Analysis Plan

Project Name: Evaluating the Effects of the Wisconsin Home Energy Assistance Program (WHEAP) Summer Fill Program

Project Code: 2407

Date Finalized: 11/4/2024

Date Revised: 2/25/2025

Note: This is an updated version of a prior Analysis Plan which was posted on 11/4/24. The original Analysis Plan is available upon request by emailing oes@gsa.gov.

Project description

This project is a quasi-experimental study to examine the effects of the Wisconsin Home Energy Assistance Program (WHEAP) Summer Fill Program on priority outcomes for Low-Income Home Energy Assistance Program (LIHEAP) households. Specifically, we will use a difference-in-differences (DiD) design with a matching component, comparing over-time change between observationally similar households that are and are not participating in the Summer Fill Program.

Since 2015, the Wisconsin Department of Administration's (DOA) Division of Energy, Housing, and Community Resources (DEHCR), who implement the state-wide WHEAP, has implemented a Summer Fill Program. Through this program, households who have been approved for WHEAP, and whose primary heating source is propane or fuel oil, receive credits on their vendor account to refill their tanks prior to the winter season.[1] The Wisconsin DEHCR was able to serve a larger number of WHEAP households because of funding increases in program years 2021-2023.[2] Prior to this, Summer Fill was rolled out intermittently since 2015, if and when funding was available.

This evaluation is a collaboration between OES, the Wisconsin DOA DEHCR, and the HHS ACF to retrospectively examine the effects of the Summer Fill Program on WHEAP households, focusing on the priority outcome of whether the program reduces the need for households to apply for "crisis assistance." Crisis assistance allows local WHEAP agencies to respond to emergency situations or act proactively to avoid serious problems. Crisis assistance is a discretionary benefit to households; even if a household is eligible, they may not receive assistance.[3] The Summer Fill Program could have an effect on households' future need for energy assistance when, for example,

[1] Summer Fill benefit payment amounts are set each year by WI DEHCR; for example, in 2022, Summer Fill benefit amounts were $400 for eligible households with propane as their primary heating fuel type, and $800 for those with fuel oil as their primary heating fuel type.

[2] $8 million was invested in the Summer Fill Program by Wisconsin DEHCR in 2023. [Source]

[3] [Source]

their tanks are filled before the winter period, thus potentially reducing the likelihood that households will be in a situation in the future where crisis assistance is needed. Avoiding the need for crisis assistance — when households have critically low supplies of heating fuel, particularly during winter months, and are unable to pay to refill their tanks — is important for protecting household well-being.

More broadly, the aim is to build evidence that will be useful to DEHCR as well as the many other grant recipients across the country to inform their programmatic decision-making.

Hypotheses

The overarching research question is: does the Summer Fill Program reduce the need for households to apply for crisis assistance in the future?

We pose two primary hypotheses:

- H1: Propane or fuel oil households who participate in WHEAP and receive Summer Fill are less likely to apply for crisis assistance during the following program year.
- H2: Propane or fuel oil households who participate in WHEAP and receive Summer Fill will request lower dollar amounts of crisis assistance during the following program year.

It is also possible that Summer Fill reduces a household's energy burden enough that it even reduces the future need for participating in WHEAP. If this is the case, we may not observe whether a household applies for crisis assistance in the following program year (making it difficult to conclude whether Summer Fill has an effect on crisis assistance requests).

It is relevant for decision makers to know whether these households are less likely to participate in WHEAP in the future. Thus, a secondary hypothesis we pose is:

- H3: Propane or fuel oil households who receive Summer Fill are less likely to receive WHEAP benefits in the following program year.[4]

Data and data structure

Data source(s):

The primary data source to be analyzed is household data sent by WI DEHCR to OES, from their Home Energy Plus System which documents all households who apply for WHEAP benefits. The specific dataset OES will be analyzing is a list of all households who were approved for any form of WHEAP heat benefit during the program years 2015 - 2023, comprised of approximately 1,783,002 household-level benefit records (with 2,345,174 million rows of data). Data elements are included in Appendix A.

[4] As we only have access to data on households approved for WHEAP benefits, we are unable to determine whether a household applied for, but was not approved for WHEAP benefits in the following program year.

During the design stage of this project, data from WI DEHCR was sent to OES in order to allow the OES team to determine the feasibility of conducting a quasi-experimental study. Data access was restricted by an OES member outside the project team in such a way that the outcome and treatment data was de-linked from each other, before the research team analyzed the data. Two 'blinded' datasets were created from the data sent by WI DEHCR to OES: (1) a 'no treatment' dataset, blind of any treatment variables, and (2) a 'no outcome' dataset, blind of any variables related to outcomes.

A secondary data source to be analyzed is the U.S. Department of Agriculture (USDA) '2010 Rural-Urban Commuting Area (RUCA) Codes' data, which provides information on the rural-urban classification for households (provided at the Census tract level).[5] To identify urban versus rural areas, we will use the Primary RUCA Code 2010 variable.[6,7] Data elements are included in Appendix B.

We also plan to use the following dataset:

- U.S. Census Bureau data on 'Physical Housing Characteristics for Occupied Housing Units' to provide information on the number of structures in a given area (i.e., zip code level) that have been built in a given date range (e.g., 2010-2019).[8] The data elements we intend to draw on are included in Appendix C.

Outcomes to be analyzed:

We will analyze the following outcomes for the confirmatory analysis:

Variable	Description
participate_WHEAP	Whether or not a household that received WHEAP benefits in year t receives WHEAP in year t+1.
crisis_request	Whether or not a household requested any crisis assistance during a given program year.
crisis_request_amount	The value of the crisis assistance benefit (in $) requested by a household (if applicable). The original variable name is 'crisisrequestedamount' (see Appendix A).

[5] See this resource for more on characterizing geographies as urban/rural.

[6] The 2020 updated RUCA data are not yet available. While there will be measurement error in using the 2010 data, we have no reason to believe this error would systematically bias our findings.

[7] RUCA code values of 1, 2, 4, and 5 will be coded as RURAL=0 and RUCA code values of 3, 6, 7, 8, 9, and 10 will be coded as RURAL=1.

[8] The U.S. Census Bureau provides publicly available data for age of home, collected via the American Community Survey [Source]. The data is accessible from the U.S. Census Bureau website.

Imported variables:

We will merge in variables from the RUCA dataset (see Appendix B), using the 'county' variable contained within both data sets. We also plan to merge in variables from the U.S. Census Bureau housing characteristics data (see Appendix C).[9]

Transformations of variables:

The treatment and outcome variables used for analysis are created as follows (recoding for all data elements is included in Appendix A):

Variable type	Variable name	Description	How coded
Treatment	received_SF	Whether or not a household received Summer Fill.	Binary variable, with value of 1 if the value of 'crisisservicetype' is 'Summer Fill Payment' or 'Summer Fill Match Payment', 0 if not
Outcome	participate_WHEAP	Whether or not a household that received WHEAP benefits in year t receives WHEAP in year t+1.	Binary variable, with value of 1 if household receives WHEAP in following program year, 0 if not. I.e., value will be 1 for all households that have a parent application in a given year and will be 0 in years where a parent application for that household doesn't appear.
	crisis_request[10]	Whether or not a household requested any crisis assistance during a given program year, excluding Summer Fill benefits.	Binary variable that is = 1 if (1) 'crisisbenefitstatus' value indicates that crisis assistance was approved or denied, AND (2) 'crisisservicetype' does not equal "Summer Fill Payment" or "Summer Fill Match Payment"; = 0 if value of 'crisisbenefitstatus' is "NA"
	crisis_request_amount	The value of the crisis assistance benefit (in $) requested by a household (if applicable)	Takes the value of 'crisisrequestedamount' if 'crisisservicetype' does not equal "Summer Fill Payment" or "Summer Fill Match Payment'; equals zero otherwise.

[9] This data is only available at the zip code level (e.g., we will know for a certain zip code the number of units/structures in that area that were built in a given year range). Therefore, we will not know at the individual level what year a given household's unit was built.

[10] The 'crisis_request' and 'crisis_request_amount' outcomes will be limited to crisis assistance paid for deliverable fuel vendors (i.e., propane and fuel oil) only, e.g., not including crisis assistance benefits paid for electricity.

Certain variables are necessary to create to take into account the transforming the "long" orientation structure of the data, which is further described below.

Raw data for variables used for matching treated and control units will be inspected as part of the matching procedure (discussed below) and transformed as needed.

Transformations of data structure:

The data is in a format such that there is at least one parent application for WHEAP benefits for each household in each program year that a household was approved for WHEAP benefits (a "household-year"), and where the household identifier is the person ID for the household head. If the household requested any crisis assistance or received Summer Fill in a particular application year, a separate observation – a "child" application – is created for that household-year such that each crisis payment is its own observation. This means that there are potentially multiple observations per household-application year: at least one parent application and zero to multiple "child" applications (the parent application is a prerequisite for the child applications).

We will transform the data in such a way that there is one observation per household-year prior to analysis; specifically, we will aggregate variables across the parent and child applications. We will code a household-year as participating in WHEAP if there is a parent application for that program year. We will code a household-year as receiving Summer Fill – the treatment indicator – in that year if they received Summer Fill at any point in that year (i.e., if the field 'crisisservicetype' takes the value "Summer Fill Payment" or "Summer Fill Match Payment" for any child application in a given year).

We will code a household-year as requesting crisis assistance if the household requested crisis assistance at any point in that year. That is, 'crisis_request' = 1 if the field 'crisisbenefitstatus' indicates that a household was approved or denied crisis assistance and the field 'crisisservicetype' takes a non-missing value other than "Summer Fill Payment or "Summer Fill Match Payment" for any child application in a given year,[11] and = 0 if there are no child applications in a given year where 'crisisservicetype' takes a non-missing value other than "Summer Fill Payment" or "Summer Fill Match Payment". We will similarly code receipt of crisis assistance, where 'crisis_pay-amount' = 1 when 'crisispaymentamount' is populated with a dollar value and the field 'crisisservicetype' takes a non-missing value other than "Summer Fill Payment" or "Summer Fill Match Payment" for any child application in a given year.

When transforming the data to a single observation for each household-year we will create a variable that is the sum of all crisis assistance requested amounts ('crisisrequestedamount') among child applications in a given year where 'crisisservicetype' takes a non-missing value other than "Summer Fill Payment" or "Summer Fill Match Payment". We will similarly create a variable for the sum of all crisis assistance received amounts among child applications in a given year, summing

[11] The field 'crisisservicetype' can take the following values: "Summer Fill Payment", "Summer Fill Match Payment", "CA Emergency Assistance Payment", "CA Prevention Assistance Payment", "Emergency Assistance Payment", "PB Emergency Electrical Assistance Payment", "PB Prevention Electric Payment", "Prevention Arrearage Assistance", "Prevention Assistance Payment", or "Spring Prevention Assistance Payment".

values of 'crisispaymentamount'. In the event that the values of 'crisisbenefitstatus', 'crisisrequestedamount', and 'crisispaymentamount' do not align with each other, we will clarify with our partners as to which variable we should use to create our outcome variables or choose the variable with the least missingness if this cannot be clarified with confidence.

Applicants have an opportunity to change information in their parent application when applying for crisis assistance later in the year. This implies that covariates can change values throughout the application year in any of the child applications; in cases where covariates might change within a household-year, we will use the most recent covariates *before the first Summer Fill payment was administered for that household-year*.

Data exclusion:

We will include all households that received the Summer Fill Program, including in program years where WI DEHCR did not officially implement the Summer Fill program (2017, 2019, and 2020).[12] We will also run a robustness check where we analyze our outcomes without observations associated with households that received Summer Fill benefits in years where WI DEHCR did not implement the program officially, to ensure that the effects are not substantively different under both sets of analyses.

Treatment of missing data:

We will not interpolate or impute missing values for our outcome. The package we propose to use for our analysis (PanelMatch package, function matched_set, value match.missing) provides us the option to match on patterns of missingness in treatment histories (see Empirical Methodology & Hypothesis Tests). If there is missingness in treatment histories, we will assess the robustness of results to both options (using this option or not) and justify one option for our primary analysis in the Technical Appendix.

If missingness is low for other covariates we will include in our estimation procedure (10% or less of the covariate's values are missing), we will drop those observations. If missingness is high for those covariates and there are some non-missing values for a given covariate across years for a household, we will use interpolation between years for that covariate for that year for that household. If missingness is high and there are missing values for all years for a given covariate for a given household, then we will include a dummy for whether a given covariate is missing ("missingness dummy") as an additional covariate and use mean-imputation for the missing values for each year (a modified version of "dummy variable adjustment") or consider dropping that covariate from our estimation strategy, justifying our decision in a Technical Appendix. One important variable with high levels of missingness that we will use for our matching procedure described below is annual heat fuel cost. WI DEHCR imputed annual heat fuel cost with proxies when it was not available. We will use the raw imputed value (that ignores any discounting they

[12] Approx. 253 households received Summer Fill in 2017, 342 in 2019, and 354 in 2020. WI DEHCR have indicated that this may be because while the Summer Fill Program was not formally offered during those years, local agencies independently administered Summer Fill benefits to households.

made to penalize applicants from not providing it), and then match on a missingness dummy for whether annual heat cost is missing.

Descriptive statistics, tables, and graphs

We will include the following descriptive statistics in the OES abstract:

1. The percentages of WHEAP households that received vs. did not receive Summer Fill, out of those eligible[13]

2. The percentage of WHEAP households that received Summer Fill who requested crisis assistance, broken down by emergency and prevention type crisis payments

3. Average heat benefit amounts per year for WHEAP households, including those who received vs. did not receive Summer Fill

4. The percentages of WHEAP households that received vs. did not receive Summer Fill, out of those eligible, broken down by program year

We may also report the following descriptive statistics to the partner[14]:

5. Median or average amounts of crisis assistance requested and received by WHEAP households that received Summer Fill, before and after receiving Summer Fill

6. Descriptive statistics related to all WHEAP households' requests for and receipt of crisis assistance, broken down by program year. Specifically:

    a. The percentage of WHEAP households that requested and received crisis assistance.

    b. The median or average amounts of crisis assistance requested and provided to each household per household-year that receives crisis assistance.

7. Table of the Summer Fill benefit amounts provided to households each program year, broken down by fuel type (i.e., propane vs. fuel oil)

8. Table summarizing the number of Summer Fill households who receive past due or disconnection notices for utilities, or are indicated as 'less fuel' (less than 20% fuel remaining) or 'out of fuel' (no fuel remaining).

[13] WHEAP households are eligible for the Summer Fill Program if they were approved for WHEAP benefits in the prior program year, their primary heating source is propane or fuel oil, and they are a homeowner.

[14] For descriptive statistics related to crisis assistance, we will examine both crisis assistance payments paid for deliverable fuel vendors (i.e., propane and fuel oil), as well as crisis assistance payments paid for all vendor types (including both deliverable and non-deliverable fuel types).

9. Table (or graph) summarizing when and how much of the Summer Fill credits are used by households (after initially being credited to their account).[15]

10. Price trend graph showing price of delivered fuel by date of delivery.

Empirical methodology and hypothesis tests

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

Empirical methodology:

Overview of methodology

Summer Fill was implemented by WI DEHCR intermittently. WI DECHR only officially implemented the Summer Fill program for three years before 2021 (2015, 2016, and 2018), and no more than two years of implementation were consecutive in that timeframe. We will investigate the viability of the difference-in-differences (DiD) estimator described in Imai, Kim, and Wang (2023) ("IKW") as our analysis strategy. IKW's methodology is suitable for our context because it allows units to be treated multiple times, for units to switch their treatment status multiple times over time, and for previous treatment status to affect current treatment status or outcomes ("carryover effects"). We plan to use IKW's R package PanelMatch to implement their estimator if we pass our diagnostic checks. We will explore other potential analysis strategies if their estimator is not viable .

The proposed approach involves applying a DiD estimator to adjust for a possible time trend after using matching as a nonparametric preprocessing step to ensure that control and treated observations in the sample are similar in terms of treatment and covariate histories. Specifically, for each treated observation, a set of control observations in the same time period is selected that has an identical *treatment* history for a prespecified number of periods (i.e. the number of lags, described further below). This treatment-history-matched set is further refined by using standard matching or weighting methods to adjust for observed confounding so that matched control observations become similar to the treated observation in terms of *covariate* histories. IKW describe their estimator as a nonparametric generalization of the difference-in-differences estimator that does not rely on the assumption of linearity; it is algebraically equivalent to a weighted two-way fixed effects estimator, where the weights represent the number of times an observation is used for matching. An advantage of the proposed matching estimator is that it is more robust to model misspecification than the standard two-way fixed effects regression estimator.

[15] Community engagement participants indicated that a key metric used by local agencies and energy vendors to measure the success and impact of the Summer Fill Program is the assessment of clients' balances and credit with their energy vendors. Participants noted that clients often have larger credits and fewer instances of running out of fuel, which demonstrates the program's positive impact on their financial stability and energy management.

Estimation

Estimand:

Our estimand is the average treatment effect of the Summer Fill program in the next application year (at time $t + 1$) among treated households ("ATT") that switch into treatment status during the current time period (time t) from control status in the previous time period (time $t - 1$); they must also have the same treatment history as control households for a prespecified number of periods. We will follow the IKW approach by creating the counterfactual outcome for each treated observation using the weighted average of the control units in the refined matched set. Then we will compute the DiD estimate of the ATT for each treated observation and average it across all treated observations. The DiD estimator measures the difference in year-over-year changes in outcomes between households that received treatment in the current year (the treated group) and those that did not receive treatment in the current year (the control group).

Estimator:

The IKW estimator involves choosing a number of leads representing F time periods after the administration of treatment, and a number of lags representing L time periods before the administration of the treatment. The number of lags is part of the identification assumption – matching treatment history is one dimension along which matching will be executed. Adding lag periods (requiring observations' treatment histories to be the same going back further in time) reduces confoundedness by removing one source of bias, downstream treatment effects. However, this potential reduction in bias comes at cost: requiring observations' treatment histories to be identical over longer time frames reduces potential matches and therefore sample size, increasing estimate variance and decreasing estimate precision.

To lay this strategy out more formally and fix terms, given a chosen L and F, we want to estimate the average causal effect of policy change on those whose policy changed (IKW eq 8):

$$\delta(F, L) = \mathbb{E}[PO_{x=1} - PO_{x=0} \mid X_{it} = 1,\ X_{i,t-1} = 0]$$

where

$$PO_{x=1} = Y_{i,t+F}(X_{it} = 1, X_{i,t-1} = 0, \{X_{i,t-l}\}_{l=1}^{L})$$

and

$$PO_{x=0} = Y_{i,t+F}(X_{it} = 0, X_{i,t-1} = 0, \{X_{i,t-l}\}_{l=1}^{L})$$

In these expressions, i indexes households, t indexes years, and X_{it} refers to a household's Summer Fill receipt in a given year. PO_{x} refers to a household's potential outcomes at time $t + F$ if

receiving Summer Fill ($x = 1$) or not ($x = 0$) at time t , assuming the household was not treated in the prior period ($t - 1$), and given the household's particular treatment history in the periods from $t - 1$ until $t - L$. As mentioned above, in our setting this corresponds to the average treatment effect on the treated, the ATT.

We cannot observe that estimand directly, since we only observe one potential outcome for each household, so we estimate it using the following expression instead (IKW eq 18):

$$\widehat{\delta}(F, L) = \frac{1}{\sum_{i=1}^{N} \sum_{t=L+1}^{T-F} D_{it}} \sum_{i=1}^{N} \sum_{t=L+1}^{T-F} D_{it} Diff_{it}$$

where

$$Diff_{it} = (Y_{i,t+F} - Y_{i,t-1}) - \sum_{i' \in M_{it}} w_{it}^{i'} (Y_{i',t+F} - Y_{i',t-1})$$

Here, i and t are used as above, Y_{it} is the observed outcome for a given household in a given year. For "treated" households receiving Summer Full in year t but not year $t - 1$, $Diff_{it}$ is the difference between (1) the change in that household's observed outcomes from $t - 1$ to $t + F$ and (2) a weighted sum of that same quantity for its matched control households, where the weights, $w_{it}^{i'}$, are generated by the chosen covariate matching procedure (see below). I.e., this is the difference between each treated household's change over time and the estimated change over time for its "synthetic" or "reconstructed" counterfactual. D_{it} is essentially an indicator for a treated observations included in estimation (treated in t, but not $t - 1$, and has matched controls). The rest of the expression for $\widehat{\delta}(F, L)$ represents averaging $D_{it} Diff_{it}$ across treated observations.

In practice, applying this estimation strategy proceeds roughly as follows:

- Step 1: Match treated households (i.e., those that received Summer Fill) in each program year with households that were untreated in that same program year but have the same treatment history in prior years.
- Step 2: Refine the set of matched control observations by applying a matching method based on observable household characteristics.
- Step 3: Estimate the effect of Summer Fill on the primary outcomes by taking the difference in differences: first, take the difference between treated (control) units in treatment years ($t + F$) and control years ($t - 1$); second, take the difference between that quantity in a treated observation and its weighted set of control observations; and third, average this across treated observations.

Identification assumption

The identification assumption behind this estimator is "parallel trends" conditional on treatment, outcome (lagged outcomes), and covariate histories. This assumption would be violated if there are unobserved confounders that vary across units and time. As described further below, we can examine whether the outcome time trends are parallel on average between the treated and matched control units for at least a subgroup for households, and whether differences in the characteristics of the treated vs control groups are constant in pretreatment (i.e., no treatment) periods.

Methodological decisions

The number of leads and lags

We will set the number of leads, F, equal to 1 because we are interested in the short-run average treatment effects of the Summer Fill program in the following application year.

We will select the number of lags that we believe optimizes the bias-variance tradeoff using the procedure described in this paragraph. The number of lags defines the number of time periods in the past where treatment history will be identical for a treated observation and the control observations in a matched set, i.e. the number of periods for which carryover effects are allowed and not confounding. A higher number of lags makes the unconfoundedness assumption more credible, but also reduces the number of potential matches. A small number of control matches (e.g. fewer than five) or unmatched treated observations could threaten the validity of estimates generated by the IKW methodology. We will first start with the highest possible number of lags and determine how many matched control observations there are per treated unit and whether there are treated observations with no matches. We will decrease the number of lags one by one such that the highest number of lags is chosen for which there is a sufficient number of matches (e.g. 30) per treated observation.

Treatment history matching

This estimation strategy involves matching treated households in each program year with households that are untreated in that program year but share the same treatment history in prior program years (up to our chosen number of lags). Exact matching on the treatment history allows us to partially control for confounding effects of past treatment on current treatment status and outcomes ("carryover effects"). This strategy also involves exact matching on time period where the matched sets only include observations from the same time period in order to adjust for time-specific unobserved confounders. If the matched sets become too small, we may partially relax these matching restrictions so that each treated observation is matched with control observations that have a similar (but not exactly the same) treatment history.

To be included in the matched set as a treated observation, the household must have switched from control (did not receive Summer Fill) at time *t-1* to treatment status (received Summer Fill) at time *t* and must have at least one matched control unit. To be included in the matched set as a control observation, the household must have remained in control status at both time *t-1* and *t* (i.e., did not receive Summer Fill in either a given program year or the year prior). We are

comparing units that switch into treatment at time *t* from *t-1* with those that stay in control; identification relies on parallel trends between these groups.

Covariate history matching

We will refine the matched sets of control observations for each treated household by matching on observable covariates. We will compare the performance of the three matching refinement methods described by IKW – Mahalanobis distance matching, propensity score matching and propensity score weighting — in terms of how much each method removes imbalance in observable confounders, the number of matches (size of matched set), and the sample size it produces. We will choose one to use for our main analysis if the imbalance, number of matches, and sample size is sufficient to produce reliable estimates.

We will explain our choice of primary refinement method in a Technical Appendix, providing information on the covariate imbalance, number of matches, and sample size associated with each matching refinement method and how we decided what was deemed acceptable. It is possible that all methods lead to concerning sample sizes, in which case we plan to evaluate the feasibility of a regression version of the IKW estimator that estimates the ATT parametrically (e.g. that adjusts for confounders via regression instead of nonparametrically via matching); if such a regression is feasible, we will similarly justify the choice of this method in our Appendix.

Below we have listed the variables we would consider using for matching treated to control units by prioritization level; we have noted which variables we propose to conduct exact matches on:

<u>Priority 1 (higher):</u>

- ***Summer Fill eligibility.*** We initially propose conducting exact matching on eligibility unless there are power concerns resulting from the sample size reduction caused by exact matching. By matching exactly on eligibility, we would be matching WHEAP-approved households who were *(in)eligible* based on their homeownership status and fuel type (WHEAP-approved homeowners with delivered fuel) but *did not participate in Summer Fill* to WHEAP-approved households who were *(in)eligible and participated*.[16]

  While we do not have full information on all the reasons eligible households might not participate, conversations with our agency partners suggest that these are households that were unable to be reached or were never contacted in the first place. Households might not be able to be reached because they changed phone numbers, moved, or were too busy. We are able to identify moves with changes in address. We do not have data to identify changes in a household's phone number but we do have data on their preferred

[16] Exact matching would imply that we are also matching Summer Fill participants who are renters or not delivered fuel customers to renters or not delivered fuel customers who did not participate. Some other reasons to consider exact matching on eligibility, fuel type (below), and homeownership (below) is that delivered fuel households might be more likely to request crisis assistance, and this likelihood of requesting might change over time in a way that is not parallel to the likelihood of requesting assistance for the control group. Furthermore, households might apply for crisis assistance for different reasons. Deliverable fuel HHs would apply due to "less fuel remaining" or "out of fuel"; electricity/natural gas HHs would apply due to "past due notice" or "disconnection notice.

communication method. How busy households are is captured by household size, student status, children in household, and having young children (aged less than 6). Homeless status, preferred communication method, preferred language, application method, and disabled household count can account for other reasons households cannot be reached. After matching on all the variables specified in this list of covariates, and potentially applying LASSO regression to help identify better predictors of Summer Fill participation (see below), idiosyncratic factors are the most likely remaining explanation why people who were *already approved for LIHEAP that year* didn't accept Summer Fill benefits for that year.

- ***Homeownership status.*** We propose exact matching on this.
- ***Fuel type.*** We propose exact matching on this.
- ***Annual heat fuel cost.***
- ***Whether annual heat fuel cost was imputed by WI DEHCR with a proxy calculated from dwelling type, size, and heating fuel in cases when it was not provided in the application.***
- ***Non-heat fuel annual cost.***
- ***Whether annual non-heating fuel cost was imputed by WI DEHCR in cases when it was not provided in the application.***
- ***The summed value of other WHEAP benefits received (heat benefit paid and public benefit paid for parent application).***[17]
- ***Income.***[18]
- ***Rural-urban classification of household's county/city.***[19]

[17] Benefit levels are primarily based on the relationship of household income to the state median income (SMI), number of rooms, dwelling type, and home energy costs. One reason cited in the information transmittals provided by WI DEHCR that Summer Fill has generally been restricted to homeowners is that homeowners are not eligible for Wisconsin Emergency Rental Assistance Program (WERA). As a result, they recommend that renters who have propane or fuel oil as their main energy source be referred to WERA. We do not observe WERA benefit amounts in our data. Nor do we have data benefit amounts from other WHEAP programs such as HE+ Program Services and Weatherization services, or non-WHEAP benefits such as payments to an energy provider using Keep Wisconsin Warm/Cool funds. Therefore, this measure of other relevant benefits received is not comprehensive.

[18] Calculated as poverty level percent, so that it is normalized to the annual poverty level threshold. The LIHEAP Report to Congress for Fiscal Year 2020 used the following income categories: (1) lowest income ($20,000 or less), (2) moderately low income (between $20,001 and $39,999), (3) higher-income households (between $40,000 and $59,000), and (4) highest-income households ($59,001 or more). As cited in the "LIHEAP Report to Congress for Fiscal Year 2020: Appendix A," according to the 2015 RECS, the mean residential energy burden for all LIHEAP income eligible households in 2015 was 11.6 percent, and the median is 9.7 percent." 2015 RECS households with incomes of $20,000 or less had average residential energy expenditures of $1,423, while those with incomes between $40,000 and $59,999 had average residential energy expenditures of $1,781. Thus, households which had more than twice as much income spent only 25 percent more on energy." This indicates a nonlinear (concave) relationship between income and energy expenditure, and thus potentially between Summer Fill participation and the likelihood of applying for crisis assistance because of cost reasons.

[19] Rural areas are less cost-effective for investing in infrastructure for natural gas and electricity, and therefore more suitable for constructing houses that use propane or fuel oil as their heating source. This map suggests that propane and natural gas are the most common fuel sources, and that propane use is concentrated in the southeastern region of Wisconsin.

- ***Heat fuel vendor.***[20] We propose exact matching on this.
- ***Local agency.***[21]

Priority 2 (lower):

- ***Heat vendor payment method.***
- ***Elderly household count.***
- ***Reason for not having annual cost for primary fuel in application (if consistently recorded).***
- ***Reason for not having annual cost for non-heating fuel in application (if consistently recorded).***
- ***Age of home, proxied by median age range of homes in household's zipcode.***
- ***Changes in residential address.***[22]
- ***Household size.***
- ***Homeless status.***
- ***Student status.***
- ***Preferred communication method.***
- ***Preferred language.***
- ***Application method.***
- ***Number of household members with a disability.***
- ***Children in household.***
- ***Number of children less than 6 years old in the household.***

All time-varying covariates listed will be included in the Mahalanobis distance and the propensity score model *unless* the inclusion of all covariates leads to unacceptable loss of sample size or matches. In that case, we will apply a combination of theory and data-driven methods (e.g. LASSO) to a subsample of our data (e.g. 2015, the first treated year in our sample) to select covariates for inclusion, prioritizing the Priority 1 level variables listed above (and choosing theoretically important variables when they are collinear with less theoretically important variables). We will use this more minimal selection in the Mahalanobis distance and the propensity score model specified with PanelMatch, reassess the performance of the matching refinement methods with the smaller set of matching covariates, and publish this assessment in our Technical Appendix.

As recommended by Chiu et al. (2024), we will also plot the raw data for the variables we plan to use for the matching step of this methodology to determine whether some variables need to be transformed first due to outliers; if a regression approach is taken, we will also determine if these variables need to be transformed due to nonstationarity or high skew in their distributions.

Diagnostic checks

[20] According to this website, propane prices in Wisconsin can vary considerably by supplier, suggesting that matching on the energy vendor would be appropriate.

[21] The local agency that processed the household's WHEAP application.

[22] We can observe whether a household head moved by identifying changes in the addressid (de-identified address variable).

We will first construct the "treatment variation plot" discussed in IKW using their R package to assess variation in the treatment and whether it is sufficient to produce unbiased estimates.[23]

Examining the parallel trends assumption for potential violations

The parallel trends assumption requires a reasonable level of confidence that, absent Summer Fill, the treatment group and matched control group would have experienced similar year-to-year trends in the need for crisis assistance, and potentially underlying financial stability. While we cannot test this assumption directly, we will examine the plausibility of this assumption and consequently our design and the feasibility of an impact evaluation by conducting covariance balance checks proposed by IKW, a placebo test, and diagnostic checks recommended by Chiu et al. If these checks fail, we may consider methodological changes to our design that would increase confidence in our design's validity (e.g. further sample restrictions).

Following IKW, we will examine the standardized mean difference of each covariate between a treated observation and its matched control observations at each pretreatment time period during the study period, including program years 2015-2023 ("covariance balance checks"). This mean difference is standardized at each pretreatment time period by the standard deviation of each covariate across all treated observations in the data so that the mean difference is measured in standard deviation units. The covariates that must be included in these balance checks include the lagged outcome variables and time-varying covariates. Standardized mean differences for the lagged outcome that remain relatively constant over the entire pretreatment period across refinement methods supports the parallel trends assumption.

We also plan to use PanelMatch to implement a placebo test that looks at the change in outcome at t-1, compared to other pre-treatment periods in the lag window. We will lastly conduct select diagnostic checks recommended by Chiu et al. for assessing the plausibility of the parallel trends assumption when there are more than three pretreatment periods (see A.4.4 in Chiu et al. for more details on these checks). To ensure we have three consecutive years of no treatment and that the results of this checks are consequently meaningful, our first check will involve constructing an event-study plot for a subgroup of households that were in our sample 2018 - 2020 but not treated prior to 2021; this check allows us to examine whether pretreatment trends are parallel visually.[24] The second check recommended by Chiu et al. that we will implement involves descriptively exploring sources of heterogeneity that might threaten the validity of parallel trends by summarizing treatment effects across time, unit (or cohort), or pretreatment covariates.

It is also important to point out that the credibility of IKW's covariate balance checks as checks for parallel trends relies on the assumption that we have data on important confounders, i.e. we have good measures of characteristics that affect the outcome and differ between treatment and control group. Indeed, we believe matching on the covariates specified below sufficiently adjusts

[23] The treatment plot will allow us to determine whether treatment varies sufficiently over time and across units to avoid impermissible comparisons and biased estimates (Chiu et al., 2024; Imai, Kim and Wang, 2023).

[24] These diagnostic tests are based on the FEct package and are also available in the *paneltools* package (detailed tutorial here).

for important sources of confounding. Important sources of confounding include group differences in:

- energy burden, which is captured by annual heat fuel costs,[25] annual non-heat fuel costs, building age, income, the dollar amount of all other WHEAP benefits (because these decrease energy burden), and fuel vendor (because fuel price affects fuel costs and fuel price is likely vendor-specific);
- fuel delivery schedule of the energy-vendors, which is captured by the rural–urban classification of the household's city (proxy for population density) and heat vendor payment method;
- eligibility for Summer Fill, which is captured by the eligibility variable constructed from homeownership status and fuel type;
- monetary and non-monetary (labor and hassle) costs associated with different types of fuel, which is captured by fuel type;
- landlords and tenants, including differential awareness about energy costs (see Myers 2018) and differential incentives between landlords and tenants to conserve energy and take up energy efficiency upgrades, all of which influence energy expenditures; this is captured by homeownership status;
- outreach about and administration of WHEAP benefits, including discretion in the amount of crisis assistance provided, which could affect benefit takeup; this is captured by local agency;
- weather and climate, which may also be captured by local agency, to the extent that a local agency's jurisdiction overlaps with geographically defined areas; and
- likelihood of not being able to be reached successfully about Summer Fill eligibility, which is captured by changes in residential address, household size, homeless status, student status, preferred communication method, preferred language, application method, disabled household count, children in household, and having young children (ages 2, or between 3 and 5 years old).

Therefore we believe that IKW's diagnostic check provides useful evidence on the plausibility of the parallel trends assumption.

Sensitivity analyses

Following IKW, we plan to examine the sensitivity of our findings to the maximum number of matches and the number of lags. In addition, we plan to conduct a sensitivity analysis of the

[25] The proxy for annual heat fuel cost takes into account dwelling type, number of rooms, and fuel type; we plan not to include dwelling type and number of rooms as additional matching variables because we would only include them because they influence fuel costs.

matching procedure to unobserved confounders that asks “How much would our results change if the unobserved sample characteristics differed a lot between those who are/are not receiving Summer Fill and also predicted requests for WHEAP crisis benefits?”.

Following Chiu et al., we also plan to conduct a sensitivity analysis to understand how robust our results are to mild parallel trend assumption violations (see Rambachan and Roth, 2023), assessed in conjunction with our sensitivity analysis.

Timeframe selection

Preliminary analysis suggests that, even without matching or refinement, households that received Summer Fill are relatively comparable in terms of observables to households that did not receive Summer Fill, and the characteristics of the WHEAP approved households in the sample are relatively constant throughout application years.

Nonetheless, the onset of the Covid-19 pandemic may provide basis for concern that trends in the treatment and control groups are not parallel in unobservables 2021 onwards; funding for both WHEAP and Summer Fill increased beginning in 2021 through 2023, which expanded access to both of these programs. In other words, it is plausible that treatment and control groups were differentially affected by the pandemic. Under the assumption that we observe all the important confounders, the balance check in observables across time IKW proposed should justify the validity of the parallel trends assumption during the pandemic just as it would for non-pandemic years. If some of these years fail the balance check, we can exclude all the years after the first year that fails our check from our analysis.

Selection bias

There is a possibility that treatment might affect exit from the sample of households we have data on (our secondary hypothesis). For example, WHEAP households that have unused credits on their energy vendor account might not apply for WHEAP the following year. To account for the selection bias caused by treatment-induced exits from our sample, we will first run our DID estimator to estimate the effect of receiving Summer Fill on being approved by WHEAP (which is equivalent to being in our sample) the following year (“WHEAP participation regression”). If a “two one-sided equivalence test” (TOST) does not let us conclude that there is no effect on WHEAP participation with 95% confidence, we will then assess whether it is still meaningful to run our DID estimation strategy for our two primary hypotheses and describe this assessment in our Technical Appendix.

Confirmatory analyses:

We will first test whether receipt of summer fill has an effect on participation in WHEAP in the subsequent program year (the secondary hypothesis). The null hypothesis is that matched treated and untreated units are no different in their likelihood of participating in WHEAP in the following program year.

If we do not reject the null hypothesis for the first confirmatory analysis (i.e, Summer Fill doesn't appear to affect WHEAP participation; see the "Selection bias" section above) we will conduct further confirmatory analyses of the two primary hypotheses:

1. Effect of receiving Summer Fill on the likelihood of requesting crisis assistance in the following year. We will test the null hypothesis that matched treated and untreated units are no different in their likelihood of requesting crisis assistance in the program year following receipt of Summer Fill.

2. Effect of receiving Summer Fill on the amount (in dollars) of crisis assistance requested in the following year: We will test the null hypothesis that there is no difference in the average amount of crisis assistance requested between matched treated and untreated units in the year following receipt of Summer Fill.

Exploratory analysis:

We will conduct exploratory analyses to understand how the Summer Fill program might affect outcomes for subgroups of households based on the following variables:

- Income[26]
- Households with members who are elderly (aged 60 or over), have a disability, and/or are a young child (under the age of 6).
- Age

Inference criteria, including any adjustments for multiple comparisons:

We will make inferences using a two-tailed t-test of the estimated coefficient of interest with an alpha rejection region of 0.05.

We will not make an adjustment for multiple comparisons when testing the hypothesis that Summer Fill has an effect on participation in WHEAP (the secondary hypothesis). We plan to control for the family-wise error rate (FWER) when testing the primary hypotheses that Summer Fill has an effect on requests for crisis assistance in the following year and that Summer has an effect on the amount of crisis assistance received in the following year. We will use a Holm-Bonferroni correction to control the FWER for these two hypotheses. Because we have two hypotheses this means we will make inferences by comparing the smaller of the two p-values against a rejection region of 0.025 (i.e., 0.05/2) and the larger of the two p-values against 0.05 (if the first p-value is statistically significant).

Limitations:

Anticipated limitations are described below:

[26] See footnote for income in C*ovariate history matching* section.

1. **Inability to link households across program years when 'head applicant' has changed:** We have a person identifier instead of a household identifier – therefore, we will be unable to identify that a household is the same when the 'head applicant' has changed throughout 2015-2023 will not be able to identified as the same household.
2. **Lack of representativeness due to data availability:** We only have data for households who were approved for some form of WHEAP heat benefit between 2015 - 2023. Therefore, we do not have data on households who (1) applied for WHEAP benefits but were denied, (2) applied and received a different WHEAP benefit (e.g., weatherization assistance, HE+ Program Services),[27] or (3) were eligible for WHEAP but did not apply.[28] This may limit the external validity of our results.
3. **Lack of data on variation of fuel prices:** Fuel prices vary within and across program years, affecting the need to apply for crisis assistance (and the value of crisis assistance requested/received). However, we do not have access to individual-level data on fuel prices for our whole sample to understand how changes in fuel prices might impact household need.
4. **Power concerns:** Preliminary data analysis suggests that requests and receipts of crisis assistance are rare outcomes; this may imply that the effect size associated with Summer Fill is small. While we have a large sample, matching may significantly reduce our sample size, limiting our ability to detect relatively small effect sizes.
5. **Utilization of results:** We could observe a null effect of the Summer Fill Program on outcomes, despite there being other benefits of the program that we are unable to observe given available data. While we are focusing specifically on outcomes related to the future need for energy assistance for this project, this is just one of several potential impacts or downstream benefits that the Summer Fill Program could have, for example:
    a. Energy vendors can serve households as part of their regular schedule of deliveries when Summer Fill credits are already applied to energy vendor accounts, which may reduce service disruptions and support efficient delivery of fuels to WHEAP households.
    b. Decreased financial stress and increasing household well-being as a result of being prepared with a full fuel tank going into the winter and potentially higher indoor temperatures
    c. Increased value of WHEAP funding and reduced administrative burden for local agency staff (e.g., reduced phone calls from households regarding crisis assistance) due to propane or fuel oil being purchased during summer months when fuel is typically cheaper,

[27] The majority of WHEAP benefits are allocated to heating assistance; for example, approx. 98% of WHEAP households were served by heat benefits in 2023. Other benefits/services include: weatherization, equipment repair/replacement, winter or year round crisis, or other crisis (WHEAP does not provide cooling assistance). WHEAP households served by heat benefits may also be served by other WHEAP benefits/services.

[28] WHEAP served between 28.47% - 31.87% of all income-eligible households across the state between 2015 - 2023. [Source].

d. Increase in real household income as more is available for other household bills, e.g., electricity bills.

It is important to note that we cannot observe these other potential benefits as part of this project; a null result of the effects of the Summer Fill Program on households' future need for energy assistance may imply that our results might be inconclusive and does not mean that the Summer Fill Program is not beneficial in other ways.

Appendices

Appendix A: List of WHEAP household data elements

Variable name	New variable name	Description	Format	Example values	Transformations
parentapplicationyear	app_year	The year in which the application was submitted.	Numeric	2018	N/A
parentapplicationnumber	app_number	A numeric value that uniquely identifies an application.	Numeric	5520044	N/A
parentapplicationdate	app_date	The date when the application was submitted.	Date	12/04/2017	N/A
parentapplicationheatbenefitawar	heat_benefit_awarded	The amount of heat benefit awarded for the application.	Numeric (in $)	$550.00	N/A
parentapplicationheatbenefitpaid	heat_benefit_paid	The actual amount of heat benefit paid for the application, which may be less than the awarded amount.	Numeric (in $)	$550.00	N/A
parentapplicationpbbenefitawarde	pb_benefit_awarded	The amount of PB benefit awarded for the application.	Numeric (in $)	$286.00	N/A
parentapplicationpbbenefitpaid	pb_benefit_paid	The actual amount of public benefit (PB) paid for the application, which may be less than the awarded amount.	Numeric (in $)	$286.00	We will create a new variable called 'WHEAP_ben_total' that is the sum of 'pb_benefit_paid' and 'heat_benefit_paid'
benefitextractiondate	benefit_extract_date	The extraction date of the benefit for the parent application. If there are multiple transactions with different extraction dates, the earliest	Date	01/17/2018	N/A

		extraction date is returned.			
crisisnumber	N/A	A numeric value that uniquely identifies a crisis.	Numeric	NA	Transformed to one binary variable: • crisis_num_received: 1 if a crisis assistance number has been received by a household, 0 if not
crisisservicedate	crisis_service_date	The date when a payment was made for the crisis.	Date	NA	N/A
crisispaymentamount	crisis_pay_amount	The payment amount that was actually paid for the crisis.	Numeric (in $)	$286.00	N/A
crisisrequestedamount	crisis_request_amount	The payment amount requested by the client for the crisis.	Numeric (in $)	$286.00	N/A
crisisbenefitstatus	crisis_benefit_status	Status of crisis benefit approval (if applicable).	Text	• BENEFIT APPROVED BASED ON APPLICATION INCOME	Transformed into three binary variables indicating whether crisis benefits were approved, denied, or not applicable (no crisis benefit): • crisis_approved: 1 if crisis assistance was approved, 0 if not • crisis_denied: 1 if crisis assistance was denied, 0 if not • crisis_NA: 1 if not applicable (i.e., crisis assistance not applied for), 0 otherwise A fourth binary variable is created called 'crisis_request' indicating whether a crisis assistance application was made, with a value of 1 if 'crisis_approved' or 'crisis_denied' is 1, and 0 if 'crisis_NA' is 1.

addressid	N/A	A numeric value that uniquely identifies an address.	Numeric	69798	We will create a new binary variable called 'moved' indicating whether or not a household has moved, where it is = 1 if the 'addressid' variable has changed between the current and previous time period, and = 0 if not.
city	N/A	The city where the applicant resides.	Text	WISC RAPIDS	N/A
county	N/A	The county where the applicant resides.	Text	WOOD	N/A
zip	N/A	The ZIP code of the applicant's residence.	String	54494-2919	N/A
caseheadageatapplicationdate	head_age	The age of the case head at the application date.	Numeric	49	N/A
housingtype	housing_type	The type of housing of the applicant's residence.	Text	Single family house	N/A
ownershiptype	ownership_rent ownership_own	The ownership type of the applicant's residence.	Text	Rent	Transformed to two binary variables: • ownership_rent: 1 if client rents, 0 if not • ownership_own: 1 if client owns, 0 if not
totalnumberofhouseholdmembers	number_household_members	The total number of people in the household.	Numeric	1	N/A
roomcount	N/A	The number of rooms in the applicant's residence.	Numeric	6	N/A
livingsituation	living_situation	Describes the current living situation of the applicant.	Text	None of the above	Transformed to one binary variable: • living_situation_homeless: 1 if client is homeless, 0 if not

governmentassistedhousing	govt_assisted_housing	Indicates whether a applicant resides in government-assisted housing.	Text	No	Transformed to one binary variable: • is_govt_assisted_hous: 1 if government assisted housing, 0 if not
guardiantype	guardian_type	If an applicant has a guardian, describe the authority granted to the guardian.	Text	Not Applicable	Transformed to one binary variable: no_guardian: 1 if no guardian, 0 otherwise
studentsituation	student_situation	Describes various situations related to the student status of the applicant.	Text	N/A Case head is 25 years or older	Transformed to two binary variables: • not_student: 1 if not a student, 0 if student • is_student: 1 if student, 0 if not student
preferredcommunication	preferred_communication	Describes the preferred communication method with the applicant.	Text	Primary Phone	N/A
preferredlanguage	preferred_lang	The preferred language of the applicant.	Text	English	Transformed to one binary variable: • preferred_lang_engl: 1 if preferred language is English, 0 otherwise
outreachindicator	outreach	The communication method used by the applicant to complete the application.	Text	Phone	N/A
militarycountperhousehold	N/A	The number of military members in the household.	Numeric	0	N/A
disabledcountperhousehold	N/A	The number of disabled members in the household.	Numeric	1	N/A
childfamilyhousing	child_housing	Indicates whether there are children in the household.	Text	No	Transformed to two binary variables:

					• is_child_housing: 1 if client has child in household, 0 if not • not_child_housing: 1 if client does not have child in household, 0 otherwise
childrenunder2count	N/A	The number of children under 2 years old in the household.	Numeric	0	N/A
childrenbetween35count	N/A	The number of children between the ages of 3 and 5 in the household.	Numeric	0	We will create a new variable called 'overall_child_count' that is the sum of 'childrenunder2count' and 'childrenbetween35count'
elderlycountperhousehold	N/A	The number of elderly individuals in the household.	Numeric	0	N/A
categorical eligibility	categorical_eligibility	Indicates whether the household meets categorical eligibility criteria.	Text	Yes	Transformed to two binary variables: • is_cat_eligible: 1 if household is categorically eligible, 0 if not • not_cat_eligible: 1 if household is not categorically eligible, 0 if household is categorically eligible
smieligibility	smi_eligibility	Indicates whether the household meets SMI eligibility criteria.	Text	Yes	Transformed to two binary variables: • is_smi_eligible: 1 if household meets SMI eligibility criteria, 0 if not • not_smi_eligible: 1 if household does not meet SMI eligibility criteria, 0 if household does meet SMI eligibility criteria
zeroincome	zero_income	Indicates whether the household has zero income.	Text	No	Transformed to two binary variables:

					• is_zero_income: 1 if household is zero income, 0 if not • not_zero_income: 1 if household is not zero income, 0 if household is zero income
referred	N/A	Indicates whether the household was referred.	Text	Yes	Transformed to two binary variables: • is_referred: 1 if household was referred, 0 if not • not_referred: 1 if household was not referred, 0 if household was referred
publicbenefit	public_benefit	Indicates whether the household received public benefits.	Text	Yes	Transformed to two binary variables: • is_public_benefit: 1 if household receives public benefits, 0 if not • not_public_benefit: 1 if household does not receive public benefits, 0 if household does receive public benefits
povertylevel	poverty_level_percent	Ratio indicating applicant's income relative to the poverty threshold.	Percent	84.95 %	N/A
annualizedtotalincome	annual_total_income	The total annual income for the household.	Numeric (in $)	$10,245.36	N/A
householddisconnected	HH_disconnected	Indicates whether the household's energy source was disconnected.	Text	NA	Transformed to one binary variable: • is_disconnected: 1 if household indicated as disconnected, 0 if not

householdoutoffuel	HH_out_of_fuel	Indicates whether the household is out of fuel.	Text	NA	Transformed to one binary variable: • is_out_of_fuel: 1 if household is indicated as out of fuel, 0 if not
pastdueordisconnectionnoticerece	HH_pastdue_disconnect	Indicates whether the household received a past due or disconnection notice for utilities.	Text	No	Transformed to one binary variable: • is_pastdue_disconnect: 1 if household is indicated as receiving past due or disconnection notice, 0 if not
lessfuelremaining	HH_less_fuel_remain	Indicates whether the household currently has less than 20% of fuel remaining.	Text	NA	Transformed to one binary variable: • is_less_fuel_remain: 1 if household indicated as less than 20% fuel remaining, 0 if not
business name	N/A	The name of the business that processed the application.	Text	WOOD CO HSD	N/A
business primaryaddress	N/A	The primary address of the business that processed the application.	Text	111 W JACKSON ST	N/A
businesssecondary address	N/A	The secondary address of the business that processed the application.	Text	FL 2	N/A
business city	N/A	The city where the business that processed the application is located.	Text	WISC RAPIDS	N/A
businesstate	N/A	The state where the business that processed the application is located.	Text	WI	N/A
business zip	N/A	The zip of the business that processed the application.	String	54495-2702	N/A

business phone	N/A	The phone of the business that processed the application.	String	(715) 421-8654	N/A
heatvendorname	N/A	The name of the vendor providing heat services to the applicant.	Text	WE ENERGIES	N/A
heatvendorprimaryaddress	N/A	The primary address of the vendor providing heat services to the applicant.	Text	231 W MICHIGAN ST	N/A
heatvendorsecondaryaddress	N/A	The secondary address of the vendor providing heat services to the applicant.	Text	NA	N/A
heatvendorcity	N/A	The city where the vendor providing heat services to the applicant is located.	Text	MILWAUKEE	N/A
heatvendorstate	N/A	The state where the vendor providing heat services to the applicant is located.	Text	WI	N/A
heatvendorphone	N/A	The phone of the vendor providing heat services to the applicant.	String	(800) 842-4565	N/A
heatvendorpaymentmethod	N/A	The method of payment used for the heat vendor.	Text	Directly pay the bill sent from the energy supplier.	N/A
nonheatvendorname	N/A	The name of the vendor providing non-heat services to the applicant.	Text	WISC RAPIDS WATER WORKS & LGT COMM	N/A
nonheatvendorprimaryaddress	N/A	The primary address of the vendor providing non-heat services to the applicant.	Text	221 16TH ST S	N/A
nonheatvendorsecondaryaddress	N/A	The secondary address of the vendor providing non-heat services to the applicant.	Text	NA	N/A

nonheatvendorcity	N/A	The city where the vendor providing non-heat services to the applicant is located.	Text	WISC RAPIDS	N/A
nonheatvendorstate	N/A	The state where the vendor providing non-heat services to the applicant is located.	Text	WI	N/A
nonheatvendorphone	N/A	The phone of the vendor providing non-heat services to the applicant.	String	(715) 422-9041	N/A
nonheatvendorpaymentmethod	N/A	The method of payment used for the non-heat vendor.	Text	Directly pay the bill sent from the energy supplier.	N/A
bothvendorname	N/A	The name of the vendor providing both heat and non-heat services to the applicant.	Text	NA	N/A
bothvendorprimaryaddress	N/A	The primary address of the vendor providing both heat and non-heat services to the applicant.	Text	NA	N/A
bothvendorsecondaryaddress	N/A	The secondary address of the vendor providing both heat and non-heat services to the applicant.	Text	NA	N/A
bothvendorcity	N/A	The city where the vendor providing both heat and non-heat services to the applicant is located.	Text	NA	N/A
bothvendorstate	N/A	The state where the vendor providing both heat and non-heat services to the applicant is located.	Text	NA	N/A
bothvendorphone	N/A	The phone of the vendor providing both heat and non-heat services to the applicant.	String	NA	N/A

bothvendorpaymentmethod	N/A	The method of payment used for the vendor providing both heat and non-heat services to the applicant.	Text	NA	N/A
primaryfuel	primary_fuel	The type of energy fuel used for heating, or for both heating and non-heating purposes.	Text	Natural Gas	Transformed to five binary variables: • primary_fuel_elec: 1 if primary fuel is electric, 0 if not • primary_fuel_oil: 1 if primary fuel is fuel oil, 0 if not • primary_fuel_gas: 1 if primary fuel is gas, 0 if not • primary_fuel_prop: 1 if primary fuel is propane, 0 if not • primary_fuel_other: 1 if primary fuel is other type, 0 if not
nonheatfuel	non_heat_fuel	The type of energy fuel used for purposes other than heating.	Text	Electric	Transformed to one binary variable: • non_heat_fuel_elect: 1 if non-heat energy fuel is electricity, 0 if not
waterheatingfuel	water_heat_fuel	The type of energy fuel used for water heating.	Text	Natural Gas	N/A
supplementalheatfuel	supp_heat_fuel	The type of fuel used for supplemental heating purposes.	Text	None	N/A
primaryfueldeliverable	primary_fuel_deliverable	Indicates whether the primary energy fuel is deliverable.	Text	No	Transformed to one binary variable: • is_primary_deliverable: 1 if the primary fuel type is deliverable, 0 if not

haselectricprovider	has_electric_provider	Indicates whether the applicant has an electricity provider.	Text	Yes	Transformed to one binary variable: • has_electric: 1 if household has electricity provider, 0 if not
sharedmeterprimaryfuel	shared_meter_primary	Indicates whether the applicant uses a shared meter for the primary fuel.	Text	No	Transformed to one binary variable: • is_shared_meter: 1 if household uses shared meter for primary fuel, 0 if not
businessmeterprimaryfuel	business_meter_primary	Indicates whether the applicant uses a business meter for the primary fuel.	Text	No	Transformed to one binary variable: • is_business_meter: 1 if household uses business meter for primary fuel, 0 if not
heatfuelannualcost	heatfuel_annual_cost	The annual cost of fuel used for heating.	Numeric (in $)	$814.82	N/A
nonheatfuelannualcost	nonheatfuel_annual_cost	The annual cost of fuel used for non-heating purposes.	Numeric (in $)	$1,228.85	N/A
bothfuelannualcost	bothfuel_annual_cost	The annual cost of fuel used for both heating and non-heating purposes.	Numeric (in $)	NA	N/A
reasonfornoannualcostforprimaryf	reason_no_annualcost_primary	Reason for not having an annual cost for the primary fuel.	Text	NA	N/A
reasonfornoannualcostfornonheatf	reason_no_annualcost_nonheat	Reason for not having an annual cost for non-heating fuel.	Text	NA	N/A
pbproxyamount	N/A	A value calculated based on dwelling type, size and heating fuel.	Numeric (in $)	$1,044.00	N/A

pbproxydescription	pb_proxy_desc	Describes how public benefit proxy amount is used.	Text	NA	N/A
heatproxyamount	N/A	A value calculated based on dwelling type, size and heating fuel.	Numeric (in $)	$746.00	N/A
heatproxydescription	heat_proxy_desc	Describes how heat proxy amount is used.	Text	NA	N/A
crisisservicetype	crisis_service_type	Indicator of the type of crisis assistance received.	Text	• Summer Fill Payment • Summer Fill Match Payment • Spring Prevention Assistance Payment • Prevention Assistance Payment • Prevention Arrearage Assistance • PB Prevention Electric Payment • PB Emergency Electrical Assistance Payment • Emergency Assistance Payment • CA Prevention Assistance Payment • CA Emergency Assistance Payment	Transformed to the following binary variables: • received_SF: 1 if value is 'Summer Fill Payment' or 'Summer Fill Match Payment', 0 if not • received_spring_prevention_assistance_payment: 1 if value is 'Spring Prevention Assistance Payment', 0 if not • received_prevention_assistance_payment: 1 if value is 'Prevention Assistance Payment', 0 if not • received_prevention_arrearage_assistance: 1 if value is 'Prevention Arrearage Assistance', 0 if not • received_pb_prevention_electric_payment: 1 if value is 'PB Prevention Electric Payment', 0 if not • received_pb_emergency_electrical_assistance: 1 if value is 'PB Emergency Electrical Assistance Payment', 0 if not • received_emergency_assistance: 1 if value is 'Emergency Assistance Payment', 0 if not

					• received_CA_prevention_assistance: 1 if value is 'CA Prevention Assistance Payment', 0 if not • received_CA_emergency_assistance: 1 if value is 'CA Emergency Assistance Payment', 0 if not

Appendix B: Rural-urban data elements

Variable	**Description**
State-County FIPS Code	Federal Information Processing Standards (FIPS) code that uniquely identifies the geographic area.
State	State of given Census tract area.
County	County of given Census tract area.
State-County-Tract FIPS Code	State-level FIPS codes of given Census tract area. State-County-Tract FIPS Codes have two digits, county-level FIPS codes have five digits of which the first two are the FIPS code of the state to which the county belongs.
Primary RUCA Code 2010	Rural-Urban Commuting Area (RUCA) code allocated to given Census tract area.
Tract Population, 2010	Population of given Census tract area.
Land Area (square miles), 2010	Land area (in square miles) of given Census tract area.
Population Density (per square mile), 2010	Population density (per square mile) of given Census tract area.

Appendix C: U.S. Census Bureau data on physical housing characteristics

Variable	Description
Geographic Area Name	Zipcode of area
Year structure built: 1939 or earlier	The number of structures within a given zip code that were built in 1939 or earlier.
Year structure built:1940 to 1959	The number of structures within a given zip code that were built between 1940-1959.
Year structure built: 1960 to 1979	The number of structures within a given zip code that were built between 1960-1979.
Year structure built: 1980 to 1999	The number of structures within a given zip code that were built between 1980-1999.
Year structure built: 2000 to 2009	The number of structures within a given zip code that were built between 2000-2009.
Year structure built: 2010 to 2019	The number of structures within a given zip code that were built between 2010-2019.
Year structure built: 2020 or later	The number of structures within a given zip code that were built 2020 or later.