] For most characteristics, the ACS provides a breakdown separately for owner-occupied units versus renter-occupied units. However, for some, the ACS does not distinguish based on renter status so we need to use the combined totals.

Table A1: Eligibility criteria for ERA1 and ERA2

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

Source: Department of the Treasury FAQs, Last updated on May 7, 2021

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

Additional details on null-hypothesis testing
To conduct the test, we assess the cumulative probability of observing the following test statistic if the null hypothesis is true:

$$\sum_{i=1}^{k} \frac{(m_i - np_i)^2}{np_i}.$$

This is a chi-square test statistic. Here, i indexes the cell (one cell for each intersection of attributes), k is the total number of cells (the total number of attributes), m is the "observed" count in a given cell (for us, the number of recipients in the i'th intersection of attributes), n is the total number of observed individuals (for us, the total number of recipients), and p is the proportion of individuals that would fall into that cell under the null hypothesis (for us, it is the proportion of *eligible* participants that have that intersection of attributes). Ignore for now that m and p are estimated and treat them as fixed for simplicity (we address this complication below).

A few notes on how the test statistic works. Suppose that the eligible and recipient population demographic profiles were exactly the same. In this case, the proportion, p, of eligible people in every cell would be exactly equal to the proportion of recipients in each cell, so np would be equal to m for every i. In this case, all of the differences in the numerator will be 0, and the value of the test statistic will be 0. Now imagine that the profiles did not resemble one another. This would mean that np were larger or smaller than m. As np grows larger or smaller than m, so too does the size of the test statistic, because it is taking a weighted sum of the squared differences of hypothesized and observed counts in each cell. Thus, the test statistic measures departures from the hypothesized model of proportions.

The null distribution that we want to obtain for our test is the one generated by randomly selecting n individuals from the k cells in the hypothesized proportions, p and then computing the test statistic each time. Because this procedure is quite easy to implement on a computer, we do not need to make reference to the chi-square distribution, which is a parametric approximation that is not always exact. Instead, we plan to use Monte Carlo methods to obtain the test statistic under the null distribution directly.

For example, suppose we had k = 4 demographic profile cells and 40 eligible potential recipients, split evenly among the cells. In that case, we would have p_1 = p_2 = p_3 = p_4 = ¼. Suppose that there were n = 4 recipients. The number of ways in which n recipients can be sorted into k cells is given by n+k-1 choose k-1. In our example, that means 7 choose 3 = 35. For each of these, we can calculate the chi-square test statistic. Of the 35 possible ways of putting four people into four cells, under the hypothesis p only one gives a chi-square of 0 (when exactly one person is in each cell), twelve give a chi-square of 2, six give a chi-square of 4, twelve give a chi-square of 6, and four give the most extreme value of our chi-square, 12 (this arises when one cell has four people in it and the rest have 0 — this is the most extreme departure from the hypothesized even split). So our null distribution can only take on five values, but these five values occur with very different probabilities.

If we saw that there were 0 recipients in m_1, 2 in m_2, and 1 in m_3 and m_4, this would represent a departure from the hypothesized proportions, which expect one person in each cell. But is this seeming departure statistically significant? In this case, we get a chi-square statistic of 2. The probability of obtaining a chi-square of at least 2 can be calculated by taking a large sample from the null distribution, and seeing how often a value this extreme occurs. The great thing about this method of hypothesis testing, aside from the fact that it provides one general test that speaks directly to our research question, is that it can handle very large n and k.

A complication with our approach is that p and m are estimated in steps 1 and 2, respectively. Recall that steps 1 and 2 produce a large number of bootstrapped replicates of p and m. Therefore, we must propagate this uncertainty into our null distribution. In addition to accounting for the way in which the random sampling of recipients from the eligible population in proportion to p could have produced different results, we also need to account for the fact that we might have observed different p and m based on the random samples and random modeling techniques we used. We describe the procedure for drawing inferences using bootstrapping and multiple imputation above.

One consideration is the potential tendency for overrejection when the multinomial test is used in small samples. According to our current understanding, there are three factors that affect the probability of falsely rejecting the null hypothesis (the False Discovery Rate – FDR) and the probability of correctly rejecting the null hypothesis (power). These are the number of cells (k), the sample size of eligible in each cell (p), and the sample size of recipients (n). In simulations conducted using the code below, we found that all test statistics exhibited rejection rates that were too high when using small sample sizes. A simple solution, also implemented in the code below, is to find an alpha level that provides the correct rejection rate for that sample size. At the sort of sample size we will be using, in the millions, we do not anticipate running into this issue, however.

Code used to investigate and address false positive rates in the proposed testing procedure

```
# helper to keep xmonte from producing message
# (this is really annoying when running multiple
# iterations)
quiet <- function(x) {
  sink(tempfile())
  on.exit(sink())
  invisible(force(x))
}

# function to identify new test level
find_new_alpha <- function(
  desired_alpha = 0.05, # the desired/original test level
  group1_freq,          # the frequency distribution for group 1
  group2_freq,          # the frequency distribution for group 2
  sims = 1000,          # no. of iterations to run
  statName = "Chisq",   # test to have xmonte estimate
  simulate = TRUE       # if FALSE will do param. chi-sq
) {

  # expected probabilities
  p <- group2_freq / sum(group2_freq)

  # simulate null
  sim.data <- lapply(
    1:sims,
    function(x) data.frame(
      group1 = as.vector(
        rmultinom(1, sum(group1_freq), p)
      ),
      group2 = as.vector(
        rmultinom(1, sum(group2_freq), p)
      )
    )
  )

  # run test
  cat("\nSimulating p-values (be patient).......\n")
  if(!simulate) {
    sim.p <- lapply(
      sim.data,
      function(data) suppressWarnings(chisq.test(
        data$group1,
        p = data$group2 / sum(data$group2)
      ))[[3]]
    )
  } else {
    sim.p <- lapply(
      sim.data,
      function(data) quiet(XNomial::xmonte(
        data$group1, data$group2, statName = statName
      ))[[5]]
    )
  }
  cat("\nFinished!\n")
  sim.p <- do.call(c, sim.p)

  # compute new level
  new_alpha <- as.vector(
    quantile(sim.p, desired_alpha)
  )
```

```
  names(new_alpha) <- "reject the null if p <= to:"

  # return
  return(new_alpha)
}

# a test run
g1 <- c(10, 20, 30, 10)
g2 <- c(20, 20, 50, 60)
set.seed(123)

 # via monte carlo
chi_alpha <- find_new_alpha(0.05, g1, g2)
chi_alpha # chi-squared

llr_alpha <- find_new_alpha(0.05, g1, g2, statName = "LLR")
llr_alpha # likelihood ratio

prb_alpha <- find_new_alpha(0.05, g1, g2, statName = "Prob")
prb_alpha # multinomial probability

 # asymptotic
asy_alpha <- find_new_alpha(0.05, g1, g2, simulate = F)
asy_alpha # asymptotic chis-quared
```