Reductions in Medicaid Spending Associated with Income-Eligible Home Weatherization in the Southeastern U.S.

Michaela Marincic, Three[3], Inc.
Erin Rose, Three[3], Inc.
Bruce Tonn, Three[3], Inc.
Palash Arora, 84.51°

ABSTRACT

Non-energy impacts (NEIs) related to health and well-being have long been supported by self-reported survey results for energy programs and income-eligible home weatherization in particular. However, recent efforts internationally and within the U.S. have sought to link health records and medical claims with weatherization participation data to examine program impacts on actual healthcare spending. The current study drew on results from a survey of participants in the Tennessee Valley Authority's Home Uplift, a low-income weatherization program in the Southeastern U.S., and linked their survey and weatherization measure data to Medicaid records on healthcare utilization and spending. Results indicated that receiving free weatherization services was associated with a short-term drop in healthcare costs in the first 12 months, as well as a longer-term stabilization of healthcare costs. This study provides initial evidence of cost reductions associated with low-income weatherization programs, which often do not pass cost-effectiveness tests based on energy savings alone. It also supports ongoing efforts to engage the healthcare sector in addressing social determinants of health (SDOH) as preventative care for their patients. Ensuring access to safe and healthy housing contributes to managing both chronic and acute health issues and reducing overall healthcare spending, particularly for Medicaid, which provides critical healthcare access for vulnerable populations.

Introduction

In addition to the primary goal of lowering energy costs, home weatherization programs can improve indoor temperatures (Alam et al., 2018; Poortinga et al., 2017); reduce infiltration of air pollution and environmental hazards (Rose et al., 2015); and increase general comfort for residents (International Energy Agency, 2025; Rose et al., 2022). Combined, these improvements in the indoor environment and the reduced financial strain from lower utility bills may contribute to better health for residents (Rose, et al., 2022, Tonn et al., 2021). The current study measures changes in health-related survey responses and Medicaid costs and claims among income-eligible households that received free, comprehensive weatherization services through the Tennessee Valley Authority's Home Uplift initiative.

Background on the Home Uplift Initiative

The Home Uplift initiative was a low-income weatherization project funded by TVA and participating local power companies (LPCs) in its jurisdiction. Energy efficiency upgrades through Home Uplift weatherization typically included air sealing and insulation measures, heating and air conditioning equipment maintenance and replacement, heat pump water heater installation, window and door replacement, refrigerator upgrades, LED bulbs, and low-flow showerheads. Pilot locations represented in the study included Knoxville, Nashville, Chattanooga, and Memphis, Tennessee; Huntsville, Alabama; and the service territories of 4-County Electric Power Association, Mississippi and Western Kentucky Rural Electric Cooperative. Home Uplift's eligibility requirements and weatherization procedures were modeled after the U.S. Department of Energy's Weatherization Assistance Program.

As part of the Home Uplift pilot, TVA contracted with Three[3] to estimate the health, well-being and social determinants of health (SDOH) benefits attributable to the Home Uplift pilot (Rose et al., 2022). This evaluation involved the design of a survey and data analysis plan to quantify the health and well-being benefits of households and to then monetize a set of NEIs based on observed practical and significant benefits. In partnership with the University of Tennessee, Knoxville's Center for Applied Research and Evaluation (UT CARE), a total of 701 Home Uplift households were surveyed across seven pilot locations between 2018 and 2021. During the first year of surveys, 300 low-income, unweatherized households were also surveyed from within the same metro areas as the treatment group households to form a control group. Follow-up surveys were administered one year after the baseline survey for both treatment and control homes.

Addressing bias in treatment assignment. A central challenge in evaluating the Home Uplift initiative is that the treatment and control groups were not equivalent at baseline. Tables 1 through 9 of the original non-energy impacts evaluation (Marincic et al. 2025) document that households selected to receive Home Uplift weatherization reported substantially worse housing and environmental conditions at the time of the initial survey than the unweatherized control households recruited from the same metro areas. At baseline, treatment households reported higher rates of broken primary heating equipment (30.8% versus 14.4% in the control group) and broken central air conditioning (35.5% versus 21.6%); they were also more likely to report that their homes were cold or very cold in winter (57.1% versus 35.9%) or hot or very hot in summer (47.7% versus 27.3%), and they reported more drafts and dust. Consistent with these worse thermal conditions, the treatment group also reported elevated baseline rates of healthcare visits for cold- and heat-related thermal stress. In short, the households that received weatherization were, on average, in worse housing and health circumstances before any intervention took place.

A true randomized holdout was not feasible in this setting. Control households were recruited from the same metro areas as treatment households, but they were not randomly assigned to forgo weatherization. Instead, selection into Home Uplift depended on income eligibility, referral timing, and application processes that themselves may correlate with the severity of a household's housing problems and health status. Households experiencing the most acute energy and comfort problems may have been more likely to seek out and qualify for the program. As a result, naive pre-post or difference-in-differences comparisons risk confounding genuine treatment effects with these pre-existing baseline differences between the groups.

These baseline imbalances motivated a complementary analysis that applies causal debiasing methods: augmented inverse propensity weighting (AIPW) and double machine learning (DML) directly to the original survey data. Rather than assuming the groups are comparable, these methods explicitly model and adjust for measured baseline differences when estimating treatment effects on health and well-being outcomes. The analysis is described below and is intended to sit alongside, not replace, the Medicaid claims analysis: the two draw on different data sources and rely on different identifying assumptions, and together they provide a more complete picture of the program's health impacts.

Data Sources and Study Participants

Data for the current study were drawn from three sources: Medicaid data from the State of Tennessee Division of TennCare; survey data from two evaluations of the Home Uplift pilot (later a full program); and data on the measures installed in Home Uplift homes. Survey data and installed measures were collected during two prior research studies: the 2017-2020 evaluation of TVA's Home Uplift (Rose et al., 2022), which was still a pilot program at the time, and a follow-on study funded by the Tennessee Department of Environment and Conservation (TDEC) from 2021-2023 that placed temperature and humidity sensors in homes about to undergo weatherization through Home Uplift (Marincic and Rose, 2023), now a fully launched program. Participants in the TDEC-funded study also completed a shorter survey with questions that partially overlapped the survey from the initial Home Uplift evaluation, allowing the team to combine results from both surveys.

Both weatherized (treatment) and non-weatherized (control) participants for the current study were recruited from the Home Uplift pilot evaluation and TDEC IEQ study. Participants gave informed consent to request and acquire their linkable medical claims data from Tennessee's Medicaid office, known as TennCare. A data sharing agreement was executed between Three[3] and the State of Tennessee's Division of TennCare for the secure transfer of identifiable Medicaid claims data for select diagnostic codes. The data sharing agreement and research design were also governed by the Institutional Review Board of the University of TN, Knoxville.

One limitation of this approach is that it does not capture any medical claims or costs billed to other forms of insurance, in the event that a participant has backup insurance or a different primary insurance. In 2021, an estimated 15.7% of the population had coverage through both Medicaid and a private insurer, and another 15.7% had coverage through both Medicaid and Medicare (Mykyta et al., 2023). A second limitation is that TennCare did not provide information on disenrollment, leaving ambiguity as to whether periods with no claims represent a time where participants are not utilizing healthcare or whether the participant became disenrolled from Medicaid temporarily or permanently. In 2024, a federal judge ruled that Tennessee had wrongfully terminated Medicaid coverage for thousands of Tennesseans beginning in 2019 (Pierson, 2024). The class action lawsuit, filed in 2020, requested the state re-enroll roughly 108,000 people who were likely wrongfully terminated from coverage (Kelman 2022).

A third limitation is that the timespan of the data covers the duration of the COVID-19 pandemic, an international health crisis. A national emergency and international travel ban was declared in the United States on March 13, 2020. Testing data from Tennessee indicate that the last known major spike in cases ended in February 2022, by which point 76% of the U.S. population had received at least one dose of a COVID-19 vaccine and 65% were fully vaccinated; in Tennessee, these rates were 61% and 54%, respectively. Using these timepoints, we can look at our participants in three groups: those weatherized before the start of the pandemic (n = 25), those weatherized during the pandemic (n = 6), and those weatherized after the pandemic (n = 16). Table 1 presents the breakout of how many months of data were before, during, or after the pandemic. The majority of pre-treatment data (70%) took place prior to the pandemic, whereas the post-treatment data were almost evenly split across the three time periods. Because more of the post-treatment data took place during and after the pandemic, we might expect higher medical costs and claims in the post-treatment period than would have otherwise been the case.

Table 1. Percentage of Pre-Treatment and Post-Treatment Data Occurring Before, During, and After the COVID-19 Pandemic

Pandemic Era	Pre-Treatment (Months of Data)	Post-Treatment (Months of Data)
Before COVID-19	496 (70%)	269 (36%)
During COVID-19	157 (22%)	214 (28%)
After COVID-19	56 (8%)	269 (36%)

Description of the Medicaid Data

Variables included in the data received from TennCare that were used for the current analysis, as well as three calculated variables, are summarized in Table 2. Variables that were provided by TennCare but not used during the course of analysis are not included in the table below.

Table 2. Summary of Key Variables Received from TennCare and Calculated Variables

Variable Name	Variable Description
DATE_OF_BIRTH	Date of birth (used to calculate age)
SERVICE_DATE_OF_CLAIM	Date services were rendered and claim was made
TYPE_OF_CLAIM	High-level description of care setting. Includes four values: Inpatient, Outpatient, Professional or Pharmacy
PAID_UNITS	Number of units paid for, if applicable (e.g. number of doses in a pharmacy prescription)
PAID_AMOUNT	High-level amount paid by Medicaid for the claim
PAID_DTL_AMOUNT	Additional detail on amounts paid by Medicaid
RX_PAID_AMOUNT	Amount paid by Medicaid for prescriptions
IMPUTED_PAID_AMOUNT	CALCULATED: Harmonized variable representing total cost for a medical encounter
INFL_IMPUTED_PAID_AMOUNT	CALCULATED: IMPUTED_PAID_AMOUNT adjusted for inflation to the year 2024
Number_of_Claims	CALCULATED: Number of individual claims/line items included in the medical encounter

The TennCare Medicaid claims data underwent four major steps to prepare for analysis:

1. *Cost data from three columns were harmonized into one variable (IMPUTED_PAID_AMOUNT).* The raw Medicaid data contained three variables with costs data: PAID_AMOUNT, PAID_DTL_AMOUNT, and RX_PAID_AMOUNT (for prescription medications). The first two variables had identical values for the majority of claims; in these cases, the value listed was imputed as the total amount paid for services. Where PAID_AMOUNT and PAID_DTL_AMOUNT did not match, only one or the other was filled in and this value was used as the cost of services. For pharmacy claims, IMPUTED_PAID_AMOUNT was populated from RX_PAID_AMOUNT. This created a unified variable for all medical costs, regardless of whether it was a cost for services or for pharmacy prescriptions.

2. *Cost data were adjusted for inflation.* Using the Consumer Price Index (CPI), the team adjusted all medical costs to 2024 U.S. dollars.
3. *Individual claims were grouped by medical encounter.* A single visit to the doctor's office can result in multiple claims for individual procedures. Often, only the primary claim/procedure would have an associated cost, and the remaining claims from the same visit would have no price listed. Therefore, all claims that took place at the same location, on the same day, with the same diagnostic codes, and for the same type of service (i.e. professional, pharmacy, outpatient or inpatient) were summarized into a single "encounter" and assigned one paid amount.
4. *Additional variables for analysis were calculated.* For analysis, participants' age was calculated in two ways: one variable represented age at time of encounter, and one variable represented age at treatment. Age at treatment was used for pre-post analyses to designate a group of "near-elderly" participants who were age 55 or higher when they received weatherization work.

Methodology

Pre-Post Comparisons

For the initial analysis, data were restricted to the period 12 months prior to weatherization and 12 months after weatherization; however, data from the month immediately following weatherization were removed to provide a buffer period for two reasons: 1) most health effects would not manifest immediately after weatherization occurs; and 2) the process of weatherization may introduce additional dust and outdoor pollen tracked in as workers go in and out of the house, but these impacts would be temporary. A 12-month period was chosen to capture all seasons only once and control for any seasonal variation in health. The same approach was used for control group homes with the exception that, since these homes did not receive weatherization during the study period, data were instead restricted to 12 months prior to their initial survey completion and 12 months after survey completion to ensure the data fell within a comparable timeline to the weatherized participants; this also helps to control for any influence the survey itself may have on behavior, though any effects from taking the survey would be expected to be minimal.

The pre-post analysis examined both medical costs and the number of medical claims reported for each participant. For each outcome, the data were analyzed for each group separately (weatherized vs. non-weatherized) and also as a comparison across groups. To compare pre and post outcomes within each group, the team used a Wilcoxon rank sum test, also known as a Mann-Whitney U. This test was preferred due to the outcomes consisting of count data that were not normally distributed; Wilcoxon rank sum is non-parametric and does not assume the data are normally distributed (bell-shaped) like many other statistical tests. Next, the change in the weatherized group was compared against the change in the control group using a difference-in-differences (DID) model; the pre-treatment data were subtracted from the post-treatment data for each group, and the resulting differences were compared using the Wilcoxon rank sum test. All pre-post and time series analyses were performed in R version 4.5.0 in RStudio 2026.01.0+392 on a standard MacBook computer.

Time Series Analysis

While the pre-post analysis was limited to a 24-month window for each person, the full Medicaid data covered nearly eight years in total. To make use of all data, the team also used Interrupted Time Series (ITS) analysis. This model focuses on trends over time and how the trend in medical costs and claims changes following weatherization. However, the current analysis differs from a traditional Interrupted Time Series model in that the “interruption” (weatherization) did not happen at the same time for all participants, but rather weatherization dates for the participants ranged from July 2018 to September 2022.

To prepare for the time series analysis, the team separately added up all medical costs and claims for each month and for each person. For months where a participant had no claims or costs, a zero was added to the data. However, because TennCare did not provide enrollment or disenrollment dates, a gap of 12 months or more was interpreted to mean the person was no longer on Medicaid during that time and zeroes were not imputed. The team additionally converted the time variable from dates to a relative index of how many months before or after treatment the observation occurred ("Time from Treatment"). As suggested in Fusi and Lechi 2020, the team then modified the Time from Treatment variable to equal zero for all months prior to weatherization; we renamed the variable “Time after Treatment” to distinguish the two. A “Month” variable uniquely numbered each month from 0 to 97 for the data, which run from January 2016 to February 2024. This numbering is the same across all households regardless of intervention date. This helps to control for any external events that may have affected some or all homes in a given month.

In addition to the adapted time variables, the team computed six new variables for inclusion in the models: a lag variable representing costs from the prior month; a lag variable representing the number of claims from the prior month; three binary season variables for summer, fall, and winter (“Season” for short); and a COVID-19 binary variable where zero indicates the time before March 2020 and after January 2022 and 1 indicates the time between March 2020 and January 2022, inclusive. The seasons were defined as: December through February for winter; March through May for spring; June through August for summer; and September through November for fall.

Using these modified time variables, the teams specified models with four different sets of variables and three different estimators for a total of twelve models. The three estimators applied were Poisson regression using glm(); Zero-Inflated Poisson (ZIP) regression using zeroinfl(); and robust regression with an M estimator using rlm() from the ‘MASS’ package version 7.3.65. Poisson regression was chosen because a Poisson distribution is specifically designed to model count data, such as number of claims or dollars in our study. Zero-Inflated Poisson (ZIP) regression was chosen as an extension of Poisson regression because a person might have zero costs/claims because they simply didn’t go to the doctor or fill a prescription that month, or a person might also have zero costs/claims because they were disenrolled from Medicaid or used secondary insurance (some participants may have private insurance coverage in addition to Medicaid). ZIP regression attempts to distinguish between the two sources of zeroes to more accurately estimate impacts on the outcome of interest. The robust regression was chosen due to the few but large outliers in the medical cost data in particular; the team chose not to remove these outliers because they are known to be real data reflecting actual costs. The four model specifications were:

1. Outcome ~ Period + Month + Time after Treatment

2. Outcome ~ Period + Month + Time after Treatment + Season
3. Outcome ~ Period + Month + Time after Treatment + Season + Lag
4. Outcome ~ Period + Month + Time after Treatment + Season + Lag + COVID

The three models estimated with ZIP regression that included lag variables returned singularity errors and test statistics could not be computed; therefore, the results below include only the Poisson and robust regression models. Statistical significance for the other nine models was estimated using Heteroskedasticity and Autocorrelation Consistent (HAC) covariance matrix estimation through the vcovHAC() function from 'sandwich' version 3.1-1 together with coeftest() from 'lmtest' version 0.9-40. Models were compared using AIC() and McFadden's pseudo R^2.

Survey-Based Causal Analysis of Health and Well-Being Outcomes

To complement the Medicaid claims analysis, the team conducted a separate causal analysis using the self-reported health outcomes collected in the Home Uplift survey. This analysis draws on a different data source (the survey rather than Medicaid claims) and a different set of methods (causal debiasing estimators rather than pre-post and time series models), and is reported here as a parallel line of evidence. Its goal is to estimate the effect of receiving Home Uplift weatherization on occupant health and well-being while formally adjusting for the baseline imbalances described above.

Outcome variables. The survey included the four items from the U.S. Centers for Disease Control and Prevention (CDC) Healthy Days measure, a widely used set of self-reported health-related quality of life questions. Respondents reported the number of days in the past 30 days that: they did not get enough rest or sleep (c1); their physical health was not good (c2); their mental health was not good (c3); and poor physical or mental health kept them from doing their usual activities, such as self-care, work, or recreation (c4). Each item ranges from 0 to 30 days and was collected at baseline (T1, before weatherization) and at one-year follow-up (T2, after weatherization).

For the physical health (c2), mental health (c3), activity limitation (c4), and sleep (c1) items, each follow-up outcome was dichotomized at a threshold of 14 or more days. For c2, c3, and c4, this corresponds to the CDC-validated definitions of "Frequent Physical Distress," "Frequent Mental Distress," and "Frequent Activity Limitation," respectively, following the CDC PLACES measure definitions (CDC n.d.) and the supporting peer-reviewed rationale (Chen et al. 2011), which applies the same threshold across all three measures and notes that clinicians and researchers commonly use a two-week symptom duration as a marker for clinically significant depression or anxiety, and that longer symptom duration is associated with greater activity limitation. The sleep item (c1) is not part of the CDC HRQOL-4/HRQOL-14 core module and has no independently validated 14-day cutpoint in the literature. For consistency with the other three Healthy Days measures, the same 14-day threshold was applied.

The focal hypotheses concern c1 (sleep) and c3 (mental health), which showed the most plausible and consistent treatment effects across estimators (see Results). The physical health (c2) and activity limitation (c4) outcomes are reported as exploratory; they did not yield significant or consistent effects and are discussed as findings warranting future work.

Augmented Inverse Propensity Weighting and Double Machine Learning. The objective of both methods is to estimate the Average Treatment Effect (ATE) of receiving Home Uplift weatherization, relative to the control group, on each Healthy Days outcome, while explicitly correcting for the non-random treatment assignment described in the Background. The analysis sample comprised 779 survey respondents (the primary respondent in each household with a valid follow-up response), of whom roughly 72% were in the treatment group and 28% in the control group. All models were estimated in Python using LightGBM gradient-boosted models with five-fold cross-fitting.

Augmented inverse propensity weighting (AIPW), also known as the doubly robust or DR-learner estimator, combines two models: a propensity score model, which predicts the probability that a household received weatherization given its baseline characteristics, and an outcome regression model, which predicts the health outcome given baseline characteristics and treatment status. The "doubly robust" property means the resulting estimate remains valid even if one of these two models is imperfectly specified, as long as the other is approximately correct. Both the propensity and outcome models were fit with LightGBM, and five-fold cross-fitting was used so that each household's prediction came from a model trained on other households, guarding against overfitting bias.

AIPW relies on the positivity, or overlap, assumption: every household must have had a non-negligible probability of being in either the treatment or the control group for a valid comparison to be possible. To enforce this, estimated propensity scores were trimmed to the range 0.05 to 0.95 (np.clip(ps, 0.05, 0.95)). Trimming limits the influence of households whose treatment status was nearly deterministic—cases that would otherwise receive extreme inverse-probability weights and destabilize the estimate.

Double machine learning (DML) was used as a complementary robustness check. DML partials out the influence of baseline covariates from the treatment indicator and from the outcome separately—by taking the cross-fitted residuals of each—and then estimates the treatment effect from the relationship between the two sets of residuals. Unlike AIPW, the DML outcome model does not include treatment status as a predictor. Because AIPW and DML rely on somewhat different identifying assumptions and model structures, agreement between them provides reassurance that a result is not an artifact of one particular specification.

Covariates for both models were selected in two steps. First, candidate predictors were ranked by LightGBM feature importance. Second, the candidate set was filtered to exclude any variable measured after weatherization, so that the models would not condition on outcomes of the treatment itself; only baseline (T1) measures were retained. The retained covariates fell into three conceptual groups: demographic and housing-tenure characteristics (age; length of residency in the home, a1; and highest education level, l2); baseline thermal conditions (winter indoor daytime temperature, b2, and summer indoor daytime temperature, b3—direct markers of the housing problem that weatherization addresses); and the outcome's own baseline value (for example, baseline sleep days when modeling follow-up sleep). Including each outcome's baseline value follows an ANCOVA-style adjustment that improves precision and is robust to partial missingness in the baseline data. The propensity (treatment) model used the same covariate set across all four outcomes. The outcome models used this same core set of covariates, with the only outcome-specific variation being the inclusion of each outcome's own baseline value, summarized in Table 8. All outcome models additionally include the binary treatment indicator. The variable codes in Table 8 represent the following:

- a1 = length of residency in the home;
- age = respondent age;
- b2 = winter indoor daytime temperature;
- b3 = summer indoor daytime temperature;
- l2 = highest education level;
- c1–c4 = baseline values of the corresponding Healthy Days items.

Table 8. Baseline Covariates Used in the Propensity and Outcome Models, by Healthy Days Outcome.

Outcome	Propensity-model covariates	Outcome-model covariates
c1 – Sleep	a1, age, b2, b3, l2	a1, age, b2, b3, c1 (baseline sleep)
c3 – Mental health	a1, age, b2, b3, l2	a1, age, b2, b3, c2 (baseline mental)
c2 – Physical health	a1, age, b2, b3, l2	a1, age, b2, b3, c2 (baseline health)
c4 – Activity limitation	a1, age, b2, b3, l2	a1, age, b2, b3, c4 (baseline activity)

Results

The results below examine the costs paid by Medicaid and the number of claims billed to Medicaid across participants as well as for specific subgroups of interest. Changes in costs and claims are presented by weatherization status, measure type, ability to pay utility bills, age, setting of care, and health condition.

Pre-Post Analysis

Total medical costs decreased in both the treatment and control groups (sample sizes: N = 50 and N = 9, respectively) when looking at both average and the median costs. The decrease was not statistically significant for either group, or when comparing across groups. Conversely, the average number of medical claims for the control group stayed constant and the median increased, whereas the weatherized homes saw a reduction in the number of claims reported (though again, neither was statistically significant).

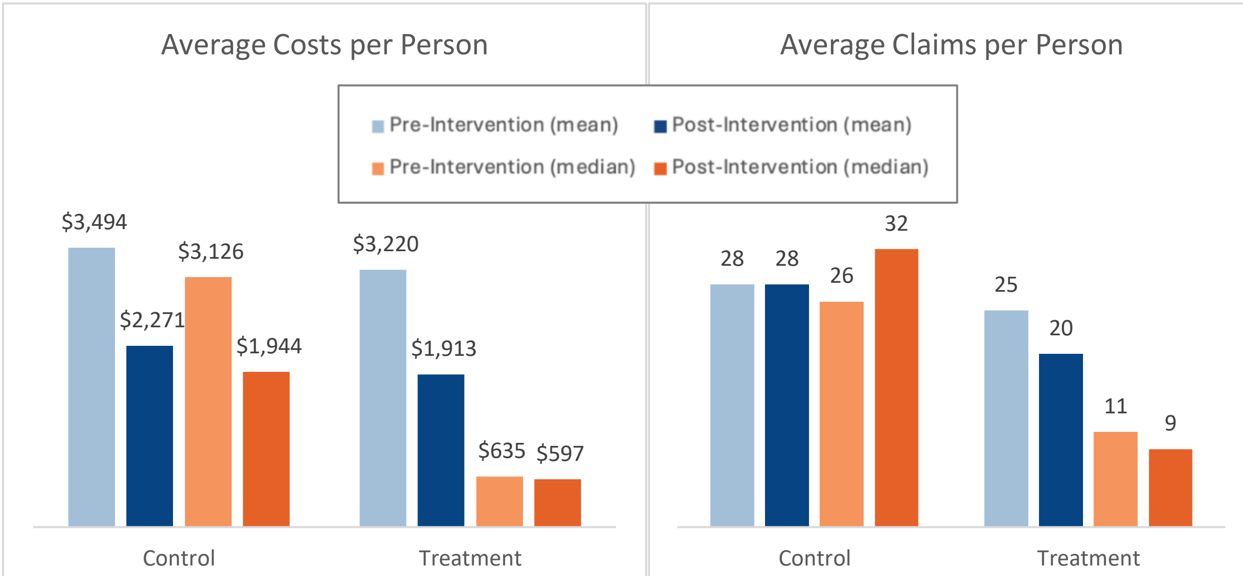

*Figure*1. *Total Medicaid Costs (left) and Claims (right) per Person, Pre- and Post-Treatment*

In addition to overall results, the team examined the set of homes that received at least one major weatherization measure, defined as new heating or cooling equipment, insulation, air sealing, or ductwork repair or replacement; 45 of the 50 treatment homes met this criterion. Impacts to costs and number of claims were similar to the previous analysis, with reductions across the board; however, this time the reduction in medical claims was marginally significant at p = .07.

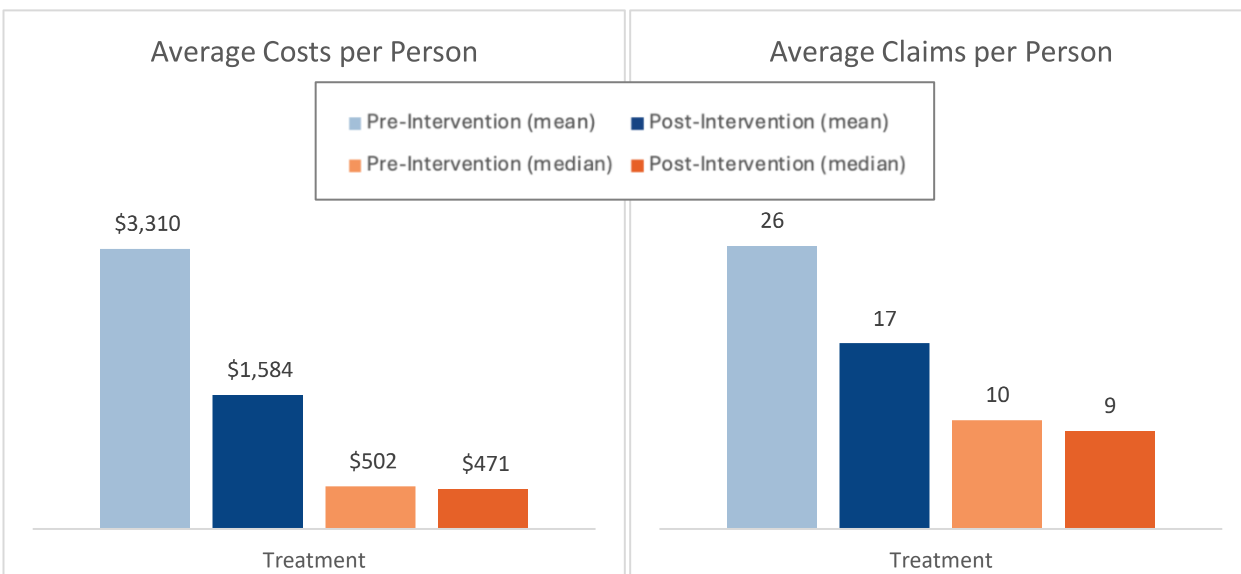

Figure 2. Medicaid Costs and Claims for Households that Received Major Weatherization Measures

The team additionally assessed impacts for households who reported they found it "Hard or Very Hard" to pay their utility bills during the initial, pre-weatherization survey (N = 13). The same pattern of reductions in costs and claims held, with no statistically significant results.

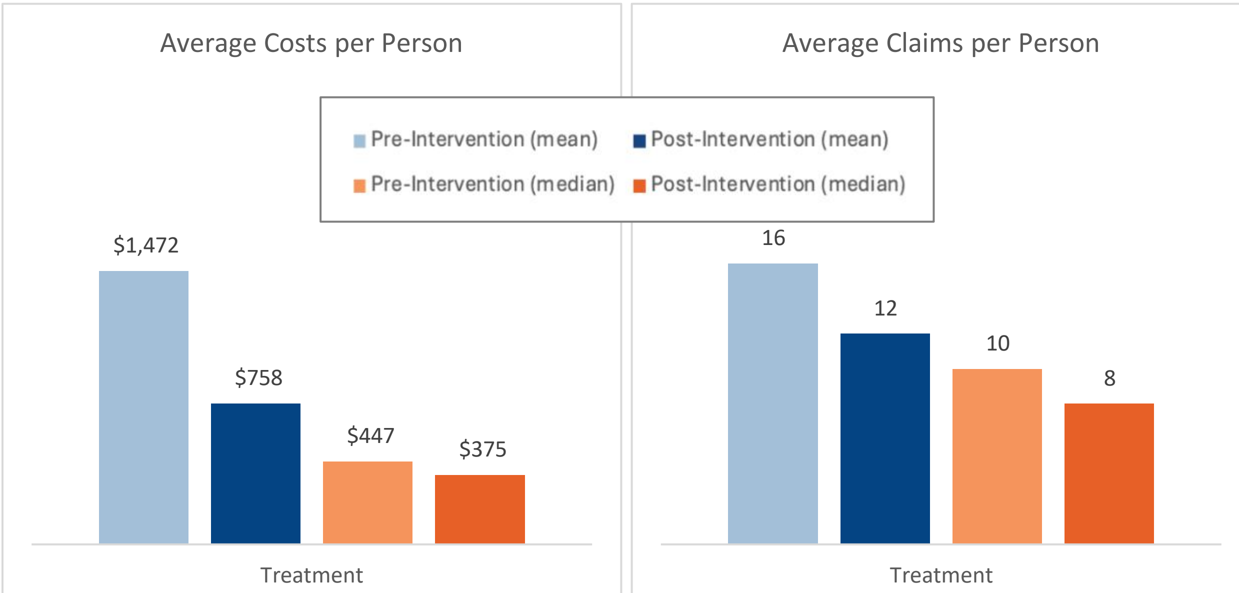

Figure 3. Medicaid Costs and Claims for Homes that Reported it was "Hard" or "Very Hard" to Pay Utility Bills prior to Weatherization

The team also examined results for participants who were age 55 or older at time of treatment (N = 11), as previous research indicates this to be a particularly vulnerable population. Beginning around age 55, people may experience the health and physical impacts of aging and find it more difficult to work at paid jobs, but the majority of the social safety net has not yet activated for them (e.g. Medicare begins at age 65). This leaves the group of "near-elderly" at risk of financial and health burdens. Notably, this group saw a drastic and statistically significant reduction (p = .01) in the average number of medical claims per person, from 43 claims before weatherization to 21 after.

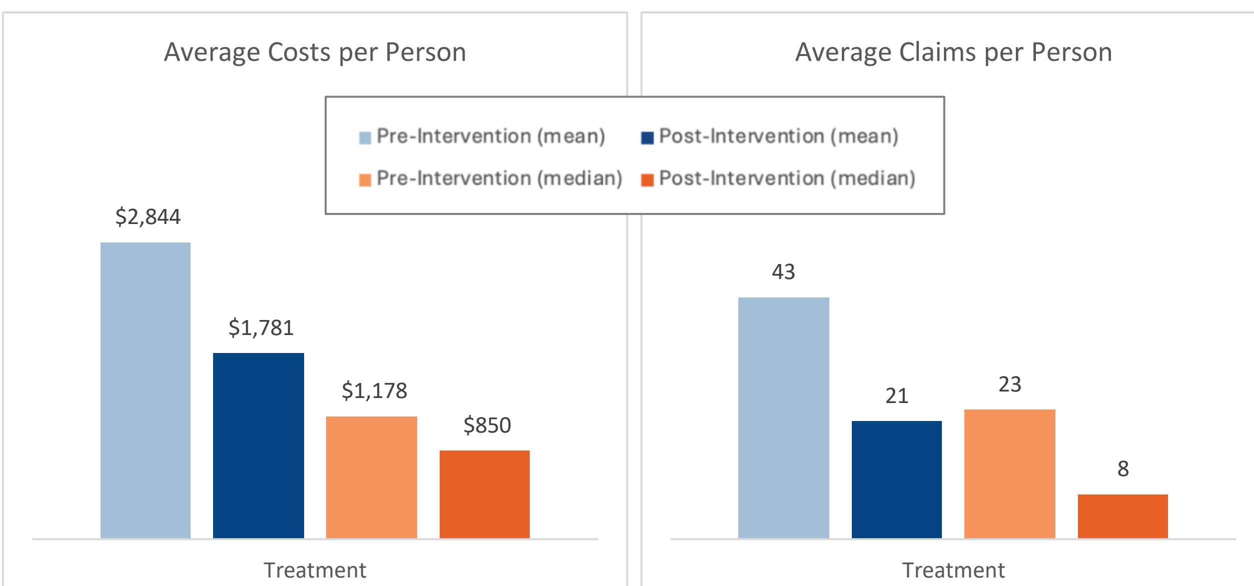

Figure 4. Medicaid Costs and Claims for Weatherization Recipients Age 55 and Over

Time Series Analysis

In the time series analyses, the Poisson regression models consistently showed a modest drop in costs immediately following weatherization as well as a change in the trend to more quickly rising costs over time; neither result was statistically significant (Table 3).

Table 3. Model Comparison for Poisson Regressions, Medicaid Costs 2016-2024

Variable	Model 1	Model 2	Model 3	Model 4
Intercept	5.932***	6.694***	6.646***	6.657***
Month	-0.004	-0.006	-0.006	-0.005
Period	-0.389	-0.335	-0.327	-0.402
Time after Treatment	0.017	0.017	0.017	0.022
Summer		-0.701	-0.670	-0.683
Fall		-1.169^	-1.135	-1.149
Winter		-1.536*	-1.506*	-1.505*
Lag			0.00003^	0.00003
COVID				-0.178
McFadden's pseudo R^2	0.009	0.049	0.054	0.055
AIC	1,542,826	1,480,388	1,472,300	1,470,497

<.001*** <.01** <.05* <0.1^

Due to the influential, high-cost hospitalization noted above in the baseline analysis, the team also employed regression with an M estimator, designed to be robust to outliers (Table 4). The four models estimated average baseline costs of $112.91 - $145.22 (intercept), with costs rising between 3-18 cents per month across the eight years of data. Costs consistently dropped immediately after weatherization, and the reduction was statistically significant in three of four models. However, the post-weatherization trend (Time after Treatment) showed costs rising more quickly than before weatherization. Spring consistently showed the highest costs as the reference category for the season variables; counter to expectations, winter showed the lowest costs. AIC scores for the robust regressions were two orders of magnitude smaller than for the Poisson regression, indicating superior models, though the lower pseudo R^2 values indicate the robust regressions explain less of the variation in the data; however, neither sets of models had high pseudo R^2 values.

Table 4. Model Comparison for Robust Regressions, Medicaid Costs 2016-2024

Variable	Model 1	Model 2	Model 3	Model 4
Intercept	129.59***	145.22***	119.61***	112.91***
Month	0.036	0.136	0.184	0.026
Period	-51.79^	-49.41^	-44.31**	-27.82
Time after Treatment	3.800*	3.504^	3.018**	1.934^
Summer		-18.56	-13.84	-11.27
Fall		-28.78	-22.49	-19.50
Winter		-80.59*	-68.53***	-66.41***
Lag			0.076	0.077
COVID				31.07^
McFadden's pseudo R^2	0.00003	0.00006	0.00016	0.00014

Variable	Model 1	Model 2	Model 3	Model 4
AIC	26,043	26,048	26,030	26,033

<.001*** <.01** <.05* <0.1^

Applying the same methods to the number of claims in each month, both the Poisson and robust regressions revealed a pattern similar to that found in the costs: a drop immediately after weatherization, followed by an increase in the upward trend of claims (Table 5, Table 6). Spring once again had the most claims and winter the fewest. The robust regressions showed improvement over the Poisson regressions according to AIC but not McFadden's pseudo R^2. Unlike costs, which were highly variable and exhibited large outliers, the claims data had a smaller range and more typical Poisson distribution, such that robust regression may provide fewer advantages; however, the improvement in AIC score is still noteworthy.

Table 5. Model Comparison for Poisson Regressions, Medicaid Claims 2016-2024

Variable	Model 1	Model 2	Model 3	Model 4
Intercept	0.743***	1.009***	0.973***	0.969***
Month	0.006	0.006	-0.006*	-0.006*
Period	-0445*	-0.441*	-0.439**	-0.424**
Time after Treatment	0.018*	0.018*	0.018**	0.017**
Summer		-0.184	-0.163	-0.161
Fall		-0.581**	-0.557***	-0.554***
Winter		-0.786**	-0.765***	-0.765***
Lag			0.00003**	0.00003**
COVID				-0.029
McFadden's pseudo R^2	0.013	0.030	0.033	0.033
AIC	9,061	8,913	8,880	8,882

<.001*** <.01** <.05* <0.1^

Table 6. Model Comparison for Robust Regressions, Medicaid Claims 2016-2024

Variable	Model 1	Model 2	Model 3	Model 4
Intercept	1.648***	2.611***	2.474***	2.418***
Month	0.006	0.003	0.004	0.002
Period	-0.659*	-0.595^	-0.589**	-0.440*
Time after Treatment	0.049**	0.048**	0.047***	0.037**
Summer		-0.937*	-0.884***	-0.856***
Fall		-1.217**	-1.148***	-1.105***
Winter		-1.517**	-1.452***	-1.433***
Lag			0.0002	0.0002
COVID				0.280
McFadden's pseudo R^2	0.0008	0.0027	0.0040	0.0036
AIC	8,241	8,232	8,218	8,223

<.001*** <.01** <.05* <0.1^

De-Biased Survey Analysis

Table 9 reports the estimated ATEs and 95% confidence intervals for each outcome under both AIPW and DML. For the two focal outcomes, both estimators produced negative point estimates indicating a lower probability of frequent distress among weatherized households. The two methods also agreed in direction and were reasonably close in magnitude. For sleep (c1), the AIPW estimate was −0.138 (95% CI −0.357 to 0.081) and the DML estimate was −0.061 (−0.135 to 0.014). For mental health (c3), the AIPW estimate was −0.159 (−0.332 to 0.014) and the DML estimate was −0.037 (−0.096 to 0.023). Neither result reached conventional statistical significance at the 95% level, as each confidence interval narrowly includes zero. However, the cross-method agreement in both direction and approximate magnitude is the key supporting evidence for a plausible—if not conclusive—protective effect of weatherization on sleep and mental health.

The physical health (c2) and activity limitation (c4) outcomes did not show this pattern. The AIPW estimate for physical health was 0.064 (−0.130 to 0.257) and the DML estimate was 0.022 (−0.043 to 0.088); for activity limitation, the AIPW estimate was 0.138 (−0.001 to 0.277) and the DML estimate was 0.05 (-0.008, 0.118). Both were non-significant, and—unlike sleep and mental health—their point estimates fell in the opposite, non-protective direction.

Table 9. Average Treatment Effect (ATE) Estimates for CDC Healthy Days Outcomes, AIPW and DML.

Outcome	AIPW ATE (95% CI)	DML ATE (95% CI)
c1 – Frequent Insufficient Sleep	−0.138 (−0.357, 0.081)	−0.061 (−0.135, 0.014)
c3 – Frequent Mental Distress	−0.159 (−0.332, 0.014)	−0.037 (−0.096, 0.023)
c2 – Frequent Physical Distress	0.064 (−0.130, 0.257)	0.022 (−0.043, 0.088)
c4 – Frequent Activity Limitation	0.138 (−0.001, 0.277)	0.055 (-0.008, 0.118)

Estimates are on the probability scale for the binarized (≥14-day) outcomes; negative values indicate a lower probability of frequent distress among weatherized households. AIPW estimates use propensity scores trimmed to [0.05, 0.95]. None of the estimates are statistically significant at the 95% level due to limited sample size. Values are reproduced exactly from the analysis notebooks.

One interpretation, offered as a hypothesis rather than a settled conclusion, is that sleep and mental health may be more responsive to weatherization over this one-year horizon for two reasons. First, they are mechanistically closer to the intervention: indoor temperature directly affects sleep quality, and reduced financial and environmental stress can directly affect mood. Second, as more acute, state-like measures, they may be less subject to recall-related measurement noise over a 30-day look-back than the more chronic, trait-like physical health and functional-limitation measures. These mechanisms are plausible but remain speculative and warrant dedicated future investigation.

Limitations and Future Work

Some limitations temper these findings, which should be read as suggestive rather than confirmatory. The primary limitation is statistical power. With the available sample size, confidence intervals are wide and none of the focal estimates reaches significance at the 95%

level. Future work should explore variance reduction techniques such as incorporating additional precision-improving covariates or applying more efficient estimators to tighten these intervals and yield more conclusive, decision-useful ranges.

Second, the analysis is restricted to respondents who completed both the baseline and one-year follow-up surveys. Differential attrition between the treatment and control groups over the intervening year (due to mortality, relocation, or worsening health) could not be fully evaluated and may bias the estimates in either direction.

Finally, sleep, mental health, and the other Healthy Days measures are established predictors of long-term healthcare utilization and cost in the broader literature. The survey-based causal framework developed here could be extended in future work to approximate the long-term cost implications of weatherization-driven health improvements, complementing the direct Medicaid cost analysis presented elsewhere in this paper.

Conclusions

Following on a history of self-reported, survey-based evidence that home weatherization can improve health outcomes, the current analysis provides one of the first times that medical claims data and weatherization evaluation survey results have been combined in the U.S. to examine whether weatherization has a tangible impact on medical costs and number of healthcare encounters. When looking at total costs and claims across weatherization participants, we found that both costs and claims consistently decreased for those receiving weatherization. Notably, TennCare funding per enrollee has steadily increased from 1994 to 2024; decreases in costs per person for study participants are not likely tied to changes in funding or payment levels (Spears 2017; Mouchette & Spears 2025; Wadhwani 2024). Interrupted time series analysis reinforced that Medicaid costs and claims are increasing over time, but that weatherization may lead to a short-term reduction as well as potential stabilization of costs, though model results were mixed. In the de-biased machine learning analysis, self-reported "Healthy Days" improved for sleep and mental health but those related to physical health did not show statistical significance. This tension between the self-reported general health and the results in the claims data merits further investigation, while the improvements in sleep and mental health show encouraging signs that weatherization can fairly quickly (within a year) positively impact outcomes that are closely linked to long-term health and well-being. Due to the small sample sizes in the study, the results are considered directional and promising for future work. In an ideal scenario, a weatherization program and health insurance program would enter a data sharing agreement to link a larger number of weatherization and healthcare records for stronger sample sizes; this was done in Victoria, Australia, for example, leading to strong and defensible evidence of energy efficiency's health benefits in that region (Sustainability Victoria 2022).

References

Centers for Disease Control and Prevention. n.d. PLACES: Measure Definitions — Health Status. Atlanta, GA: U.S. Centers for Disease Control and Prevention. https://cdc.gov/places/measure-definitions/health-status.html

Chen, H.-Y., Baumgardner, D. J., and J. P. Rice. 2011. Health-Related Quality of Life Among Adults with Multiple Chronic Conditions in the United States, Behavioral Risk Factor

Surveillance System, 2007. Preventing Chronic Disease 8(1):A09. https://www.cdc.gov/pcd/issues/2011/jan/09_0234.htm

Fusi, F., and J. Lecy. 2020. *Foundations of Program Evaluation: Regression Tools for Impact Analysis*. Scottsdale, AZ: Arizona State University. https://ds4ps.org/pe4ps-textbook/docs/p-020-time-series.html

Kelman, B. (2022, August 3). They lost Medicaid when paperwork was sent to an empty field, signaling the mess to come. KFF Health News. https://kffhealthnews.org/news/article/tennessee-medicaid-tenncare-paperwork-lawsuit-disenroll-pandemic-emergency/

Marincic, M., and E. Rose. 2023. *Impacts of Weatherization on Temperature, Humidity, and Occupant Well-Being in the Tennessee Valley: Final Report*. Knoxville, TN: Three^3, Inc. www.threecubed.org/uploads/2/9/1/9/29191267/three3_tdec_ieq_report_final_11.15.23.pdf

Marincic, M., E. Rose, B. Tonn, and R. Beteta-Benson. 2025. *Effects of Home Uplift Weatherization on TennCare Utilization and Costs*. Knoxville, TN: Three^3, Inc. www.threecubed.org/uploads/2/9/1/9/29191267/tva_tenncare_final_report_10.1.25.pdf

Mouchette, B., and M. Spears. 2025. Federal safety net programs in Tennessee [Interactive infographics]. The Sycamore Institute. https://sycamoretn.org/federal-programs-dashboard/

Mykyta, L., Keisler-Starkey, K., Lindstrom, R., & Bunch, L. (2023, July 19). About 43 million people in the U.S. had multiple health plans in 2021. U.S. Census Bureau. https://www.census.gov/library/stories/2023/07/multiple-health-coverage-plans-in-2021.html

Pierson, B. (2024, August 27). Tennessee wrongly kicked thousands of Medicaid, judge rules. Reuters. https://www.reuters.com/legal/government/tennessee-wrongly-kicked-thousands-off-medicaid-judge-rules-2024-08-27/

Rose, E., Marincic, M., Tonn, B., and B. Hawkins. 2022. *TVA Home Uplift Metro Areas Final Report on Non-Energy Impacts*. Knoxville, TN: Three^3, Inc. https://www.threecubed.org/uploads/2/9/1/9/29191267/tva_home_uplift_metro_areas_nei_report_march_2022.pdf

Spears, M. 2017. Understanding Medicaid and TennCare. The Sycamore Institute. https://sycamoretn.org/what-is-tenncare/

Sustainability Victoria. (2022, August). *The Victorian Healthy Homes Program: Research findings*. Melbourne, Victoria, Australia. https://assets.sustainability.vic.gov.au/susvic/Report-Energy-Victorian-Healthy-Homes-program-research.pdf

Wadhwani, A. 2024. TennCare seeks $975M in additional funding tied to medical costs, price of new obesity drugs. The Tennessee Lookout. https://tennesseelookout.com/2024/11/18/tenncare-seeks-975m-in-additional-funding-tied-to-medical-costs-price-of-new-obesity-drugs/