1-1-2017

Analyzing Electricity Use of Low Income Weatherization Program Participants Using Propensity Score Analysis and a Hierarchical Linear Growth Model

Ksenia Polson
University of Denver

ANALYZING ELECTRICITY USE OF LOW INCOME WEATHERIZATION PROGRAM PARTICIPANTS USING PROPENSITY SCORE ANALYSIS AND A HIERARCHICAL LINEAR GROWTH MODEL

A Dissertation

Presented to

the Faculty of the Morgridge College of Education

University of Denver

In Partial Fulfillment

of the Requirements for the Degree

Doctor of Philosophy

---

by

Ksenia S. Polson

June 2017

Advisor: Dr. Antonio Olmos

Author: Ksenia S. Polson
Title: ANALYZING ELECTRICITY USE OF LOW INCOME WEATHERIZATION PROGRAM PARTICIPANTS USING PROPENSITY SCORE ANALYSIS AND A HIERARCHICAL LINEAR GROWTH MODEL
Advisor: Dr. Antonio Olmos
Degree Date: June 2017

Abstract

This evaluation utilized propensity score matching methods and a longitudinal hierarchical linear growth model to determine the effect of residential energy efficiency upgrade(s) on household electricity use for the low-income community over the course of a year in the City and County of Denver, Colorado. Propensity score analysis with risk set matching was performed at each month under analysis applying nearest neighbor and nearest neighbor with caliper approaches by balancing covariates across the treatment and control groups. Following the completion of propensity score analysis, the data were aggregated to form a data set that was used in a hierarchical linear growth model. A hierarchical linear growth model was used to examine mean differences in electricity use between the treatment groups after controlling for a set of covariates.

Results indicated that electricity consumption was best predicted with the propensity score matched subsample. Conditional growth models produced a statistically nonsignificant difference in electricity use following residential energy efficiency upgrade(s) after controlling for variables such as sex, age, primary heating fuel, square footage of household, water heater fuel type, number of household members, type of household, status of home ownership, disability status, race, unworked income, and method of payment. None of the covariates were statistically significant in predicting electricity consumption for the subsample. As a final stage of analysis, another

longitudinal hierarchical linear model was used with the entire data set, both matched and unmatched cases, to compare the results across the two data sets. The results for this model indicated a statistically significant effect of treatment with number of household members, type of dwelling, and unworked income serving as statistically significant predictors of electricity use. Since a subsample based on propensity score analysis was to simulate a randomized control trial, which is considered the gold standard in experimental research, and it is more difficult to obtain statistically significant results with a smaller sample, the results from this subsample take precedence over the results obtained from the entire sample.

This evaluation contributes to the fields of energy efficiency and evaluation practice through the application of propensity score matching algorithms to monthly longitudinal data to be able to accurately isolate the effect of treatment on the outcome. The results from the propensity score-based sample highlight the need to utilize techniques, such as propensity score matching to control for confounding variables in a quasi-experimental study. This evaluation demonstrates that in the absence of these types of selection techniques, results could be biased. Finally, the results have informed the direction of future research and focus areas at the local level for analyzing energy efficiency programs for a low-income population.

Acknowledgements

I would like to acknowledge my husband, Zach, for his continued support and understanding throughout my Ph.D. program and this dissertation. Thank you for being there for me and being patient with the whole process and challenges along the way. In addition, I would like to thank my mom, Vera and grandma, Valentina, for their consistent encouragement and positivity. Thank you to the members of my study group, you have helped me tremendously in reaching this milestone.

I thank the members of my committee Dr. Kathy Green and Dr. Kellie Keeling. It has been a pleasure to work with you and have your guidance and expertise throughout this Ph.D. program. I would like to thank members of the City and County of Denver's Evaluation team for their input, critiques, and involvement with this project. Thank you, Dr. Krystina Finlay, Dr. Ken Seeley, and Miriam Peña for your time, efforts, and perseverance in getting enough interest in this study and providing me the opportunity to see it until the end. Without your support this evaluation would not have been possible.

I would like to especially acknowledge my advisor and dissertation director, Dr. Antonio Olmos. You have significantly contributed to my success in the field of research methods and statistics. Thank you for spending so much time and effort advising and guiding me throughout the program and the dissertation process. Thank you for believing in my abilities and for giving me opportunity to partner with the City and County of Denver for this evaluation. Your genuine care for your students and research expertise are admirable and have inspired me in many ways.

Table of Contents

List of Tables

Chapter 1: Introduction and Literature Review

Introduction

With steadily rising levels of greenhouse gas (GHG) emissions globally, nationally, and locally, authorities are pressed to implement programs that promote and sustain energy efficient practices. The general public as well as public officials have a role to play in reducing energy consumption. In doing so, individuals would benefit by saving on their monthly energy costs. According to the U.S. Energy Information Administration (2013), in 2012, the U.S. residential sector accounted for 22% of the country's total primary energy consumption and about 20% of its total GHG emissions. In 2011, 71% of residential sector emissions were attributable to electricity consumption for operating appliances, lighting, heating, and cooling (US Department of State [DOS], 2014). National U.S. Laboratories, states, and cities have been tasked to perform evaluations of energy efficiency programs to identify best practices and recommendations on energy efficiency practices to consumers. An often-overlooked population for energy use reduction is low income households within the residential sector (Murray & Mills, 2014). Low income households are particularly vulnerable to energy costs and programs aimed to assist this population are especially important to ensure their energy security. While there was a comprehensive national evaluation of the Weatherization Assistance Program (WAP) by the Department of Energy and associates, evaluations at the city and

county levels are rare. This evaluation is important because it is one of the first evaluations conducted at the local city and county level and provides useful outcomes and recommendations to a local government agency.

To date, the Low Income Home Energy Assistance Program (LIHEAP) is the only federal program that has been designed to assist the poor with household energy costs. LIHEAP is a block program that was established through Title XXVI of the Omnibus Budget Reconciliation Act of 1981 (Kaiser & Pulsipher, 2003). This program is administered by the U.S. Department of Health and Human Services (HHS) which uses funds to assist eligible households in paying a portion of their energy costs. Another federal program, with similar eligibility criteria, the Weatherization Assistance Program (WAP) is administered by the U.S. Department of Energy (DOE). The WAP assists low-income consumers including the elderly, persons with disabilities, and children by weatherizing households and implementing health and safety precautions (Tonn *et al.*, 2003). Since this study was community - focused, Xcel-sponsored Low - Income Weatherization Program (LIWP) was examined in the analysis. LIWP is a free weatherization service for low-income customers aimed to reduce participants' energy bills. Xcel Energy partners with Energy Outreach Colorado, a local non-profit whose goal is to lower energy bills through customized weatherization services. This study focused on LIHEAP participants' evaluation of energy efficiency upgrade(s) and its impact on electricity consumption in kilowatt hours (kWh) for low - income households.

Statement of the Problem

Most of the increase in GHG emissions in the atmosphere over the course of the last 150 years have been the result of human activities. In the U.S., the main source of GHG emissions is from burning fossil fuels for electricity, heat, and transportation needs. GHGs function by trapping heat and making the planet unseasonably warm. In the growing world population, GHG emissions are widespread and cause climate change (DOS, 2014). There is a need to study programs that improve energy efficiency in residential households, particularly in low - income households, to determine whether energy efficiency practices result in energy savings and, thus, reduce air pollutants. Weatherization programs, and to a larger extent LIHEAP, have received little attention from political scientists and social scientists regarding their effectiveness to address the needs of low-income consumers. The implications from these programs and their impact for public policy regarding energy cost support to low-income households have not been thoroughly discussed. One-third of all workers in the U.S. earn wages that place their income below the poverty level (Carnevale & Rose, 2001). This portion of the U.S. population is especially in need of social programs to support its well-being and to avoid poverty. There is a scarcity of studies that focus on the effects of retrofit interventions on energy use in low-income residential households. In this study, the segment of the population who were enrolled in LIHEAP and LIWP were evaluated to determine whether (if at all) energy efficiency upgrades impact electricity use of low - income households. Households are eligible to participate in both programs, and for this evaluation all households were enrolled in both LIHEAP and LIWP. This evaluation

focused on the impact of LIWP energy efficiency upgrades on low - income households' electricity use.

Purpose of the Study

It is imperative to define explicit goals for the evaluation of a program at the start (Fitzpatrick *et al*., 1997). The primary purpose of this evaluation was to determine whether there was a statistically significant change in mean electricity consumption in kWh over the course of a year for low - income households enrolled in Xcel's LIWP upgrade(s) in comparison to participants who had not installed the upgrade(s). This evaluation also examined whether a set of covariates related to the outcome had an impact on households' electricity use. Finally, the evaluation extended the application of propensity score matching analysis and a hierarchical linear growth model in the field of energy efficiency.

The objectives of this evaluation were three - fold:

(1) Use propensity score matching and a hierarchical linear model to determine whether mean electricity consumption statistically significantly differs between households who had not installed upgrade(s) in comparison to households who had installed the upgrade(s).

(2) Assess whether household-level covariates statistically significantly impact electricity consumption of LIWP participants.

(3) Compare the results of propensity score matching and hierarchical linear growth model to analyze longitudinal monthly electricity data and household

level covariates over a period of twelve months with the results obtained from performing a hierarchical linear growth model without propensity score matching.

This evaluation applied quantitative statistical methods to draw inferences about the effectiveness of LIWP to contribute to a change in electricity consumption for low - income consumers with 12 months of data. In the initial phase, propensity score matching was performed to create a subsample of low - income households that were similar on a set of covariates. In the second phase, the data set compiled from propensity score matching was used to perform hierarchical growth modeling to assess if electricity use changed between households that had installed energy efficiency upgrades in comparison to households that had not yet installed them. Then, a new hierarchical growth model was fit with the entire data set, matched and unmatched cases, to determine if households' electricity use changed as compared to results obtained from the previous propensity score-based model.

Evaluation Questions

The following evaluation questions were used to guide the researcher throughout this study.

1) Do low-income residential households experience a statistically significant change in electricity use over the course of a year following participation in LIWP upgrade(s) in Denver County, CO by implementing propensity score matching and a hierarchical linear growth model after controlling for covariates, such as

sex, age, primary heating fuel, square footage of household, water heater fuel type, number of household members, type of household, status of household, presence of a disability, race, unworked income, and method of payment?

2) How do the results using a subsample of households matched on propensity scores and using the entire data set (without propensity score matching) compare in terms of a statistically significant change in electricity consumption over the course of a year using a hierarchical linear growth model after controlling for covariates, such as sex, age, primary heating fuel, square footage of household, water heater fuel type, number of household members, type of household, status of household, presence of a disability, race, unworked income, and method of payment?

Hypotheses

The hypotheses formulated for this study are as follows:

1) H_1 = Households that participate in LIWP experience a statistically significant decline in their electricity use in kWh following the installation of upgrade(s) after controlling for the covariates listed above.
2) H_2 = A subsample of households that are matched on a propensity score experience a greater statistically significant decline in electricity consumption as compared to the entire data set following the installation of upgrade(s) after controlling for the above covariates.

The first evaluation question was addressed with the application of propensity score matching analyses including nearest neighbor with a ratio of 1:1 and nearest neighbor with a caliper value of 0.1 and a hierarchical linear growth model. The second evaluation question was explored with an implementation of a hierarchical growth model on the propensity score- based data set that allowed for specification of time, intercept, and growth parameters. Prior to fitting the hierarchical growth model, exploratory examination of data was conducted to establish support for this model and provide visual graphs of the outcome across households that installed upgrades and those that did not. Then, a hierarchical growth model was applied on the entire data set to compare the results of the model to that developed in the previous evaluation question. Table 1 provides the reader with definitions of terms used throughout the study.

Table 1

Definition of Terms Relevant to the Study

Term	Definition
Balanced Risk Set Matching	An entity receives treatment at time *t* and that entity is matched to another entity with a similar set of covariates up to time *t* that has not received the treatment up to time *t*. Marginal distributions of covariates are forced to be balanced across the matched treatment and control groups (Li, Propert & Rosenbaum, 2001).
Energy Efficiency	A way of managing and restraining the growth in energy consumption.

Feedback Effect	It is measured in terms of a change in electricity consumption, in kilowatt-hours (kWh), based on electricity meter information before and after the introduction of an intervention program (U.S. Department of Energy Office of Project Management Oversight & Assessments, 2016)
Propensity Score Analysis	It offers a useful approach to the analysis of evaluation data when randomized trials are not possible or in times when researchers need to measure treatment effects due to intervention(s). This procedure assigns participants to treatment and control group and matches them on a set of key covariates which forms a propensity scores based on distance between the covariates (Guo & Fraser, 2015).

This dissertation is organized as follows. The literature review that follows this introduction includes a brief overview of common examples of energy efficiency practices that have been implemented. It highlights previous research on WAP and LIHEAP and their impact on consumer energy use. Chapter 1 concludes with the theoretical framework of propensity score analysis and hierarchical linear growth modeling and provides background for the method section in Chapter 2. Chapter 2 presents the method used to address the evaluation questions and hypotheses. Chapter 3 reports the descriptive statistics, results of propensity score matching algorithms, and observations from hierarchical linear growth models. Chapter 4 concludes with main findings examined through this research study, discussion, strengths and limitations of the study, and areas of further research. This chapter, also, mentions the value of this dissertation to the field of research methods and statistics as well as evaluation practice in energy efficiency.

Literature Review

A literature review provides a background for a research study. The American Psychological Association (APA) defines literature reviews as "critical evaluations of material that has already been published" (2016, p. 10). It combines previous literature reviews, draws comparisons to the current study, and establishes the importance and significance of the current study.

Worldwide, electricity use accounts for approximately 40% of GHG emissions. Effective conservation programs have been implemented to reduce GHG emissions and fossil fuel pollutants (Delmas *et al*., 2013). Studies that examine the effects of energy efficiency strategies on energy consumption have been conducted since the 1970s. Interventions can be divided into two categories: antecedent and consequence techniques (Abrahamse *et al*., 2005). Antecedent interventions are those that implement methods that influence energy reduction prior to embarking on environmental practices at home. These methods have been proven effective because they use personalized information that is relevant to a household participating in the environmental intervention. For example, households that receive individualized information on energy saving practices or home energy audits tailored to the needs of a home are more likely to reduce their energy use. However, providing information on its own is not sufficient and other forms of outreach are necessary (Geller, 1981).

Consequence interventions are actions taken *following* the energy efficiency activities by means of providing a consequence which is dependent on the end result. A common energy efficiency strategy, also a consequence intervention, is called a feedback

effect. Often, it is defined as "any procedure in which subjects are taught to discern their own behavior through contingent stimulation" (Hayes & Cone, 1981, p. 81). In this procedure, residents are taught to determine the cost, amount, and use of energy to meet their energy savings needs. The idea is that by providing households feedback on their energy savings facilitates further reductions in energy. Hayes and Cone (1981) observed that households that received monthly feedback reduced electricity use on average by 4.7%, while the control households increased their use by 2.3%. However, studies citing comparative feedback among neighbors did not show better results than individual feedback. This is likely due to participants having individual electrical and natural gas consumption patterns that cannot be meaningfully compared to their neighbors. Overall, the results have been mixed, with some studies noting positive effects of energy feedback strategies causing a decrease in consumption levels and others reporting minimal significant treatment effects.

For example, research by Grønhøj and Thøgersen (2011) found that families who participated in an intervention program that focused on energy feedback effects achieved an 8.1% reduction in electricity usage against a 0.7% reduction by a comparable control group. The study concluded with a model for use of feedback systems that allows energy consumers to connect their everyday actions to their energy consumption. This feedback tool produced particularly positive effects in terms of energy savings in families with teenage children. On the other hand, an energy feedback program in the U.K. found only marginal statistical significance in changes in residential behaviors (Brandon & Lewis, 1999). This study concluded that industrialized societies, such as the U.K., have

favorable intentions towards conserving the environment, but these intentions are not always translated into appropriate behaviors. This is due to several factors that include a lack of opportunities to save, financial costs, and comfort in a chilly climate.

There are many disparities that arise from studies that attempt to assess energy efficiency of households because they differ in methodological quality and experiment with different efficiency strategies. Nevertheless, there is an awareness of the detrimental impacts of GHGs and the benefits of implementing building retrofits to achieve energy savings (Metz *et al*., 2007). However, Altan (2010) reports that there is still a general lack of rigorous methodologies for viable user interventions that consider the interconnections of social and behavioral factors that affect energy consumption. It has been extensively reported that influencing household energy consumption is challenging. This is due to the fact that the level of consumption is heavily dependent on household behaviors (Kua & Wong, 2012). For example, Kua and Wong examined the effectiveness of combining three types of instruments of outreach including pamphlets, stickers, and face-to-face interactions to influence household energy conservation. The results indicated that self-reported behavioral changes were strongly correlated to the degree of trust placed in the energy conservation information provided, the ease with which conservation measures could be implemented, and satisfaction associated with implementing these measures. When the actual reductions in energy consumption were examined, 60.7% of participants in the treatment group observed an average reduction of only 2% in energy use. The discrepancy between self-reported behavior and actual savings was attributed to interactions with households. Households comprising more

than five residents, households having an elderly member above 60 years old or children aged below 12 years old were less consistent in following the recommended measures. So, any conservation achieved by the rest of the household members was set back by family members in these age groups. As such, success or failure of an energy intervention program is often determined by the household's characteristics and occupant consumption patterns.

The barriers to participating in household energy efficiency retrofit programs are dependent on socio-economic, financial, and personal factors. For example, lack of knowledge about the extent of energy savings that could be achieved due to retrofits and relatively low energy prices. The financial piece is tied to education as consumers are concerned with costs and financing that is required to perform equipment upgrades and renovations. Personal factors play a role since a decision to upgrade a home is influenced by one's aptitudes, attitudes, and availability of information on local contractors and vendors of materials.

Due to these factors, many programs that aim to encourage households to perform energy improvements to building stock often include incentives to home owners. These incentives can be in the form of providing education on home energy evaluation programs to financial incentives to perform the upgrades. However, these types of building retrofits require time, effort, and finances from both the household and entities performing the activities. Stern (1992) reports that program participation rates vary according to the type and size of financial reward, with larger financial incentives resulting in greater program participation. As such, there can be many barriers for

consumers to adopt energy saving retrofits unless government-administered and government-sponsored programs are implemented.

Overview of Weatherization Assistance Program

Few studies have examined the costs of energy use for consumers and potential savings that could be obtained by weatherizing their households. A rise in fuel prices in the mid to late 1970s prompted U.S. government response with the creation of WAP. In 1979, WAP was authorized by Congress under Title IV of the Energy Conservation and Production Act and is now funded by the DOE. The purpose of the program, as stated in the Weatherization Assistance for Low-Income Persons Rule (2001), is to "…increase the energy efficiency of dwellings owned or occupied by low-income persons, reduce their total residential energy expenditures, and improve their health and safety…" This program serves grantees, i.e., 50 states and the District of Columbia, some Indian tribes, and territories, to increase energy efficiency of households. The grantees transfer grants to their sub-grantees, such as local weatherization agencies to perform the work. Historically, households are eligible for WAP if they meet the following criteria: income at 150% of the federal poverty level or income 60% or less of the state median income (Tonn *et al*., 2003). In the past few years, the guidelines were revised and define eligibility as household income at or below 200% of the federal poverty level (Eisenberg, 2014).

Due to its nature, the program can produce several positive economic effects. It allows for a demand in labor and materials to implement the program. The program also

results in a decrease in consumption of primary home heating or cooling fuels, such as electricity, natural gas, and oil. In theory, the program contributes to an increase in household income by an amount that would have been used on household's energy consumption had there not been any measures installed. To date, the program has contributed to upgrades of more than 6.9 million homes. Energy efficiency retrofits save households an average of $437 annually in heating and cooling costs with additional savings from lighting and appliance upgrades. For every $1 invested in WAP, it returns approximately $2.51 in benefits to consumers (State of Rhode Island Energy Efficiency and Resource Management Council, 2016).

In 2009, a stream of funding for the WAP was increased based on the federal stimulus resources being provided through the American Recovery and Reinvestment Act (ARRA) of 2009. DOE allocated a total of $227.2 million per towards WAP grants to 51 grantees which included the 50 states and the District of Columbia. This study was an impact evaluation of program outcomes using data from researchers at Oak Ridge National Laboratory (ORNL) who were joined by a team of independent evaluators (APPRISE, Dalhoff & Associates, Blasnik & Associates, Energy Center of Wisconsin) to perform the evaluation (Tonn *et al*., 2014). A review of those studies is presented in the following section.

The primary method of analysis to estimate energy impact as a result of WAP consists of pre/post treatment research designs with a comparison group. The analyses used weather normalized utility billing data. Tonn *et al*. (2014) explained that the weather normalization approach led to the estimation of "…weather-adjusted annual energy

consumption for each home based on monthly usage data and daily outdoor temperature using a variable degree day base regression analysis" (p. 25). Total energy savings per household was calculated as the difference in the normalized annual consumption between the pre-treatment and post-treatment periods.

For this evaluation for the 2008 program year, nearly 98,000 units were weatherized, of which 59% were single family homes, 18% mobile homes, 5% small multifamily homes, and 18% large multifamily buildings (Tonn *et al*., 2014). The primary fuel source for 60% of the weatherized single family homes was natural gas, followed by 26% bulk fuels, and 14% electricity. Over 50% of upgraded homes were built before 1980. The evaluation reported that, on average, the clients of WAP were more likely to be elderly, have a disabled person living in the home, have a child less than 5 years of age, be a single parent household, and be less healthy as compared to the general U.S. population. One of the primary goals of the study was to estimate energy cost savings to consumers based on weatherization of homes. On average, energy cost savings were 11.9% for single family homes or about $264 per year. Since heating fuels, such as fuel oil and propane are higher priced than natural gas and electricity, energy cost savings from these fuels are higher. The results from a nationally representative random sample of weatherized households and a control group support that weatherization can be effective in decreasing energy burdens of low-income homes. Post-weatherization, more participants reported a variety of positive effects including an easier time paying for utility bills, were less likely to have their electricity or natural gas disconnected, and had enough funds to both purchase food and pay energy bills monthly. However, even

following weatherization most households still experienced an energy burden, according to Tonn *et al*. (2014).

Findings from a national evaluation of WAP indicate that it is a cost-effective federal investment. A summary of the studies based on the listed evaluation data for the 2008 program year is provided in Table 2.

Table 2

WAP Evaluation Data and Results Summary for 2008 (Tonn *et al*., 2014)

Evaluation Data	Impact Evaluation Outputs	Outcomes
Housing characteristics and weatherization measures installed in about 20,000 households and mobile homes	Approximately 35 million households were eligible for WAP in 2008	Estimated first year program energy savings were 2,270,000 MM Btus, equivalent to 400,000 barrels of oil
Building characteristics and weatherization measures installed in 10,000 multi - family units	WAP funds supported upgrades of 97,965 units with 59% single family, 18% mobile homes, and 23% multi-family	Large variations in energy savings were more influenced by occupant behavior and changes in primary heating fuel and use of secondary fuel sources than by the quality of upgrades
Fuel type and occupant characteristics for about 20,000 households	DOE total expenditures on WAP were $236,000,000	The net present value of the program energy cost savings was $420,000,000 (2013 dollars)
Electricity and natural gas billing histories for about 8,000 weatherized homes and homes from 1000 natural gas and electric utilities that were used as a comparison group	The average cost to weatherize a unit was $4,695	78% of the savings accrued to households and 22% to rate payers of utilities
Program implementation survey data from 50+ grantees and about 900 sub-grantees	WAP and leveraged expenditures supported 8,500 jobs and increased national economic output by $1.2 billion	Carbon emissions were reduced by 2,246,000 metric tons (about the amount of carbon emitted from 600,000 automobiles in the U.S.)

Demographic, energy use behavior, and health impacts data for about 1400 households		The surveyed households reported that after weatherization their homes were better insulated; general health of occupants improved; respondents suffered fewer asthma symptoms; experienced less pesticides; fewer instances of thermal stress; and fewer missed days of work
Demographic and employment-related data from 600 energy auditors, crew leaders, and crew		
Indoor environmental quality data measures prior and following weatherization for a nationally-representative sample of 500 treatment and control homes		
Detailed in-field observations of 450 audits, installation processes, and final inspection reports by 19 sub-grantees		
In-field assessments of 105 upgraded households from 2008 and energy saved (if any)		
14 case studies of high-performing weatherization agencies		
Training experiences from a survey of 800 individuals who received training at DOE weatherization centers		

Even though the national evaluation found the program was effective for program participants, local agencies vary substantially in their management strategies and the range of energy savings differs across households. Research by Brown and Berry (1995) shows that, "nearly three quarters of the variation in energy savings across agencies could be explained by the types of weatherization measures installed and the average level of

gas consumption per dwelling prior to weatherization" (p. 742). In particular, these researchers reported the greatest energy savings from energy users who have weatherized their homes and whose household's envelope or heating systems were compromised. This study documented that energy efficiency measures, such as installing attic, wall, and floor insulation were especially cost-effective. In addition, installation of water heater measures resulted in a decrease of base-load consumption of natural gas. Households that replaced furnaces were more likely to achieve higher-than-average savings and new furnaces acted to promote health and safety. Finally, distribution system problems, such as air leakage, when left unfixed, posed health and safety concerns and contributed in a reduction of energy savings. This study indicated that weatherization of residential households typically results in reduced energy consumption and substantial energy savings to their occupants.

Overview of Low Income Home Energy Assistance Program

LIHEAP, closely tied to the WAP based on eligibility criteria and as a source of weatherization funding, warrants further investigation. The impacts of LIHEAP on low-income home energy savings have not been thoroughly vetted in the literature. A precursor to LIHEAP was the Low - Income Energy Assistance Program (LIEAP) which was established in 1974 through the Emergency Energy Conservation program, a part of the Economic Opportunity and Community Partnership Act (Warren, 2003). At the time, the program's aim was to assist households with their weatherization needs. By 1977, the Community Services Administration continued to provide assistance as the needs of low -

income consumers grew during a hike in fuel energy costs. In 1981, LIEAP transitioned into LIHEAP.

Eligibility for LIHEAP is based on household income. The criterion that the households must meet to qualify for LIHEAP include income at 150% of the federal poverty level or income 60% or less of the state median income (Tonn *et al*., 2003). However, states may choose to set the limit above 110% of the poverty level. States exercise discretion regarding the amount of financial assistance they provide to households. LIHEAP offers three different levels of assistance that include paying a portion of a household's heating and cooling costs, providing financial assistance in times of crisis, and providing weatherization services. The U.S. Department of Health & Human Services (HHS) reports that during fiscal year 2007, the residential energy burden for low-income households was 13.5%, 16% for LIHEAP recipients and 7% for all households (HHS, 2008). This indicates that low-income vulnerable populations spend almost twice as much of their income on home energy costs as households in general.

Research by Murray and Mills (2014) states that participating in LIHEAP significantly reduces households' energy insecurity. More importantly, reductions in LIHEAP have worsening ramifications for low-income households and utility companies. This study showed that eliminating LIHEAP significantly decreases the number of energy secure households by 17%. This indicates that more energy insecure low-income households could default on their payments to the utilities, thereby causing loss of profits, and higher rates for paying customers. However, these findings do not necessarily lead

one to advocate for more funds for this program as the cost - benefit trade - off needs to be further analyzed to provide concrete recommendations.

Intersection of WAP and LIHEAP

Another study, the first of its kind in 2003, assessed the relationships between WAP and LIHEAP. This study's aim was to determine if there is an impact from weatherizing homes of LIHEAP participants based on the level of assistance they receive. The study implemented a research design that included pre- and post- treatment conditions with random assignment and a control group to assess a sample of households that only received LIHEAP assistance compared to households that received both LIHEAP and weatherization assistance (Tonn *et al*., 2003). The level of LIHEAP benefits provided was determined based on household income category; benefits increased as household incomes declined and household size increased. In addition, a high - energy benefit was allocated to households that met an energy burden based on a calculation of primary heating fuel expenditures (Tonn *et al*., 2003). The study concluded that weatherization of houses resulted in energy savings for participants, with decreases in LIHEAP high - energy benefits. However, the findings also indicated that weatherization did not suggest that the need for standard LIHEAP benefits be relinquished, as there are other factors that contribute to household energy insecurity. To improve knowledge in the field of energy efficiency and inform future efforts to facilitate energy savings, researchers need to implement methodologically rigorous designs and have access to resources that would enable them to conduct studies. Some of the crucial

resources that need to be available to researchers include provision of funding, availability and access to data, and processes that ease in navigating data privacy issues. There is a scarcity of studies that document the impact of LIHEAP and WAP on household energy consumption.

Analytical Framework

Propensity Score Analysis and Matching.

An evaluation is defined as the systematic investigation of the merit, worth, or significance of any "object" (Tyler *et al*., 1967). It is an assessment of the implementation and effects of a program. In theory, the goal of an effective evaluation is to achieve program improvement and benefit the recipients of a program's funds. Evaluation assesses the utility of specific components of a program in detail to inform future decision - making about a program. This research is focused on outcome evaluation and monitoring. From a program evaluation perspective, the goal was to determine whether the set of upgrades performed through the LIWP resulted in electricity use reductions thus giving merit to the program in achieving its intended purpose.

The purpose of many studies is to explore relationships between variables. The exploration can take an experimental form in which an active independent variable (IV) is manipulated or a study can examine individual differences with an attribute IV. The attribute IV is typically a naturally occurring phenomenon. The experimental form can be a randomized control trial or a quasi-experimental research design. The randomized control trial design has the greatest ability to discern causality while a quasi-experimental

design requires close inspection of threats to internal validity (Johnson & Christensen, 2012). The two designs are similar due to the administration of a treatment. If a design does not have a treatment it is considered a non - experimental design. If in addition to treatment, a design has a control group or multiple measures, but the individuals are not randomly assigned to groups, then it is considered a quasi - experimental design. The randomized control trial research design is the 'gold standard' and causality is supported because many variables are under control by the researcher and groups are considered probabilistically equivalent on uncontrolled variables due to the randomization.

Causation indicates that one event is the result of the occurrence of the other event. This is also referred to as cause and effect relationship between the two variables (Shadish, *et al*., 2002). In an experimental design, making causal inferences is dependent on a comparison between a treatment group and a control, or counterfactual. A worldview paradigm that encompasses a research agenda based on the cause and effect relationship is the positivist tradition. Positivist theorists claim that for causation to be supported, three criteria must be met: 1) temporal precedence of cause, 2) relationship between cause and effect, and 3) no other variable can explain the association between the independent and dependent variable through control or isolation (Cook & Campbell, 1979). It is only when these criteria are met and (the third criterion is more difficult to meet than the other two) that we can provide a cause and effect explanation. The first criterion is logical as it establishes that the cause precedes the effect chronologically.

The second criterion that mandates an association between cause and effect requires further inquiry. An association between cause and effect does not imply

causation. A major element in establishing causation is control. In an experimental study, ideally, many variables are ruled out leading to a study in which the effect of treatment can be isolated. The researcher has control over the setting of a study, control over who receives treatment and when this treatment is administered, and controls threats to valid inference (Johnson & Christensen, 2012). Essentially, the researcher attempts to learn whether the treatment that was applied caused the effect that was observed. For example, a correlation or a relationship between variable A and variable B could be due to the influence of variable C on variables A and B, or it could be that A causes B and *vice versa*. However, ruling out alternative explanations makes a stronger case for a cause - and - effect relationship. This criterion is usually best met with a randomized-control trial.

In true experiments, selection, also known as method of sampling, is controlled. However, these factors may imply that obtained effects are specific to the population under interest and do not apply to the populations to which one wishes to generalize findings (Campbell, 1957). In quasi - experimental studies, selection bias can exert a pivotal effect on the outcome. Since assignment to groups is not random, selection bias is associated with systematic differences between the treatment and control groups. The threats to internal validity are central to the drawbacks of quasi - experimental studies because they relate to the central question of whether treatment made a difference. Threats, such as maturation, selection, mortality, diffusion, or imitation of treatments, among others, relate to assignment to the treatment and control groups and help in identifying other explanations for observed differences on the outcome. For example,

Hawthorne effects are present in subjects who behave differently because they are being studied (Allcott & Mullainathan, 2012). In summary, random assignment remains the gold standard because the probability of individuals being assigned to the control or the treatment group is equal, and alternative explanations of treatment effect are eliminated more easily.

However, often in the social and physical sciences, randomized - control designs are not possible and can be costly and sometimes unethical. As such, quasi - experimental designs are used as the next best alternative to randomized-control designs. Propensity score methods have proven useful for evaluating treatment effects with quasi - experimental designs or observational studies. The role of propensity score methods in observational studies is to reduce the bias created by nonrandom assignment and making the adjusted estimates closer to those from randomized experiments. There are six important steps involved in the design of observational studies, as shown in (Figure 1).

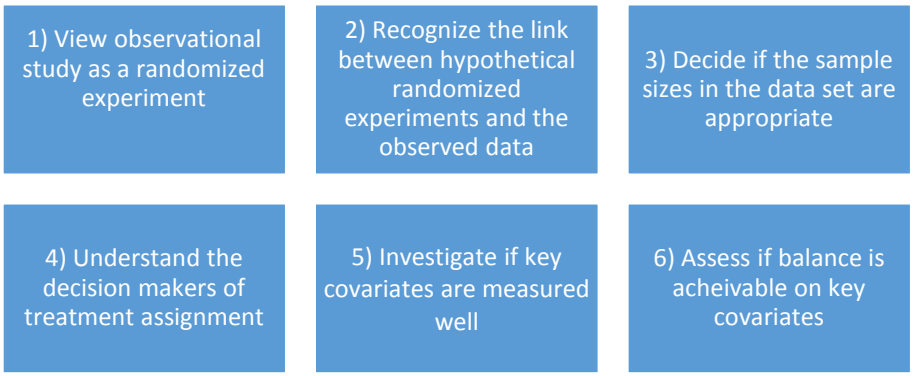


Figure 1. *Six Steps to Designing Observational Studies* (Rubin, 2008)

Having described the requirements necessary to assert causality in observational studies, reviewed here are each of the steps listed above in the context of conducting an observational study. The first step is to view an observational study as a randomized experiment. Many of the components of randomized designs can be duplicated when designing observational studies whose purpose is to obtain the closest possible answer that would have been obtained in a randomized experimental design comparing the same treatment and control conditions in the same population. The impetus underlying a solid research design is the ability to draw causal inferences based on the data. A conceptual framework developed to facilitate causal inferences is the Neyman - Rubin counterfactual framework. This framework explores whether assumptions used in random assignment experiments apply to observational studies. In this context, a counterfactual is defined as "a potential outcome, or the state of affairs that would have happened in the absence of the cause" (Shadish, Cook & Campbell, 2002). That is, for a participant in the treatment group, a counterfactual is the potential outcome under the control condition and *vice*

versa for a participant in a control group. Thus, the counterfactual is not observed in real data, it is a missing value. Neyman - Rubin's framework states that "individuals selected into either treatment or non-treatment groups have potential outcomes in both states: the one in which they are observed and the one in which they are unobserved" (Guo & Fraser, 2015, p. 24).

The Neyman - Rubin framework can be summarized in the following model:

$$Y_i = W_i Y_{1i} + (1 - W_i) Y_{0i} \quad (1)$$

Where i represents each person under analysis who would have two potential outcomes (Y_{0i} and Y_{1i}) that indicate potential outcomes in the untreated and treated states. $W_i = 1$ denotes the receipt of treatment, $W_i = 0$ denoted non-receipt. Y_i represents the measured outcome variable (Guo & Fraser, 2015).

The rationale for this equation serves as support for drawing causal inferences between W_i (the cause) and Y_i (the outcome) in which one investigates the outcome of Y_{0i} under the condition $W_i = 0$ and compares Y_{0i} with Y_{1i} (Guo & Fraser, 2015). This is a simplified explanation of the counterfactual-based framework and there are many confounding factors that may impact the outcome.

According to Rubin (2008) the second step in conducting a quasi - experiment involves understanding by the researcher regarding the nature of the hypothetical randomized experiment that led to the observed dataset. The key is knowing exactly what the treatment conditions and outcome variables were. This step ensures that the researcher is fully aware of the experiment that is being approximated by the data. The third step is to determine whether the sample sizes of the dataset are adequate. This step

is met by performing power calculations, taking into consideration the level of power desired, level of significance desired, and the sample size. When conducting a randomized control trial, it is not possible for the researcher to look at any outcome measures since the experiment has not been carried out. Rubin (2008) believes that by removing any outcome measurements from the dataset, the researcher introduces an important feature of randomized experiments to observational studies. This step is done to provide objectivity to the study at hand until the design phase of the study is complete.

The fourth step is to understand why some subjects received the active treatment condition versus the control treatment condition. For example, identifying the background variables measured on the experimental subjects that led the decision - makers to assign subjects to one group over another determines the key covariates for the study. It is important for the researcher to understand what rules were used in assigning the treatment condition to individuals. The fifth step concerns the availability and quality of key covariates' measurements. If the covariates are poorly measured or even non-existent, any analysis conducted is futile. Unless there is sufficient scientific evidence that supports using a certain set of covariates, having completed this step provides support for validity to the analysis.

The sixth step concerns the extent to which balance can be achieved on key covariates. In this step, the goal is to find matched pairs of treated and control units such that the treated and the control units appear to be balanced on their distributions of key covariates. In some scenarios achieving this balance could be difficult and inferences may be restricted to a subpopulation of units. Having met the six steps described does

not fully mean that the researcher will attain an answer similar to the one that would have been attained in a randomized experiment, but it, at least, makes the study at hand more objective and approximates randomized conditions in which the probabilities of treatment versus control assignment vary little across the matching pairs.

In addition to the steps described above, critical assumptions of causality must be met. One of them is the ignorable treatment assignment assumption (ITAA). Since the counterfactual acts as a missing value in evaluations with observational data, there are several sources of error that contribute to bias on outcome difference. For the Neyman - Rubin counterfactual framework to work correctly, the ITAA must be met. ITAA refers to assignment to treatment or control conditions that is independent of the potential outcome, if we hold observable covariates constant.

This assumption can be expressed as:

$$(Y_0, Y_1) \perp W \mid \boldsymbol{X} \qquad (2)$$

Where conditional on covariates $\boldsymbol{X}$, the assignment of evaluation participants to binary treatment conditions (i.e. treatment vs. non-treatment) is independent of the outcome of non-treatment (Y_0) and the outcome of treatment (Y_1) (Guo & Fraser, 2015).

It is typical for the ITAA to be violated in quasi-experimental conditions because the creation of a comparison group follows a process that introduces endogeneity bias to group assignment where the outcomes are not independent of treatment. Usually, the ITAA is tested before the treatment is implemented, by a chi-square test in the case where *X* is a categorical variable or an independent samples t - test if *X* is a continuous variable. If significant differences are detected, ITAA is not upheld and there is evidence for

endogeneity bias. This is an indication that action is warranted in the form of alternative approaches that correct for the endogeneity bias. To address the issue of selection bias and estimates of Average Treatment Effects (ATEs), propensity score methods have been applied to data in observational studies (Stürmer *et al*., 2006). A propensity score is defined as the conditional probability of assigning a unit to a particular treatment condition given observed covariates (Rosenbaum & Rubin, 1983). In other words, the propensity score is a balancing score representing a combination of observed covariates. So, a pair of treated and control participants who share a similar propensity score are seen as equivalent, even though their values may differ on individual covariates.

Propensity Score Matching Procedure

Propensity score matching can be expressed as a three-step process as described in Guo and Fraser (2015). More detail on the process of conducting a propensity score analysis is provided in the method section.

1. Selection of key covariates and estimation of propensity scores. The model starts with an estimation of the conditional probability of receiving treatment. This procedure is carried out with logistic regression that analyzes a dichotomous treatment variable and covariates that are perceived to be causing an imbalance between the treated and control groups. The goal of this procedure is to decide on covariates that contribute to selection bias and arrive at an optimal estimate of propensity scores.

2. Matching with propensity scores. Once propensity scores are created they are used to match participants in the treatment group with participants in the control group. Matching based on multiple covariates is difficult and may result in a dimensionality problem. Matching based on a single propensity score addresses this concern. The goal of matching is to make participants across the groups as similar as possible on the propensity scores. Matching techniques used include greedy matching. A type of greedy matching, nearest neighbor with a ratio of 1:1 and nearest neighbor with a caliper of 0.1 was utilized in this study.
3. Analysis using propensity scores. a) In theory, the sample produced in step 2 corrects for selection bias on observed covariates and violations of statistical assumptions, such as independence between the IV and the regression equation's error term. At this phase, multivariate analyses can be performed with the sample as if it had been a part of a randomized experiment. A caveat to this is that most multivariate analyses can only be performed with samples created by greedy matching.

   b) An alternative to performing multivariate analyses is using stratification of the propensity scores. Stratification is a comparison of the mean difference of an outcome between treatment and control conditions within a stratum. The goal of stratification is to estimate a mean and variance for the sample. The ATE and its statistical significance are also estimated.

Balanced Risk Set Matching

Li, Propert, and Rosenbaum (2001) were the first researchers to develop an approach to propensity score analysis in which an individual receives treatment at time *t*, and that individual is matched to another individual on a similar set of covariates up to time *t* who has not received treatment up to time *t*. Through this approach, marginal distributions of covariates are forced to be balanced in the matched treated and control groups. Among all balanced matches, a propensity score is selected that minimizes the distance between covariates within matched pairs. When the treatment and some pre-treatment covariates are time dependent, balanced risk matching is the recommended technique. In this evaluation time was a critical component of the study, thus, participants in the control group became participants in the treatment group after the installation of upgrade(s), balanced risk set matching was applied to the treated and control groups.

Li (1999) developed two types of risk set matching. The first type involves risk set matching with untreated controls, so only matches that were never treated could serve as controls. In this case, all matched controls are selected from a pool of untreated patients. The second type of risk set matching is performed when an individual treated at time *t* can be matched to any individual not yet treated prior and up to *t*. Thus, the individuals are considered controls if they have not received treatment yet. A crucial point in risk set matching is that each individual is either treated once or never, assuming that treatment time varies among individuals. As such, the time of treatment for different individuals may vary and the treatment is time dependent.

In the field of energy efficiency, especially for low - income households, the time that the upgrades are installed varies across households as they may request upgrades at any time if they qualify for the services. So, the data on energy efficiency is longitudinal and collected in monthly increments. As such, individuals who were not treated at a time point served as controls and once they became treated they moved into the treatment group.

Types of Propensity Score Matching

The following section describes the advantages and disadvantages of greedy matching which was used to match participants in the control group with participants in the treatment group based on the computed propensity score. Greedy matching is a linear matching algorithm that matches participant in the treatment group with a first case in the control group based on the criteria for matching (Rosenbaum, 2002). Greedy matching can be disadvantageous because when a match between a treatment and control is created, the participant in the control group is removed from any further consideration for matching. For this study, nearest neighbor 1:1 and nearest neighbor with caliper 0.1 were recommended given a large sample size of low-income households and the possibility to conduct follow-up analyses.

Greedy matching.

Greedy matching refers to a procedure in which a match for a participant in the treatment group is based on the first case of the control group that meets the criteria for

matching (Olmos & Govindasamy, 2015). It is useful to note that even if there is a more optimal match in the control group, the algorithm will still choose a match based on the first case. Most algorithms select participants from both the control and treatment groups at random, so multiple runs will result in different groups with varying degrees of matching. All greedy matching algorithms divide matching into smaller, simpler decisions which are handled optimally one at a time. In other words, once the algorithms find a match the decision is final and reconsiderations are not possible. This approach may or may not find the best match for a participant. With all greedy matching, researchers encounter a problem of a common support region that needs to be set for it to work. The common support region refers to a set of propensity scores in terms of logits across the treated and non-treated participants. As such, participants who fall outside of the common support region have no matches and are excluded. The common support region is sensitive to the covariates used to predict propensity scores, thus careful model specification is necessary. On a positive note, these issues can be resolved by testing multiple models and performing sensitivity analyses. Despite its disadvantages, greedy matching is a useful technique that allows for follow-up multivariate analyses.

Nearest neighbor matching.

The nearest neighbor matching is a type of greedy matching. The nearest neighbor procedure matches based on the nearness of propensity scores of participants in the treatment and control groups. P_i and P_j are propensity scores for the participants in the

treatment and control groups respectively (Guo & Fraser, 2015). The equation for nearest neighbor matching can be expressed as:

$$C(P_j) = min_j \parallel P_i - P_j \parallel, j \in I_0. \quad (3)$$

Where a control participant *j* is a match for a treated participant *i,* if the absolute difference of propensity scores is the smallest among all the possible pairs of participants in the control and treatment groups. Typically, one of the participants in the control group is matched to one participant in the treatment group, also known as 1-to-1 matching. In other scenarios, multiple participants in the control group can be matched to a participant in the treatment group, also known as n-to-1 matching. More matches for participants in the treatment group means better estimates for the counterfactual in the control group (Olmos & Govindasamy, 2015). One of the drawbacks of this technique is whether a large sample size is available, specifically, regarding the need for a greater proportion of participants in the control group in comparison to the treatment group. However, if the researcher experiments with nearest neighbor matching with a caliper approach and conducts sensitivity analyses, this technique has been proven helpful.

Nearest neighbor matching with a caliper.

The nearest neighbor matching with a caliper is the same technique as described above, but with specifications. In this technique, researchers place a restriction based on the absolute distance of propensity scores between the two participants that meet the following condition:

$$\parallel P_i - P_j \parallel < \varepsilon, j \in I_0 \quad (4)$$

Where P_i and P_j represent propensity scores for the participants in the treatment and control groups and ε is the caliper. Rosenbaum and Rubin (1985) recommend using a caliper size of a quarter of standard deviation of the sample estimated propensity scores defined as $\varepsilon < 0.25\sigma_p$ where σ_p indicates the standard deviation of the estimated propensity scores of the sample.

Sensitivity Analysis

A sensitivity analysis in an observational study asks what unmeasured covariate would have to be included in the study to alter the conclusions that the study has reached. In other words, a researcher attempts to determine the extent to which results are susceptible to change in the presence of another variable initially excluded from the study. A problem created by the omission of important variables is known as selection bias. Thus, it is useful to perform a sensitivity analysis to derive a range of possible values attributable to hidden bias. Hidden bias, in this context, refers to an unknown set of values that are unable to be measured. Several sensitivity tests have been developed, each one is a randomization test specific to the type of outcome being analyzed. Typically for non-parametric tests, Wilcoxon's rank sign test has been used and it assumes interval data and a symmetric distribution.

Sensitivity tests are performed on the matched sample using Wilcoxon's rank sign test which evaluated the degree of change in p-values from significant to non-significant or *vice versa* for increasing values of gamma. Gamma is a measure of the extent of departure from random assignment. In other words, a statistically significant change in

the odds of lower/upper bounds demonstrates the magnitude of the treatment effect changes with increasing values of gamma. A study's results are considered sensitive if values of gamma that are near 1 cause changes in significance in comparison to studies that do not possess hidden bias (Rosenbaum, 2002). Thus, the greater the degree of departure from a gamma value of 1.0 the more robust the study's results are.

Longitudinal HLM and Two-Level Growth Model

The development of hierarchical linear models (HLM) has created progress in terms of expanding an array of techniques for conducting research on individual change. The use of HLM is documented in fields of sociology, biometrics, and econometrics since the early 1970s. HLM is an advanced method for the analysis of hierarchical data with complex patterns of variability with a focus on nested sources of variability (Snijders & Bosker, 1999). The goal of HLM is to isolate causal effects by specifying models that statistically control for variables at different levels of aggregation. Since HLM can handle analysis of longitudinal data it can discern the directionality of effects. One of the major advantages of HLM is its ability to accommodate data with a nested structure or hierarchical structure. For example, nested data can refer to repeated observations nested within individuals. HLM handles this hierarchy by defining levels of data within temporally nested data sets. This is possible if there is an abundant level of variability present within a data set. At its core, HLM's objectives are two-fold: 1) create separate models of variable relationship within each level and 2) examine how variables at higher levels predict relationships at lower levels.

A growth curve model (GCM) can represent many instances of individual change. In these models, subjects are measured repeatedly over time to study individual growth. GCM can explain change at the individual and cluster levels, appraise change over time, and account for the impact of personal characteristics. Despite the availability of data on energy consumption and variables related to estimating its effects, the application of hierarchical linear models to data on energy consumption is extremely limited in scope. This is due to varying house built environments which make it difficult to accurately identify all possible variables and interactions that impact the way in which homes are constructed, occupied, and renovated (Hsu, 2015). Another challenge concerns generalizability of findings to different settings because of a lack of data arising from randomized control trials, differences in infrastructure systems, policies, and regulations. This evaluation attempts to advance the fields of energy efficiency and research methods by performing analyses that include data matched on a set of covariates matched through the propensity score matching procedure. In theory, using matched data in the study should provide more accurate, statistically - sound, and practical results. Another important note of modeling of individual change on an outcome measure may involve different patterns of change over time. The data may fit a variety of patterns including linear change, quadratic change (data demonstrates curvature), cubic change (data demonstrates rising and falling patterns), or nonlinear change (data demonstrates floor and ceiling effects) (Hesser, 2015; Raudenbush & Bryk, 2002).

Linear Change

A linear change is a pattern of data that is represented by a straight line that is fit to all time points. The linear change is associated with time and the outcome measure through the specification of two parameters. The two growth parameters refer to the intercept or individual's initial status and a slope or an individual's rate of change. The prediction equation representing linear change is displayed below:

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 * TIME \quad (5)$$

Here, the $\hat{y}$ represents the average outcome measure at time *t* for person *i*. The intercept ($\hat{\beta}_0$) indicates the true ability of person *i* at time of 0. The slope or the average constant rate of change refers to the growth rate for person *i* over the period of data collection. The average constant rate of change is the expected change during one unit of time. In terms of interpretation, as time progresses and the slope is positive, the measure of outcome increases. If the slope is negative and time progresses, the measure of outcome decreases. *TIME* refers to the study's specified period of time under examination (Raudenbush & Bryk, 2002).

Quadratic Change

A quadratic change model is an expansion of the linear change model with three parameters. The third parameter accounts for the growth rate that changes across time. Here the growth rate is time dependent and, thus, changes at each time point. The prediction equation displays the quadratic change below:

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1(TIME) + \hat{\beta}_2(TIME)^2 \quad (6)$$

Where at any given time point, the measure of outcome is estimated with $\hat{\beta}_0$ being the initial status and $\hat{\beta}_1$ denoting an estimate of linear change. While β_2 is representative of an estimate of quadratic change and $\beta_1 + 2\beta_2\ (TIME)$ characterizes the acceleration over time (Raudenbush & Bryk, 2002).

Cubic Change

The use of higher order polynomials to represent individual change is possible at almost any level of complexity. A cubic change is usually observed with data that display rising and falling patterns in the outcome. In a model that captures a cubic change, an additional parameter is added to explain the rate of change in outcome. As in the quadratic change, there are high points and low points that can occur at any point in the data. The new fourth parameter ($\hat{\beta}_3$) provides information on when changes in outcome are gradual or rapid and the direction in the rate of change (Singer & Willet, 2003).

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 * (TIME) + \hat{\beta}_2(TIME)^2 + \hat{\beta}_3(TIME)^3 \qquad (7)$$

Unconditional Model

Initially, in the model building process an unconditional model is estimated which is a model characterized by a lack of level - 2 predictors. This model was developed and analyzed prior to adding predictors to provide empirical evidence for determining a proper specification of the individual growth equation and baseline statistics for evaluating subsequent level - 2 models. In multilevel analyses, the researcher decides if he or she will include fixed or random effects in the models. Fixed effects are equivalent

to regression coefficients representing the average individual effect at initial status and average individual rate of change. Random effects represent the underlying structure of the data which is usually represented by estimates of variability, such as variance and standard deviation. In other words, in a 2 - level HLM, random effects represent variability across level - 1 units and within level - 2 units around the fixed effect. As such, each individual's average initial status and average rate of change randomly varies around the average initial status and group change. Finally, it is imperative to evaluate the degree of variation of initial status and rate of change as significant variability leads to meaningful interpretations and supports the reliability of models under analysis.

Conditional Model

In HLM models, typically level - 1 and level - 2 predictors are assumed to be fixed. Level - 1 intercepts and slopes are left to vary randomly across groups. Due to the assumptions about their error distributions, their variances are called random coefficients. In simpler terms, these coefficients can be seen as coefficients obtained from level - 1 regressions as a type of random variable that comes from and generalizes to a distribution of possible values. In this context, groups are viewed as a subset of the possible groups. In evaluating the random effects of the unconditional model, if there is evidence of statistical significance, additional variance of growth parameters can be explained with the addition of covariates. This results in the conditional model, i.e. conditional on the relationships between the covariates and growth parameters. It can be useful to center continuous covariates so that the average of the covariate is transformed to zero. This

allows for a more meaningful interpretation of growth parameters. Categorical covariates do not need to be centered, however an assignment of a reference category is necessary. The assignment of a reference category enables the researcher to compare the reference category to other levels of the covariate. Estimates of variability in intercept and slope by the addition of covariates to the conditional model can be compared to the estimates of variability in intercept and slope in the unconditional model. In addition, the evaluation of model fit due to the addition of covariates is performed to determine whether the model fits data the best. A comparison of model fit using the chi - square difference test was used to manually determine model fit with a chi - square table of significance (Schermelleh-Engel *et al*., 2003).

Chapter 2: Method

All components of the research methodology used in this study are reported in this chapter. The research design, data sources, and means of data analysis are explained. Electricity use in kWh for low - income households constituted the dependent variable for this study. The information is organized into the following sections: 1) Research Design, 2) Procedure, 3) Participants, 4) Study variables 5) Descriptive statistics, and 6) Data Analysis.

Research Design

This study employed quantitative - based research methods in the form of propensity score matching and hierarchical linear growth models to draw inferences from the data. Due to the nature of the program, all households participated in both the LIHEAP and LIWP. Thus, to measure the impact of upgrades the data were separated into two groups: (1) households who had not had a set of upgrades installed constituted the control group, (2) households who had a set of upgrades installed became the treatment group. This research was undertaken to explore any changes in electricity use over the course of the study. A propensity score method was applied at each month matching the outcome on a set of covariates for a period of twelve months. Then, the

matched monthly data were aggregated culminating in a data set ready for a hierarchical linear model. A hierarchical linear growth model was used to examine the trajectory of electricity consumption for the control and treatment groups after controlling for covariates. To compare the results from propensity score - based models, another hierarchical growth model (without propensity score matching) was fit on the entire sample to detect any differences in electricity consumption and whether the effect of treatment was statistically significant.

Procedure

Enrollment of participants in LIHEAP occurs over a period of 6 months starting on November 1 and ending on April 30th of the following year. As part of this application, consumers consent to disclose their energy consumption data that includes utility account payment history and general energy usage data for up to 24 months to the LIHEAP office. Once the disclosure of data is authorized by the participant, data collection begins on the date that the participant signs the application and ends when energy assistance program participation is terminated. The LIHEAP application collects information regarding applicant's demographics, age, number of members in the household, type of residence, disability status, sex, and unworked income. In addition, personal information, such as household income of the applicant and other household members is collected. These data are collected via a paper application filled out by the consumer and either mailed or hand-delivered to the LIHEAP office.

The application for low - income weatherization services contains some of the same data as the LIHEAP application, but also lists more detailed information on households. This information includes data on heating system type, type of fuel, square footage of home, type of water heater fuel, and occupant status. The information on these covariates was included in the propensity score model and came from the Salesforce database maintained by the City and County of Denver's Department of Environmental Quality and Environmental Health (DEQEH). The data on customers' electricity use was obtained from Xcel Energy through a partnership with the Denver Office of Strategic Partnerships (DOSP). DOSP used a utility release form to capture consent to disclose utility customer data. The data for this evaluation were obtained through a submission of a data use agreement form to the LIHEAP office. The data use agreement detailed data user obligations including confidentiality, public release, and data ownership. A confidentiality statement was signed to ensure that the data are used solely for business purposes as intended. Thus, the dataset for this evaluation was assembled from multiple sources including the data from the LIHEAP office, Xcel, DEQEH and DOSP. Since the data have already been collected by another party for purposes other than this evaluation, they are considered secondary data. The University of Denver's Institutional Review Board (IRB) was contacted to determine if a formal approval for the use of these data was necessary. Since the data had already been collected, an application for secondary data with waiver of personal consent was used. The IRB determined that the application was exempt.

Participants

The sample consisted of households in Denver County in Colorado with at least 2 months and up to 12 months of outcome data. The criteria for inclusion based on LIHEAP and LIWP eligibility were discussed in the previous section. The total sample of the study consisted of 813 households and 7897 observations of participant data. The matched sample following propensity score analysis consisted of 813 households and 4022 participant observations. Variables that were provided to the researcher about households included applicant's age, primary heating fuel, square footage of home, type of water heating fuel, number of household occupants, type of residence, whether applicant owns or rents a home, whether any member of the household is disabled, applicant's sex, applicant's race, whether any member of the household receives non-work income (i.e., public assistance programs), and whether the applicant submits payment for his or her electricity use through a vendor or a client.

Study Variables

Energy efficiency upgrades defined treatment for this study and acted as an independent variable (indicated whether an individual was in the control group = 0, or treatment group = 1). The treatment for this study consisted of enrollment and participation in LIWP through a set of upgrades performed on a household. Upgrades included one or more of the following: air-sealing (professional), ceiling/attic insulation, compact fluorescent lamp (CFL) in 40 Watt, 60 Watt, 75 Watt or 100W Equiv., clothes washer, dish washer, weather stripping, door replacement, floor/crawl insulation, LED,

refrigerator, solar photovoltaic system (PV), tank water heater, thermostat, wall insulation, water heater blanket, window replacement, and water heater pipe insulation. The evaluation examined whether enrollment in LIWP contributed to a change in electricity use (in kWh) over the period of examination.

In terms of covariates, the variables were chosen based on their theoretical and research - based relationship to electricity use. A total of 14 covariates were included to measure their impact on both the propensity score model and the hierarchical linear growth models under examination. The covariates for the study included the following characteristics: age of applicant (Judson & Maller, 2014); number of members in the household (Tonn *et al*., 2014), primary fuel for heating (Eisenberg, 2014); unworked income, such as enrollment in public assistance programs, i.e., Temporary Assistance for Needy Families (Hsu, 2015; Tonn *et al*., 2014); type of dwelling (Davis, 2011; Hsu, 2015); race (Davis, 2011; Hsu, 2015); size of dwelling in square feet (Eisenberg, 2014); fuel for water heater (Eisenberg, 2014); status of home ownership (Davis, 2011); sex (Tonn *et al*., 2014); and method of payment for electricity use. Age, size of home in square feet, and number of household members served as continuous predictors. All other covariates served as dichotomous predictors of electricity use.

Dichotomous covariates were assigned a reference category that allowed for drawing comparisons between the reference category and other levels of the covariate. For example, for primary fuel for heating the reference category was defined as natural gas otherwise primary heating fuel was classified as electric. For unworked income, if an applicant indicated that he or she receives public assistance (aside from LIHEAP

assistance) the reference category was defined as yes, otherwise no. For type of dwelling, a house or modular home served as the reference category, otherwise the categories included townhome, duplex/triplex/fourplex, and apartment. For race, two variables were created: the first variable indicated whether the applicant was Black; otherwise the category included individuals of White, Hispanic, American Indian, Asian, or Native Hawaiian descent; the second variable indicated whether the applicant was Hispanic, otherwise race categories were operationalized as indicated for the first variable above for race. For water heater fuel type, natural gas was the reference category which was compared to electric as the other level of the variable. For status of home ownership, those who reported they rent their residence were classified as the reference category in comparison to those who reported that he or she owns it. For sex, those who identified as males were in the reference category in comparison to females. For method of payment, those who paid for their electricity use directly to the utility provided were in the reference category, otherwise they were categorized as a client for payment purposes.

Descriptive Statistics

The descriptive statistics for longitudinal HLM analyses were calculated in SPSS 23 while the descriptive statistics for propensity score analyses were calculated in RStudio 3.1.0. The initial exploration of data included variables that were assessed through univariate and bivariate statistics, such as correlation matrices, means, and standard deviations (SD). A correlation matrix and means and standard deviations of

continuous variables in the matched data set are provided in Table 3. A descriptive summary of categorical variables is displayed in Table 4.

Table 3

A Correlation Matrix and Means (SD) for Analysis Variables

Variables	a	b	c	d	e	f	g	h	i	j	k	l	m
a. Total Usage	1												
b. Age	-.07**	1											
c. Primary Heating Fuel	-.07**	.02	1										
d. Square Feet	.11**	-.08**	-.03	1									
e. Water Heating Fuel	-.06**	.05**	.72**	-.02	1								
f. Number of HH Members	.21**	-.52**	.02	.16**	-.03	1							
g. House	.08**	.05**	.09**	.03	.09**	.08**	1						
h. Rent Own	.04*	-.36**	-.00	-.01	.00	.17**	-.30**	1					
i. Disabled	.00	.22**	.02	-.09**	.04*	-.16**	-.04**	.05**	1				
j. Sex	.00	.16**	-.00	-.05**	.02	-.03*	.01	-.13**	.06**	1			
k. Hispanic	-.02	-.10**	.02	-.12**	.07**	.15**	.11**	.00	-.05**	-.04*	1		
l. Black	.08**	-.01	-.05**	.08**	-.05**	-.04**	-.10**	.17**	.07**	-.04*	-.50**	1	
m. Unworked Income	.06**	-.43**	-.07**	.08**	-.07**	.42**	.05**	.12**	-.35**	-.06**	.07**	-.05**	1
n. Payment Method	-.01	-.00	-.02	.00	-.02	.11**	.05**	-.08**	-.11**	-.05**	.06**	-.12**	.01
Mean	654.58	54.41	+	1184.50	+	2.66	+	+	+	+	+	+	+
SD	436.40	16.68	+	652.95	+	1.67	+	+	+	+	+	+	+

Table 4

Descriptive Summary of Dichotomous Variables

Dichotomous Variable	Frequency	Percent (%)
Intervention		
Control	2011	50%

Treatment	2011	50%
Primary Heating Fuel		
Electric	134	3%
Natural Gas	3888	97%
Water Heating Fuel		
Electric	87	2%
Natural Gas	3935	98%
Type of Dwelling		
House/Modular Home	3352	83%
Townhome/Duplex/Triplex/Apartment	670	17%
Occupant Status		
Own	1557	39%
Rent	2465	61%
Household Member Disability Status		
No	2857	71%
Yes	1165	29%
Sex		
Male	856	21%
Female	3166	79%
Race		
Hispanic	1814	45%
Black	932	23%
White	586	15%
American Indian, Asian, and Native Hawaiian	289	7%
Other	404	10%
Unworked Income		
No	2735	68%
Yes	1287	32%
Payment Method		
Vendor – paid	3977	99%
Client – paid	45	1%

Data Analysis

Propensity Score Matching

RStudio 3.1.0. was used to perform this part of the analyses. In the case of observational studies, often there is a relatively small group of participants who receive treatment in comparison to a much larger group of participants in the control condition.

When the costs of conducting a randomized - control trial are high or not possible, selective sampling of participants in the control condition is advised. The controls for the study are matched so that they are similar to the treated subjects on a set of measured background variables (Rosenbaum & Rubin, 1985). Before a propensity score model was estimated, diagnostic tests were performed to determine the degree of similarity of groups on covariates. The crucial step in this phase is the specification of covariates for the propensity score model. This is because the estimation of the treatment effect is dependent on the covariates used. As such, prior studies were analyzed in the determination of key covariates that were included in the model. Once the covariate section was complete, the researcher assessed whether the groups were balanced on the covariates. The researcher applied statistical techniques to determine if the covariates were balanced across groups. For example, Imbens and Woolridge (2009) recommend reporting the difference in averages by treatment status, scaled by the square root of the sum of the variances for each covariate used in the model. This equation is also known as the normalized difference and is expressed below:

$$\Delta_X = \frac{\bar{X}_1 - \bar{X}_0}{\sqrt{S_0^2 + S_1^2}}, \tag{8}$$

Where for ω = 0,1, $S_\omega^2 = \sum_{i:Wi=\omega}(x_i - \bar{x}_\omega)^2 / (N_\omega - 1)$, the sample variance of x_i in the subsample with treatment $W_I = \omega$. A propensity score matching model was used to balance the control and treatment groups on the observed covariates. The data on electricity consumption over the course of a year were assigned propensity scores through greedy matching of 14 covariates at each month of analysis. Then, the treatment effect

was estimated by taking the average of treatment effects at all 12 months under analysis. The vectors of observed covariates or propensity scores were computed based on the conditional probability of a household's status of installation of upgrades(s). Thus, the households that installed the upgrades(s) were designated the treatment group and households that were yet to install the upgrade(s) were classified as the control group.

The most common technique that estimates the conditional probability of receiving treatment using a vector of observed covariates is binary logistic regression. The equation for the binary logistic regression can be expressed as follows (Guo & Fraser, 2015):

$$P(W_i|\boldsymbol{X}_i = x_i) = E(W_i) = \frac{e^{x_i\beta_i}}{1+e^{x_i\beta_i}} = \frac{1}{1+e^{-x_i\beta_i}} \quad (9)$$

Where $\boldsymbol{W}_i$ indicates binary treatment condition ($\boldsymbol{W}_i$= 1, if the participant is in the treatment condition and $\boldsymbol{W}_i$= 0, if the participant is in the control condition) for the ith case. The vector of covariates is denoted as $\boldsymbol{X}_i$. The vector of regression parameters is denoted as β_i.

When we assume that there are only two conditioning variables x_1 and x_2 , the log-likelihood model function of the equation above can be expressed as follows:

$$log_{e^l}(\beta_0, \beta_1, \beta_2) = \sum_{i=1}^{n} W_i(\beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i}) - \sum_{i=1}^{n} log_e[1 + exp(\beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i}).] \quad (10)$$

Where estimated values of β_0, β_1, and $\beta_2$are logistic regression coefficients that maximize the propensity to reproduce sample observations. In applying the binary logistic regression equation logistic regression coefficients: β_0, β_1, and β_2 become, $\widehat{\beta_0}$ $\widehat{\beta_1}$, and $\widehat{\beta_2}$ and thereby estimate the propensity scores for each sample participant i.

Hmisc (Harrell *et al*., 2016) statistical package within R was used to assess the balance of groups prior to matching through histograms and tests of standardized difference. A function within the MatchIt (Ho *et al*., 2011) package calculated risk set pairs between the control and treated individuals. Thereby, this algorithm estimated propensity scores for each pair of groups (one from treated and one from control).

Propensity score analysis was performed using two matching approaches: nearest neighbor 1:1 and nearest neighbor with caliper. Rosenbaum and Rubin (1985) suggest one quarter of one standard deviation should be set as a caliper threshold. The caliper specifies a vector within which the matched units may be positioned. So, a caliper of .25 indicates that matches must be within .25 of one standard deviation to be kept, otherwise the match is dropped. Another option is 1-n where treatment is matched to multiple controls units. For this option, the specifications designate how many times units within the treatment group can be reused. For example, in 2:2, participants in the largest group are matched up to two times and the participants in the second largest groups are matched up to two times. Both the caliper and the 1- n options were used to assess balance on the covariates included in the model.

Power Analysis

Prior to performing hierarchical linear models a power analysis was performed *a priori* using Optimal Design software to evaluate the probability of detecting the effect of treatment if a true effect is present (Spybrook *et al*., 2011). Power was estimated using

the repeated measures design for treatment on quadratic change. For a given power of .80, a total sample size of 178 is necessary to detect an effect size of .65.

Hierarchical Linear Growth Model

HLM 7.01 (Scientific Software International, Inc.) was used to perform analyses involving hierarchical linear growth models. The outcome of the study was depicted graphically over time with the treatment and control groups included. The trajectory of electricity use indicated a quadratic trend due to curvatures in the monthly data (Figure 2). The outcome demonstrated increases and decreases over the course of a time period of June 2013 to May 2014.

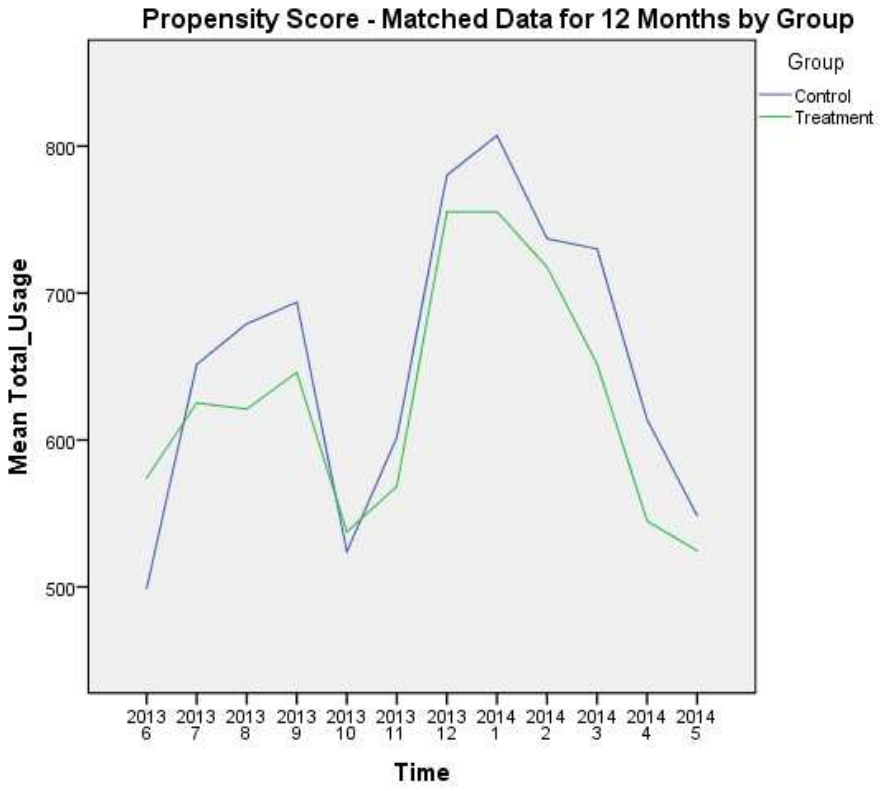


Figure 2. Graph of the Propensity Score- Based data for 12 months by Group

Unconditional and Conditional Models

Prior to fitting the quadratic model, an unconditional model with a linear trajectory was fit in which no covariates were included. Since the unconditional model resulted in statistically significant variability across intercepts and slopes, covariates were included to further explain variance in the growth parameters. The addition of covariates resulted in a model that was conditional on the relationships between the covariates and growth parameters. Three continuous predictors were grand mean centered. Grand mean centering subtracts the grand mean of the predictor using the mean from the full sample. Typically, centering makes the value of initial status more interpretable because the expected value of Y when *x* is zero represents the expected value of Y when X is at its mean (Algina & Swaminathan, 2011).

Chapter 3: Results

The primary goal of this evaluation was to determine whether enrollment in LIWP had an impact on electricity consumption of households resulting in a statistically significant change between the treatment and control groups after controlling for a set of covariates. The secondary goal concerned the application of propensity score matching and a hierarchical linear growth model to analyze longitudinal monthly electricity data over twelve months as compared to a hierarchical linear growth model without propensity score matching.

This chapter is organized into three sections following a summary of descriptive statistics. The first section focuses on the results of propensity score matching utilizing two matching techniques. Balanced risk set matching was performed at each month under analysis in RStudio 3.1.0. Matching was performed with optimal and full matching techniques, but the models did not converge, so matching with nearest neighbor and nearest neighbor with caliper approaches were reported. The second section focused on tests of assumptions, estimates of intercepts and growth parameters for the HLM model with the propensity score matched subsample. HLM 7.01 (Scientific Software International, Inc.) was used to perform longitudinal HLM models that included unconditional and conditional models using households' electricity consumption as the

outcome. In the third section, results of longitudinal HLM analyses without propensity score matching are presented. The preliminary phase of analysis involved analyses of data through univariate and descriptive statistics. The dependent variable and covariates were examined in SPSS 23. Covariates included in the models included age, sex, primary heating fuel, square footage of household, water heater fuel type, number of household members, type of household, status of household, presence of disability, race, unworked income, and method of utility payment. The assumption of normality was tested with skewness for continuous variables. Of the continuous variables, size of home in square feet and electricity use were noted for issues of skewness. Violations of assumptions of normality and heterogeneity of variance (HOV) were found, but the analysis was deemed robust due to a large sample size of 4022 observations and 813 cases. In addition, a balanced design with 2011 of households in the treatment group and 2011 in the control group minimized effects of violation of assumptions.

Propensity Score Analyses

The results from the propensity score analyses using nearest neighbor with a ratio 1:1 and nearest neighbor with a caliper 0.1. Results were reported in terms of standardized differences observed on covariates prior to matching and post matching. The assignment of the treatment and control groups were determined based on households' status on the intervention. So, the months over the course of which households had not yet had the upgrade(s) installed were considered controls. The months that followed the installation of upgrade(s) were determined in the treatment

condition. Since all households in the sample performed the upgrade(s) at some point, all households had data prior to and following the intervention.

Prior to matching, descriptive statistics were performed to examine the balance among the covariates at each month starting in June of 2013 through May of 2014. Standardized differences were calculated for each covariate at each month under analysis. Thus, propensity score matching was performed on the treatment and control groups with a set of covariates at each month over a course of a year. The propensity score is a balancing score representing a combination of observed covariates. So, if a pairing of households across the treatment and control groups shared a similar propensity score, the pair is considered equivalent despite having potentially different values on any of the covariates. The propensity scores were used in matching of households across the treatment and control group with the goal of making them equivalent to more accurately estimate the effect of treatment.

In the process of estimating propensity scores, standardized differences were computed to evaluate balance following the application of the matching technique. Standardized differences were estimated at each month prior and following the matching procedures. Balance between the groups was also assessed visually with the use of back - to - back histograms of the covariate distributions. Results demonstrated lower values on distances between the treatment and control groups for the nearest neighbor matching technique for June of 2013 – March of 2014. To avoid redundancy in results of standardized difference tests only results for four months of comparisons are reported here. These include June of 2013, July of 2013, April of 2014, and May of 2014 (the first

two months and the last two months of analysis). Results of standardized difference tests for August of 2013 – March of 2014 are provided in table format in Appendix A. Visual representations of back - to - back histograms are provided for each month under analysis in Appendix A. Sample sizes of the whole and matched samples are presented in Table 5. Results of standardized difference tests for the months of June and July in 2013 (first two months of analyses) prior to and following nearest neighbor matching are displayed in Table 6. Additionally, results of standardized difference tests for the months of April and May 2014 (last two months of analyses) prior to and following nearest neighbor matching with caliper are displayed in Table 7.

Table 5

Whole and Matched Sample Sizes across the Treatment and Control Conditions

Time Point		**Sample Size**			
		Control (All)	Control (Matched)	Treatment (All)	Treatment (Matched)
2013	June	512	33	35	33
	July	571	39	41	39
	August	585	46	48	46
	September	496	96	99	96
	October	550	184	197	184
	November	436	206	227	206
	December	441	225	249	225
2014	January	505	249	268	249
	February	423	209	227	209
	March	361	331	471	331
	April	251	216	526	216
	May	192	177	552	177
Total			2011		2011

Table 6

Evaluation of Standardized Differences Pre - and Post- Matching for Covariates in June and July of 2013 Using Nearest Neighbor 1:1 Matching

Time point	June 2013		July 2013	
	Pre – Matching	After NN 1:1	Pre – Matching	After NN 1:1
Age	-0.01	-0.02	0.03	0.01
Primary heating fuel	-0.10	0.36	-0.08	0.11
Square feet	-0.74	-0.08	-0.64	0.14
Water heating fuel	-0.04	0.17	-0.02	0
Number of household members	-0.27	-0.24	-0.27	-0.07
Type of dwelling	-0.09	0	-0.05	-0.06
Ownership status	-0.11	0.06	-0.25	0
Disability status	0.16	0.06	0.12	0.20
Sex	-0.17	0.15	-0.24	0.07
Hispanic	-0.13	0	-0.12	-0.05
Black	0.13	0.12	0.19	0.15
Unworked income	-0.15	-0.12	-0.13	-0.27
Payment method	-0.23	-0.12	-0.21	-0.11

A total of 13 covariates were included in calculating standardized mean differences between the treatment and control groups in June of 2013 (Time 1). As race had more than two categories, two dummy variables were created to represent Black and Hispanic categories. Age, square feet, and number of household members served as continuous variables and the rest of the variables were classified as dichotomous. In June of 2013, the standardized mean difference prior to matching was high in reference to square feet (-.735) followed by number of household members (-.272) and payment method (-.226) (Table 6).

Following the completion of preliminary analyses, the propensity score model was estimated using a general linear model with the covariates listed above. Results

demonstrated lower values on distance between the treatment and control groups for the nearest neighbor matching technique for June of 2013 – March of 2014. Lower mean standardized differences were observed between most, but not all covariates. In June of 2013, notable improvements were found for square feet (-.080) followed by number of household members (-.242), disability status (-.058), and ownership status (-0.058), sex (.150), Black (.121), unworked income (-.123) and payment method (-.121). In this month, only primary heating fuel (.364) fell short in the assessment of balance based on an absolute standardized difference rule of thumb of above .25. Therefore, the inclusion of this variable into outcome analyses did not necessarily aid in the ability of the model to infer treatment effects since the covariate balance was not improved following matching. However, back - to - back histograms indicated that matching improved the overall balance between covariates between the treatment and control groups (See Appendix A).

The 13 covariates under analysis were assessed for balance in July of 2013 (Time 2). Age, number of household members, and square feet served as continuous variables. As race had more than two categories, Black and Hispanic were used as dummy variables. The rest of the variables were dichotomous. In July of 2013, the standardized mean difference prior to matching was high in reference to square feet (-.640) followed by number of household members (-.265) and ownership status (-.248) (Table 7).

The propensity score model was estimated using a general linear model with the covariates listed above in July of 2013. Results demonstrated lower values on distance between the treatment and control groups for the nearest neighbor matching technique in July of 2013. Lower mean standardized differences were observed between most, but not

all covariates. In July of 2013, notable improvements were found for square feet (.142) followed by number of household members (-.065), age (-.014), sex (.068), Black (.152), Hispanic (-.049), and payment method (-.112). In this month, only unworked income (-0.272) fell short of in the assessment of balance based on an absolute standardized difference rule of thumb of above .25. Therefore, the inclusion of this variable into outcome analyses did not necessarily aid in the ability of the model to infer treatment effects as the covariate balance was not improved following matching. However, back - to - back histograms indicated that matching improved the overall balance between covariates between the treatment and control groups (See Appendix A).

Table 7

Evaluation of Standardized Differences Pre- and Post- Matching for Covariates in April and May of 2014 Using Nearest Neighbor 1:1 Matching with Caliper

Time point	April 2014		May 2014	
	Pre – Matching	After NN 1:1	Pre – Matching	After NN 1:1
Age	0.08	0.03	-0.11	0.02
Primary heating fuel	-0.07	-0.02	0.01	-0.06
Square feet	-0.28	-0.07	-0.22	-0.11
Water heating fuel	-0.06	0.08	-0.02	-0.07
Number of household members	-0.07	-0.02	0.05	-0.06
Type of dwelling	-0.03	0.08	-0.05	-0.09
Ownership status	0.13	0	0.34	0.06
Disability status	0.04	0.04	-0.12	-0.07
Sex	0.04	-0.07	-0.02	-0.05
Hispanic	0.08	-0.01	0.13	-0.03
Black	-0.06	-0.07	-0.01	0
Unworked income	-0.09	-0.13	-0.02	0.01
Payment method	-0.02	0	0.03	-0.09

The 13 covariates under analysis were assessed for balance in April of 2014 (Time 11). Age, number of household members, and square feet served as continuous variables. As race had more than two categories, Black and Hispanic were used as dummy variables. The rest of the variables were dichotomous. In April of 2014, the standardized mean difference prior to matching was high in reference to square feet (-.275) followed by ownership status (-.127) (Table 7).

The propensity score model was estimated using a general linear model with the covariates listed above in April of 2014. There was a convergence issue with the nearest neighbor matching approach in April of 2014. The distance between the groups prior to matching was less than the distance reported following the matching in this month. Thus, for this month the nearest neighbor algorithm was not optimal. So, the nearest neighbor with caliper matching was used in April of 2014 to complete the data set for the next steps of analysis. Results demonstrated lower values on distance between the treatment and control groups for the nearest neighbor with caliper matching technique in April of 2014. Lower mean standardized differences were observed between most, but not all covariates. In April of 2014, notable improvements were found for square feet (-.074) followed by Hispanic (-.008), age (.028), primary heating fuel (-.022), and number of household members (-.022). In this month, there were no variables that fell short of in the assessment of balance based on an absolute standardized difference rule of thumb of above .25. Improvements in balance were also displayed visually via back - to - back histograms indicating that matching improved the overall balance between covariates between the treatment and control groups (See Appendix A).

The 13 covariates under analysis were assessed for balance in May of 2014 (Time 12). Age, number of household members, and square feet served as continuous variables. As race had more than two categories, Black and Hispanic were used as categories. The rest of the variables were dichotomous. In May of 2014, the standardized mean difference prior to matching was high in reference to ownership status (-.219) followed by square feet (.335) (Table 7).

The propensity score model was estimated using a general linear model with the covariates listed above in May of 2014. There was a convergence issue with the nearest neighbor matching approach in May of 2014. The distance between the groups prior to matching was less than the distance reported following the matching in these two months. Thus, for this month the nearest neighbor algorithm was not optimal. So, the nearest neighbor with caliper matching was used in May of 2014 to complete the data set for the next steps of analysis. Results demonstrated lower values on distance between the treatment and control groups for the nearest neighbor with caliper matching technique in May of 2014. Lower mean standardized differences were observed between most, but not all covariates. In May of 2014, notable improvements were found for ownership status (.059), square feet (-.110) followed by age (.024), and Hispanic (-.033). In this month, there were no variables that fell short of in the assessment of balance based on an absolute standardized difference rule of thumb of above .25. Improvements in balance were, also, displayed visually via back-to-back histograms indicating that matching improved the overall balance between covariates between the treatment and control groups (See Appendix A).

Overall, the two matching algorithms resulted in a decrease in distance in the balance of covariates between the treatment and control conditions during the period of June 2013 through May 2014. Not all variables improved after matching was performed, but for most variables improvements were observed. The sample sizes of whole and matched samples indicated that the months that retained the most cases were in March of 2014 with 624 observations across the treatment conditions, followed by 536 in January of 2014, and 498 in December of 2013. The matched samples across the 12 time points were aggregated to form a data set for outcomes analysis that followed with hierarchical linear modeling.

Hierarchical Linear Model with the Propensity Score-based Sample

Initially, unconditional models were fit to map the average trajectory of electricity consumption across the propensity score-matched sample. Results indicate that the mean intercept of households' electricity use was not significantly different from zero, $\beta_{00} = 33.09$, $p = .529$. This denotes the true electricity consumption of household*i* at time of 0. The mean slope for growth rate predicting electricity use was significantly different from zero, $\beta_{10} = 185.62$, $p < .001$. The mean acceleration rate for growth rate predicting electricity use was significantly different from zero, $\beta_{20} = -12.07$, $p < .001$.

The chi-square test of level-2 residual variance of the intercept was 1689.64 ($df = 254$, $p < .001$). The corresponding chi-square hypothesis test on the residual variance for growth rate was, $\chi^2 = 1220.09$ ($df = 254$, $p < .001$). The $\chi^2 = 20.52$ ($df = 254$, $p < .001$) for the rate of acceleration was also significant (Table 8). This provides evidence that there

was significant between group variance in the intercept, slope, and acceleration parameters across groups. The intercept term represents the between household group variance in electricity consumption. Statistical significance indicated substantial variation in electricity consumption at initial status and each growth parameter providing evidence for further exploration of household-level predictors to model the variation at initial status and the growth rate.

Table 8

Summary of Fixed and Random Effects for the Unconditional Model (Matched Sample)

Fixed effects	**Coefficient**	**SE**	***t***	***P***
Mean electricity use, β_{00}	33.09	52.51	0.63	0.529
Mean growth rate, β_{10}	185.62	15.14	12.26	<.001
Mean acceleration rate, β_{20}	-12.07	0.93	-12.92	<.001
Random effects	**Variance Component**	***d.f.***	χ^2	***P***
Initial status, r_{0i}	1054551.28	254	1689.64	<.001
Growth rate, r_{1i}	108339.18	254	1220.09	<.001
Acceleration rate, r_{2i}	421.15	254	950.02	<.001
Level-1, e_{ti}	23112.40			
Deviance = 57672.01 with 7 df				

Conditional models were built because statistically significant variability was reported in the growth parameters of the unconditional model. Covariates at level-2 were included in the conditional model to explain the variability and to examine the household-level trajectories. Covariates that were related to the linear growth and quadratic growth parameters were included in the model. Conditional models were

constructed using the treatment variable as a level-2 predictor of electricity use. The variables that were included in the propensity score model were also included in the hierarchical linear model based on their association to the outcome and treatment. The covariates were included because of their actual correlation to the outcome and treatment or were considered related to the outcome and treatment based on theoretical evidence. Due to these factors, the following set of covariates were included to reduce selection bias in the estimate of the effect of treatment. Additional covariates included were age, primary heating fuel, square feet, water heater fuel, number of household members, dwelling type, ownership status, disability status, sex, race, unworked income, and payment method. The model converged with all covariates included, so all variables were retained.

The parameter estimates for continuous variables were interpreted as the amount of change in electricity use for a 1 unit change in the average value on the covariate in the sample. The intercept of households' electricity use was not significantly different from zero, $\beta_{00} = 916.33$ ($p = .113$). The estimated mean growth rate for electricity use in kWh was not statistically significant, $\beta_{10} = -5.13$ ($p = 0.966$). This means that average electricity consumption decreased by an average of 5.13 units per month over the course of a year. The results were not significant indicating that mean electricity use did not significantly differ across households and the mean rate of electricity use did not change in a given month holding all other variables constant. The mean acceleration was $\beta_{20} = -0.81$ which accounts for the rate of change that varies at each time point ($p = .909$). So, on average, households' electricity use varied in a negative direction at each time point.

At initial status, sex of the applicant, (β_{010} = -322.59, p = .035) was reported as a significant predictor of electricity use. So, in this case whether the applicant's sex was female or male made a difference in his or her electricity use after holding all other variables constant. The treatment variable was not a significant predictor of electricity use, with the treatment group having a reduction of 22.26 units (p =.371) in comparison to the control group. All other variables were not significant.

The chi-square test on level-2 residual variance of the intercept was 4105.62 (df = 736, p <.001). The corresponding chi-square test on the residual variance for growth rate was, χ^2 = 2571.02 (df = 736, p <.001). The chi-square tests indicate that there was statistically significant between group variance in the intercept and slope parameters across groups. The intercept term represents the between household variance in electricity use after controlling for the impact of treatment, age, primary heating fuel, square feet, water heater fuel, number of household members, dwelling type, ownership status, disability status, sex, race, unworked income, and payment method. For a summary of fixed and random effects of the linear component of the conditional model see Table 9. A summary of fixed effects for the quadratic component of the conditional model is provided in Table 10.

Table 9

Summary of Fixed and Random Effects for the Linear Conditional Model (Matched Sample)

Fixed effects	**Coefficient**	**SE**	***t***	***P***
Mean electricity use, β_{00}	916.33	576.95	1.59	0.113
Treatment, β_{01}	40.12	92.84	0.43	0.666

Age, β_{02}	-3.66	3.88	-0.94	0.346
Primary heating fuel, β_{03}	-344.26	469.18	-0.73	0.463
Square feet, β_{04}	0.02	0.11	-0.73	0.878
Water heater fuel, β_{05}	266.08	546.34	0.49	0.626
Number of household members, β_{06}	-17.45	50.61	-0.35	0.730
Dwelling type, β_{07}	5.22	121.22	0.04	0.966
Ownership status, β_{08}	-135.98	121.13	-1.12	0.262
Disability status, β_{09}	26.98	110.04	0.25	0.806
Sex, β_{010}	-322.59	152.63	-2.11	0.035
Hispanic, β_{011}	-35.03	107.53	-0.33	0.745
Black, β_{012}	96.64	122.59	0.79	0.431
Unworked income, β_{013}	95.07	143.34	0.66	0.507
Payment method, β_{014}	-445.52	438.14	-1.02	0.310
Mean growth rate, β_{10}	-5.13	120.33	-0.04	0.966
Treatment, β_{11}	-22.26	24.84	-0.90	0.371
Age, β_{12}	0.93	0.97	0.96	0.336
Primary heating fuel, β_{13}	52.86	107.03	0.49	0.622
Square feet, β_{14}	0.01	0.03	0.50	0.615
Water heater fuel, β_{15}	-105.12	124.72	-0.84	0.400
Number of household members, β_{16}	15.87	11.29	1.41	0.160
Dwelling type, β_{17}	37.55	30.70	1.22	0.222
Ownership status, β_{18}	37.65	31.20	1.21	0.228
Disability status, β_{19}	-2.97	28.39	-0.11	0.917
Sex, β_{110}	74.17	39.32	1.89	0.060
Hispanic, β_{111}	13.70	26.83	0.51	0.610
Black, β_{112}	5.22	31.92	0.16	0.870
Unworked income, β_{113}	-20.46	33.20	-0.62	0.538
Payment method, β_{114}	111.62	82.07	1.36	0.174

Random Effects	**Variance Component**	***d.f.***	χ^2	***p***
Initial status, r_{0i}	556532.10	736	4105.62	<.001
Growth rate, r_{1i}	3526.04	736	2571.02	<.001
Level-1, e_{ti}	32714.22			
Deviance = 58041.21 with 4 df				

Table 10

A Summary of Fixed Effects for the Quadratic Component of the Conditional Model (Matched Sample)

Fixed effects	Coefficient	SE	*t*	*p*
Mean acceleration rate, β_{20}	-0.80	7.06	-0.11	0.909
Treatment, β_{21}	2.08	1.51	1.38	0.168
Age, β_{22}	-0.04	0.06	-0.71	0.479
Primary heating fuel, β_{23}	-2.59	5.95	-0.44	0.664
Square feet, β_{24}	-0.001	0.001	-0.79	0.428
Water heater fuel, β_{25}	5.95	7.11	0.84	0.403
Number of household members, β_{26}	-0.81	0.70	-1.17	0.243
Dwelling type, β_{27}	-2.89	1.84	-1.57	0.117
Ownership status, β_{28}	-2.15	1.85	-1.16	0.246
Disability status, β_{29}	0.03	1.67	-1.02	0.987
Sex, β_{210}	-3.88	2.30	-1.68	0.093
Hispanic, β_{211}	-0.96	1.63	-0.59	0.557
Black, β_{212}	-0.60	1.91	-0.31	0.754
Unworked income, β_{213}	0.56	2.00	0.28	0.778
Payment method, β_{214}	-6.37	4.67	-1.36	0.713

The proportion of variance explained (PVE) by treatment and covariates indicates the amount of variance in electricity consumption at intercept and for the growth parameter. The estimated PVE by the listed above covariates was .47 for initial status. This indicates that the proportion of variance calculation resulted in 47% of the variance at initial status is accounted for by treatment, age, primary heating fuel, square feet, water heater fuel, number of household members, dwelling type, ownership status, disability status, sex, race, unworked income, and payment method. This indicates that 53% of the variance remained unexplained in the households' differences in electricity use at initial status. The estimated PVE explained by the listed above covariates for the linear growth rate was 0.96. So, 96% of the variance in the slope of the mean linear growth rate of

electricity use was explained by treatment, age, primary heating fuel, square feet, water heater fuel, number of household members, dwelling type, ownership status, disability status, sex, race, unworked income, and payment method. So, only 4% of the variance in electricity use at the growth rate remained unexplained.

Hierarchical Linear Model with the Entire Sample

Initially, unconditional models were fit to map the average trajectory of electricity consumption across the whole sample of households across 12 months of data. Results indicate that the mean intercept of households' electricity use was significantly different from zero, $\beta_{00} = 440.78$, $p < .001$. The mean slope for growth rate predicting electricity use was significantly different from zero, $\beta_{10} = 87.25$, $p < .001$. The mean acceleration rate for growth rate predicting electricity use was significantly different from zero, $\beta_{20} = -6.49$, $p < .001$.

The chi-square test of level-2 residual variance of the intercept was 2946.25 ($df = 658$, $p < .001$). The corresponding chi-square hypothesis test on the residual variance for growth rate was, $\chi^2 = 2265.48$ ($df = 658$, $p < .001$). The $\chi^2 = 1689.82$ ($df = 658$, $p < .001$) for the rate of acceleration was also significant (Table 11). This provides evidence that there was significant between group variance in the intercept, slope, and acceleration parameters across groups. The intercept term represents the between household group variance in electricity consumption. Statistical significance indicated substantial variation in electricity consumption at initial status and at each growth parameter

providing evidence for further exploration of household-level predictors to model the variation at initial status and the growth rate.

Table 11

Summary of Fixed and Random Effects for the Unconditional Model (Whole Sample)

Fixed effects	**Coefficient**	**SE**	***t***	***p***
Mean electricity use, β_{00}	440.78	20.05	21.98	<.001
Mean growth rate, β_{10}	87.25	6.40	13.63	<.001
Mean acceleration rate, β_{20}	-6.49	0.41	-16.00	<.001
Random effects	**Variance Component**	***d.f.***	χ^2	***p***
Initial status, r_{0i}	205852.82	658	2946.25	<.001
Growth rate, r_{1i}	22301.65	658	2265.48	<.001
Acceleration rate, r_{2i}	9.00	658	1689.82	<.001
Level-1, e_{ti}	203.04			
Deviance = 110220.87 with 7 df				

Conditional models were built because statistically significant variability was found in the growth parameters of the unconditional model. Covariates at level-2 were included in the conditional model to explain the variability and to examine the household-level trajectories. Covariates that were related to the linear growth and quadratic growth parameters were included in the model. Conditional models were constructed using the treatment variable as a level-2 predictor of electricity use. Additional covariates included were age, primary heating fuel, square feet, water heater fuel, number of household members, dwelling type, ownership status, disability status, sex, race, unworked income, and payment method. The model converged with all covariates included, so all variables were retained.

The estimated mean intercept indicated that households' electricity consumption was not significantly different from zero, $\beta_{00} = 592.72$, $p = .211$. The mean growth rate for electricity use in kWh was also not significant, $\beta_{10} = 101.45$, $p = .328$. This means that the average electricity use increased by an average of 101.45 units per month over the course of 12 months. The results were not significant indicating that mean electricity use did not significantly differ across households and the mean growth rate of electricity use did not change in a given month holding all other variables constant. The mean acceleration was also not significant, $\beta_{20} = -8.09$, $p < .187$ which accounts for the rate of change in electricity use that varies at each time point. So, on average, households' electricity use varied in a negative direction at each time point.

At initial status, the number of household members was a significant predictor of electricity use, $\beta_{06} = 34.35$, $p < .048$. Since age, number of household members, and square footage were continuous covariates they were centered around the grand mean. A one unit change in the covariate is the predicted electricity use for an average number of household members of the sample. So, a one unit increase in the number of household members from the mean of 3.0 household members resulted in an increase of 34.35 units in electricity use. For the linear growth rate, the treatment variable was a significant predictor of electricity use, $\beta_{11} = 36.29$, $p = .035$. So, the treatment group experienced a reduction of 36.29 units in comparison to the control group. In addition, type of dwelling was reported as a statistically significant predictor of the households' electricity consumption growth rate, $\beta_{17} = 37.95$, $p = .006$. Households who were classified as a house/modular home had an increase of 37.95 units in electricity use in comparison to

non - modular homes. Unworked income was reported as a statistically significant predictor of households' electricity consumption growth rate, $\beta_{17} = 29.91$, $p = .032$. Those applicants who had indicated a source of unworked income (other than LIHEAP benefits) were associated with a reduction of 29.91 units in comparison to those who indicated they do not receive any kind of public assistance. Finally, the type of dwelling was reported as a statistically significant predictor of households' electricity consumption acceleration rate, $\beta_{27} = -2.92$, $p < .001$. Thus, the rate at which households' use varied was in a negative direction for those residing in a modular home in comparison to those who resided in a non - modular home. All other variables were not significant.

The chi-square test on level-2 residual variance of the intercept was 5194.36 ($df = 759$, $p < .001$). The corresponding chi-square test on the residual variance for growth rate was, $\chi^2 = 2906.24$ ($df = 759$, $p < .001$). The chi-square tests indicate that there was statistically significant between group variance in the intercept and slope parameters across groups. The intercept term represents the between household variance in electricity use after controlling for the impact of treatment, age, primary heating fuel, square feet, water heater fuel, number of household members, dwelling type, ownership status, disability status, sex, race, unworked income, and payment method. For a summary of fixed and random effects of the linear component of the conditional model see Table 12. A summary of fixed effects for the quadratic component of the conditional model is provided in Table 13.

Table 12

Summary of Fixed and Random Effects for the Linear Conditional Model (Whole Sample)

Fixed effects	**Coefficient**	**SE**	***T***	***p***
Mean electricity use, β_{00}	592.72	473.11	1.25	0.211
Treatment, β_{01}	60.05	66.87	0.90	0.369
Age, β_{02}	1.11	1.56	0.71	0.476
Primary heating fuel, β_{03}	-151.45	237.80	-0.64	0.524
Square feet, β_{04}	0.03	0.04	0.89	0.374
Water heater fuel, β_{05}	391.39	234.83	1.67	0.096
Number of household members, β_{06}	34.35	17.35	1.98	0.048
Dwelling type, β_{07}	9.24	50.19	0.18	0.854
Ownership status, β_{08}	6.21	46.40	0.13	0.894
Disability status, β_{09}	34.09	42.91	0.79	0.427
Sex, β_{010}	-75.07	58.72	-1.28	0.201
Hispanic, β_{011}	-13.35	45.76	-0.29	0.771
Black, β_{012}	-24.55	54.62	-0.45	0.653
Unworked income, β_{013}	91.30	49.51	1.84	0.066
Payment method, β_{014}	-380.68	454.72	-0.84	0.403
Mean growth rate, β_{10}	101.45	103.74	0.98	0.328
Treatment, β_{11}	-36.29	17.14	-2.12	0.035
Age, β_{12}	-0.34	0.46	-0.73	0.466
Primary heating fuel, β_{13}	-10.59	56.49	-0.19	0.851
Square feet, β_{14}	-0.0007	0.009	-0.08	0.938
Water heater fuel, β_{15}	-125.52	66.87	-1.88	0.061
Number of household members, β_{16}	5.05	5.04	1.00	0.316
Dwelling type, β_{17}	37.95	13.66	2.78	0.006
Ownership status, β_{18}	3.67	13.58	0.27	0.787
Disability status, β_{19}	0.89	12.97	0.07	0.946
Sex, β_{110}	19.46	16.83	1.16	0.248
Hispanic, β_{111}	-2.04	13.47	-0.15	0.879
Black, β_{112}	28.99	16.69	1.74	0.083
Unworked income, β_{113}	-29.91	13.94	-2.14	0.032
Payment method, β_{114}	75.24	85.84	0.88	0.381

Random Effects	**Variance Component**	***d.f.***	χ^2	***P***

Initial status, r_{0i}	161346.94	759	5194.36	<0.001
Growth rate, r_{1i}	1126.41	759	2906.24	<0.001
Level-1, e_{ti}	48236.92			
Deviance = 110845.48 with 4 df				

Table 13

A Summary of Fixed Effects for the Quadratic Component of the Conditional Model (Whole Sample)

Fixed effects	**Coefficient**	**SE**	***T***	***P***
Mean acceleration rate, β_{20}	-8.09	6.13	-1.32	0.187
Treatment, β_{21}	2.92	1.09	2.68	0.007
Age, β_{22}	0.03	0.03	1.04	0.299
Primary heating fuel, β_{23}	1.52	3.38	0.45	0.653
Square feet, β_{24}	-0.00	0.00	-0.11	0.910
Water heater fuel, β_{25}	6.93	4.17	1.66	0.096
Number of household members, β_{26}	-0.26	0.32	-0.80	0.426
Dwelling type, β_{27}	-2.92	0.84	-3.49	<0.001
Ownership status, β_{28}	-0.31	0.86	-0.36	0.718
Disability status, β_{29}	-0.29	0.81	-0.36	0.721
Sex, β_{210}	-0.89	0.99	-0.90	0.369
Hispanic, β_{211}	0.26	0.87	0.30	0.746
Black, β_{212}	-1.76	1.04	-1.70	0.090
Unworked income, β_{213}	1.54	0.89	1.72	0.085
Payment method, β_{214}	-3.38	4.81	-0.70	0.482

The proportion of variance explained (PVE) by treatment and covariates indicates the amount of variance in electricity consumption at intercept and for the growth parameter. The estimated PVE by the listed above covariates was .22 for initial status. This indicates that proportion of variance calculation resulted in 22% of variance in electricity use at initial status accounted for by treatment, age, primary heating fuel, square feet, water heater fuel, number of household members, dwelling type, ownership status, disability status, sex, race, unworked income, and payment method. This indicates

that 78% of the variance remained unexplained in the households' differences in electricity use at initial status. The estimated PVE explained by the covariates listed above for the linear growth rate was 0.95. So, 95% of the variance in the slope of the mean linear growth rate of electricity use was explained by treatment, age, primary heating fuel, square feet, water heater fuel, number of household members, dwelling type, ownership status, disability status, sex, race, unworked income, and payment method. So, only 5% of the variance in the electricity use growth rate remained unexplained.

Chapter 4: Discussion

Synopsis of the Evaluation

The primary purpose of this evaluation was to determine if the change in electricity use differed for households who had not installed upgrade(s) with households who had installed upgrade(s) using a propensity score matching method and hierarchical linear modeling. A related objective focused on assessing the impact of household - level covariates on electricity use. Finally, this evaluation illuminated the application of propensity score matching and hierarchical linear modeling to longitudinal data in the field of energy efficiency and contributed to research methods, statistics, and evaluation practice.

Quantitative Findings

Propensity score analysis with balanced risk matching was implemented with two matching approaches: nearest neighbor 1:1 and nearest neighbor with caliper. The nearest neighbor 1:1 was implemented to balance covariates among households across the control and treatment groups starting in June of 2013 through March of 2014. Due to a suboptimal performance of the nearest neighbor 1:1 algorithm in April and May of 2014, nearest neighbor with caliper algorithm was used for these months. Propensity score matching was performed at each of the 12 months under analysis and aggregated to form a subsample of matched households on a set of covariates. Then, a hierarchical

linear model was performed on the subsample to determine the effect of treatment on households' electricity use. In addition, a hierarchical linear model was used with the entire sample to compare the results from models performed on a propensity score - based sample to detect any differences. The entire sample consisted of 7897 observations and 813 households.

The focus of this evaluation concerned the examination of the following two questions: 1) Do low-income residential households experience a statistically significant change in electricity use over the course of a year following participation in LIWP in Denver County, CO by implementing propensity score matching after controlling for covariates, such as sex, age, primary heating fuel, square footage of household, water heater fuel type, number of household members, type of household, status of household, presence of a disability, race, unworked income, and method of payment? Hypothesis 1 stated that households that participate in LIWP experience a statistically significant decline in their electricity use in kWh following the installation of upgrade(s) after controlling for the covariates listed above. The null hypothesis was retained as the results showed that households that participated in LIWP did not experience a statistically significant decline in their electricity use in kWh following the installation of upgrade(s) after controlling for the above covariates.

2) How do the results using a subsample of households matched on propensity scores and using the entire data (without propensity score matching) set compare in terms of a statistically significant change in electricity consumption over the course of a year using a hierarchical linear growth model after controlling for covariates, such as sex, age,

primary heating fuel, square footage of household, water heater fuel type, number of household members, type of household, status of household, presence of a disability, race, unworked income, and method of payment? Hypothesis 2 stated that a subsample of households that are matched on a propensity score experience a greater statistically significant decline in electricity consumption as compared to the entire data set following the installation of upgrade(s) after controlling for the above covariates. The null hypothesis was retained as the results from the propensity score - based sample did not indicate a statistically significant change in electricity consumption.

Interestingly, the results from the hierarchical linear growth model with the entire sample produced different results when compared to the propensity score - based model. The results differed in terms of treatment and covariate impacts on electricity use on the intercept and growth parameters. Thus, the covariates impacted the estimation of treatment effects between the control and treatment conditions. This provides evidence in support of balancing treatment conditions before performing outcome analyses and comparing effects between groups (Rubin & Rosenbaum, 1983; Rosenbaum, 2002). A review of the literature indicated that systematic comparisons of the different strategies to apply propensity score analysis with respect to validity and with specific attention to exclusion of participants are limited (Stürmer *et al*., 2005). Further, excluding a large proportion of treated subjects because of a lack of untreated matches may severely alter the composition of the study population (Stürmer *et al*., 2006). This means that the two samples could be considered inherently different and may not be best suited for the purposes of comparison. An assessment of balance on the covariates was performed and

improved the standardized differences between the treatment groups. Finally, a power analysis indicated sample values far below those obtained from the matched sample were needed to detect an effect of treatment given the presence of a true effect. Based on these observations, the results from the propensity score - based sample are considered accurate and valid.

Even though the results from the entire sample are different from the results obtained from the propensity score - based sample they are still focal to understanding differences between the two sets of results. The results from the entire sample indicated a statistically significant effect of treatment. After taking account of the mean rate of acceleration, the mean electricity consumption was 33.37 kWh per month lower for the treatment group in comparison to the control group. At the current electricity rate of $0.05461 (Xcel, 2017), the mean savings of the group who had performed the upgrade(s) was $1.82 per month. This rate excludes service and facility charges. Thus, over the course of 12 months, approximately $21.84 was saved on electricity. This finding presents a logical conclusion given that upgrade(s) meant households could have received a varying number of upgrades ranging from installing CFL bulbs to professional air sealing. This means that the range of potential savings on electricity use varied dramatically and was dependent on the number and type of upgrades installed.

For the model that was estimated for the propensity score - based subsample, considering the growth parameters, the effect of treatment was not significant on mean electricity consumption. The only predictor that made a statistically significant impact on electricity use was applicant's sex at initial status. Interestingly, for the propensity score-

based sample up to 96% of the variance in electricity use was explained by the addition of the covariates.

Further, the results from the entire sample could have been influenced by selection bias in the data. Selection bias is referred to as the selection of individuals or groups for analysis where non - randomization is not achieved, thereby ensuring that the sample obtained is not representative of the population to be analyzed (Guo & Fraser, 2015). Selection bias leads to confounding variables that have an impact on both the treatment and outcome. The presence of confounding variables can influence analysis results by increasing Type I error rate which is the incorrect rejection of a true null hypothesis (Rosenbaum, 2002). Thus, the influence of confounding variables can falsely attribute the effect of treatment to the intervention. I believe that selection bias may have impacted the hierarchical linear growth performed on the entire sample and may have biased the results, thus, incorrectly attributing the effect of treatment to the intervention. For these reasons, I infer that the results from the propensity score - based sample are considered accurate and valid.

Relevance of this Evaluation

Even though findings based on the propensity score matched sample did not detect statistically significant differences in electricity use, the findings are relevant for several reasons. First, most electric upgrades that were evaluated comprised minor upgrades that affect home electricity consumption. The upgrades included activities such as installing compact fluorescent light bulbs (CFL) and replacing a shower head. These

types of upgrades are considered minor and, typically, do not result in large energy savings for participants. For example, installing a CFL bulb yields an average savings of $3 per year and replacing a 3+ gpm shower head with a 1.6 gpm or lower results in savings ranging from $15 to $30. The criteria for upgrade designations was acquired from a local non-profit, Groundwork Denver, that works to improve the environment and public health.

Second, the findings from this evaluation, while unexpected, were not entirely surprising. National evaluation of WAP found large disparities in energy savings between local agencies. It was reported that some agencies achieved savings of 30 to 40% of pre-weatherization consumption, while others produced no measurable savings (Tonn *et al*., 2014). So, there was a lot of variation in the way in which agencies approach weatherization services, the resources that are made available to them, and their evaluation approaches to inform future program decisions. In fact, past research found that low - income weatherization programs were twice as costly, per unit of electricity saved, as the average utility efficiency program (Schweitzer *et al.*, 2003). It appears that potentially characteristics of the low - income population have an impact on their use. Since data on employment were not provided for this evaluation, the proportion of individuals who were unemployed was unknown. It is possible that if a substantial portion of participants in the study were unemployed and they consented to installing upgrade(s) any reductions in their electricity consumption would be offset by the fact that they remain at their home and continuously use appliances throughout the day. This is in

comparison to individuals who are employed, typically are not home during the day, and would be expected to have greater reductions in electricity use.

Third, to date, most of the national and state evaluations of weatherization programs used data derived from pre - defined algorithms and specialized software. Most of the evaluations were conducted using specialized software that used estimates of energy data and computed savings based on pre - defined models. This indicates that in many cases, the models analyzed under the "one size fits all" mentality may not have addressed the nuances specific to programs being analyzed. Furthermore, the methods behind these software models are often not explained in detail. This may pose difficulties for researchers attempting to replicate results of the studies. Since the impact of programs is dependent on a broad array of factors, capturing as many of the variables affecting electricity consumption for a particular program under investigation is crucial to obtain relevant and timely analyses. Thus, this evaluation contributed to the fields of evaluation practice and energy efficiency because it used *actual* electricity data to perform the analyses.

Fourth, this evaluation is a contribution to the field of Research Methods, Statistics and Evaluation because of its application of both a propensity score analysis and hierarchical linear growth modeling to longitudinal monthly electricity data. To date, there are no analyses based on a propensity score model or a hierarchical linear growth model to estimate the effect of treatment of a weatherization program for a low - income population. Specifically, ways and techniques to minimize selection bias were discussed.

Fifth, despite a non-significant change in electricity use, the emphasis of this evaluation was not to undermine the effectiveness of installing weatherization upgrade(s) in homes. Besides a decline in electricity use there are many other benefits of weatherization upgrades that improve the lives and public health of a low-income population. Low-income programs are not solely designed to be cost effective, but also aim to help low-income individuals pay for heating costs and improve their quality of life.

Limitations

The present evaluation has several limitations. Since secondary data were used in this evaluation, the quality of data collected and accuracy of data reported were not assessed. Many of the issues encountered as part of the data cleaning process were resolved. This included removing duplicate entries, merging data, and assumption checking. Nevertheless, the data used for this evaluation were accessed through extensive collaboration with multiple stakeholders on this project. The data were acquired through data sharing agreements and permissions to use data for research and evaluation purposes. The long process of obtaining access to the data for this project illuminated the need for more seamless cooperation and communication between the leading actors. Another limitation concerned the outcome being modelled using only data collected by the collaborating entities. Different specifications for the time period of analysis may have resulted in different models, and thereby produce different results. Since both LIHEAP and LIWP are voluntary programs, not all individuals who qualify for the

benefits apply for these programs. So, the data used were collected from participants who applied for and enrolled in the two programs.

A limitation pertaining to weather patterns concerned the lack of outdoor air temperature (OAT) measurements for this analysis. Since propensity score matching was performed at each of the 12 months under analysis, the inclusion of OAT did not seem necessary as households had identical OAT values over the course of a month in which electricity use measurements were recorded. However, the analysis of the hierarchical linear model with the entire sample may have been influenced by temperature variations in the data given the exclusion of OAT.

Another limitation concerned the generalizability of this evaluation's findings. This study only examined Xcel- sponsored weatherization efforts in the City and County of Denver and these results don't necessarily apply to all types of low-income residential efficiency programs. State weatherization programs can vary from region to region so what happens in one state may not translate to another state.

Another limitation concerned the size of samples at each month of propensity score matching analysis. Even though the overall sample size was sufficient for power and quantitative analyses, the sample size differed from month to month. For example, the beginning of the study was associated with sample sizes of matching pairs under 100 observations while in the months completing the study the sample sizes ranged in the 300's and up.

Suggestions for Data Collection and Reporting

In the process of assembling data sets for the analyses I encountered errors and inconsistencies in the data. For example, there were duplicate entries recorded for each household. The format of key variables varied across the data sets. The process of getting data ready for analysis is summarized in a chart in Appendix B. To make data assembly take less time and alleviate some of the burdens associated with compiling data sources, it is suggested to store data in Access files. This would allow for easier manipulation of data and merging of other data necessary to perform the analyses. To expedite the process of data assembly at the city level, it is advisable to create a data inventory to track the data that are available for analysis including time points and relevant variables and identify any additional data that need to be obtained from other sources to facilitate getting access to these data for evaluation purposes.

Recommendations for Future Research

The findings from this evaluation contribute to the broader fields of evaluation practice at the city and county level and energy efficiency. The results from this evaluation spurred many questions that aim to elucidate and expand upon the present findings. The following types of questions may be pursued because of this research: What specific electric upgrades contributed the most to a reduction in electricity use for a low-income population? What factors influenced a statistically non-significant decline of electricity consumption? Would conducting qualitative analyses involving interviews of participants shed more light on the evaluation's findings? What upgrade(s) should the

City and County of Denver invest more financial resources towards to get the greatest amount of energy savings? This study illustrated that including the type and number of upgrades as covariates in the models is suggested for future studies assessing the impact of weatherization assistance programs. Other ideas include conducting a small - scale randomized control experiment to see if results differ from propensity score-based models. Given a sufficient timeline and planning, participants could be put on a wait list for upgrades, thus allowing for outcome data to be collected prior to the installation of upgrades.

Studies involving qualitative analyses are especially suggested due to the critical role of participant behavior in energy consumption. In addition, potential explanations concerning the rebound effect could be explored. The rebound effect refers to the forecasted reduction in energy use due to a set of energy efficiency upgrades that is influenced by consumer and market responses (Gillingham *et al*., 2014). So, following energy efficiency measures that are aimed to reduce energy consumption, consumers and/or the market respond by increasing use or increasing prices. Another avenue for examination focuses on the contractors performing the upgrades and inquiring why there are few differences in electricity consumption between the treatment groups. This speaks to the contractors' direct access to participants and their willingness to ask participants about potential reasons why their electricity use did not change as result of the upgrade(s). Were there any significant changes in behaviors within the household towards electricity consumption? Did the number of occupants change after the upgrade(s) were installed? The results have informed the direction of future research and

focus areas at the local level for analyzing energy efficiency programs for low-income communities.

Conclusion

The results of this evaluation are particularly useful for local community and state leaders who guide policy agendas and monitor program effectiveness. This study illustrated the extent of savings that could be possible from the installation of electricity upgrades as part of the weatherization assistance program. There were important lessons that were learned from analyzing electricity consumption of low-income residential households. The analysis of household patterns of electricity consumption examined in this study allowed for assessing the impact of covariates on the outcome and longitudinally. This study is unique in its design and method and is well situated to further research in the fields of energy efficiency and evaluation practices and methods. This study allowed for comparisons between the propensity score - based sample and the whole data set and reconciled some of the differences between the two sets of results.

The use of propensity score analysis matching methods enabled comparisons across the treatment groups that were matched on a set of covariates related to the outcome. This facilitated a more accurate assessment of the effects of treatment on electricity use. Efforts were made by the researcher to control as many confounding variables as possible. This was accomplished by including as many covariates as possible that were associated with the outcome. Subsequently hierarchical linear growth models allowed for analyses of households compared to each other and to the overall

group. Trajectories of electricity consumption were modelled to determine the shape that best fit the data under investigation. Knowledge of electricity consumption patterns for households and at group level provided information about the non - effectiveness of the intervention on the outcomes.

References

Abrahamse, W., Steg, L., Vlek, C., & Rothengatter, T. (2005). A review of intervention studies aimed at household energy conservation. *Journal of Environmental Psychology*, *25*(3), 273–291. doi:10.1016/j.jenvp.2005.08.002

Allcott, H., & Mullainathan, S. (2012). *External validity and partner selection bias*. (NBER Working Paper No. 18373). Cambridge, MA: National Bureau of Economic Research.

Algina, J., & Swaminathan H. (2011). Centering in two-level nested designs. In J. Hox & J. K. Roberts (Eds.), *The handbook of advanced multilevel analysis* (pp. 285-312). New York: Routledge.

Altan, H. (2010). Energy efficiency interventions in UK higher education institutions. *Energy Policy*, *38*(12), 7722–7731. doi:10.1016/j.enpol.2010.08.024

American Psychological Association. (2016). *Literature review guidelines*. Retrieved from http://www.apa.org/pubs/journals/men/literature-review-guidelines.aspx

Brandon, G., & Lewis, A. (1999). Reducing household energy consumption: a qualitative and quantitative field study. *Journal of Environmental Psychology*, *19*(1), 75–85. doi:10.1006/jevp.1998.0105

Brown, M. A., & Berry, L. G. (1995). Determinants of program effectiveness: Results of the national weatherization evaluation. *Energy*, *20*(8), 729–743. doi:10.1016/0360-5442(95)00027-E

Campbell, D. T. (1957). Factors relevant to the validity of experiments in social settings. *Psychological Bulletin*, *54*(4), 297-312. doi: http://dx.doi.org/10.1037/h0040950

Carnevale, A., & Rose, S. (2001). Low earners: Who are they? Do they have a way out? In M. Miller (Ed.), *Low-Wage Workers in the New Economy* (pp. 45-66). Washington D.C.: The Urban Institute Press.

Cook, T. D., & Campbell, D. T. (1979). *Quasi-experimentation: Design and analysis for field settings*. Chicago: Rand McNally.

Davis, M. (2011). *Behavior and energy savings evidence from a series of experimental interventions*. Retrieved from Environmental Defense Fund: http://blogs.edf.org/energyexchange/files/2011/05/BehaviorAndEnergySavings.pdf

Delmas, M. A., Fischlein, M., & Asensio, O. I. (2013). Information strategies and energy conservation behavior: A meta-analysis of experimental studies from 1975 to 2012. *Energy Policy*, *61*, 729–739. doi:10.1016/j.enpol.2013.05.109

Eisenberg, J. (2014). *Weatherization assistance program technical memorandum background data and statistics on low-income energy use and burdens* (Report No. ORNL/TM-2014/133). Oak Ridge, Tennessee.

Fitzpatrick, J., Sanders, J., & Worthen, B. (1997). *Program Evaluation: Alternative Approaches and Practical Guidelines* (4th ed). N.J.: Pearson Education, Inc.

Geller, E. (1981). Evaluating energy conservation programs: Is verbal report enough? *The Journal of Consumer Research*, *8*(3), 331-335. doi: 10.1086/208872

Gillingham, K., Rapson, D., Wagner, G. (2014). The rebound effect and energy efficiency policy. *Review of Environmental Economics and Policy*, *10*(11), 68-88. doi: 10.1093/reep/rev017

Grønhøj, A., & Thøgersen, J. (2011). Feedback on household electricity consumption: learning and social influence processes. *International Journal of Consumer Studies*, *35*(2), 138–145. doi:10.1111/j.1470-6431.2010.00967.x

Guo, X. S., & Fraser, M. W. (2015). *Propensity score analysis: Statistical methods and applications* (2nd ed.). Thousand Oaks, CA: Sage Publications, Inc.

Harrell, F.E. (2016). Package "Hmisc". Retrieved from http://biostat.mc.vanderbilt.edu/Hmisc

Hayes, S. C., & Cone, J. D. (1981). Reduction of residential consumption of electricity through simple monthly feedback. *Journal of Applied Behavior Analysis*, *14*(1), 81–88. http://doi.org/10.1901/jaba.1981.14-81

Hesser, H. (2015). Modeling individual differences in randomized experiments using growth models: Recommendations for design, statistical analysis and reporting of results of internet interventions. *Internet Interventions*, *2*(2), 110–120. http://doi.org/10.1016/j.invent.2015.02.003

Ho, D. E., Imai, K., King, G., & Stuart, E. A. (2011). MatchIt: Nonparametric preprocessing for parametric causal inference. *Journal of Statistical Software, 42*(8), 1-28.

Hsu, D. (2015). Identifying key variables and interactions in statistical models of building energy consumption using regularization. *Energy*, *83*, 144–155. http://doi.org/10.1016/j.energy.2015.02.008

Imbens, G.W. & Woolridge, J.M. (2009). Recent developments in the econometrics of program evaluation. *Journal of Economic Literature*, *47*(10), 5-86. doi: 10.1257/jel.47.1.5

Johnson, B. & Christensen, L. (2012). *Educational research: Quantitative and qualitative approaches*. Needham Heights, MA, US: Allyn & Bacon Educational research.

Judson, E. P., & Maller, C. (2014). Housing renovations and energy efficiency: insights from homeowners' practices. *Building Research & Information*, *42*(4), 501–511. http://doi.org/10.1080/09613218.2014.894808

Kaiser, M. J. & Pulsipher, A. G. (2003). LIHEAP reconsidered. *Energy Policy*, *31*(14), 1441–1458. doi:10.1016/S0301-4215(02)00200-8

Kua, H. W., & Wong, S. E. (2012). Lessons for integrated household energy conservation policies from an intervention study in Singapore. *Energy Policy*, *47*, 49–56. doi:10.1016/j.enpol.2012.04.009

Li, Y. (1999). Optimal balanced matching (Doctoral dissertation). Retrieved from ProQuest Dissertations and Theses. (Accession Order No. 9953558).

Li, Y. P., Propert, K. J., Rosenbaum P. R. (2001). Balanced risk set matching. *Journal of American Statistics Association, 96*, 870–882.

Metz, B., Davidson, O.R., Bosch, P.R., Dave, R. & Meyer, L.A. (2007). *Climate Change 2007: Mitigation. Contribution of Working Group III to the Fourth Assessment Report*. Cambridge University Press. Retrieved from https://www.ipcc.ch/publications_and_data/publications_ipcc_fourth_assessment_report_wg3_report_mitigation_of_climate_change.htm

Murray, A. G., & Mills, B. F. (2014). The impact of low income energy assistance program participation on household energy insecurity. *Contemporary Economic Policy*, *32*(4), 811–825. doi:10.1111/coep.12050

Olmos, A., & Govindasamy, P. (2015). Propensity Scores: A Practical Introduction Using R. *Journal of Multi-Disciplinary Evaluation, 11*(25), 68-88. Retrieved from http://journals.sfu.ca/jmde/index.php/jmde_1/article/view/431

Raudenbush, S.W. & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods*. London, England: Sage Publications, Inc.

Rosenbaum, P.R. (2002). *Observational Studies*. NY: Springer.

Rosenbaum, P. R. & Rubin, D.B. (1983). The central role of propensity score in observational studies for causal effects. *Biometrica, 70*(1), 41-55.

Rosenbaum, P. R. & Rubin, D.B. (1985). Constructing a control group using multivariate matched sampling methods that incorporate the propensity score. *The American Statistician*, 39, 33-38.

Rubin, D. B. (2008). For objective causal inference, design trumps analysis. *The Annals of Applied Statistics*, *2* (3), 808-840. doi:10.1214/08-AOAS187

Schermelleh-Engel, K., Moosbrugger, H., & Muller, H. (2003). Evaluating the fit of structural equation models: Tests of significance and descriptive goodness-of-fit

measures. *Methods of Psychological Research Online*, *8* (2), 23–74. Available from http://www.dgps.de/ fachgruppen/methoden/mpr-online/

Schweitzer, M., Jones, D. W., Berry, L. G., & Tonn, B. E. (2003). *Estimating energy and cost savings and emissions reductions for the state energy program based on enumeration indicators data* (Report No. ORNL/CON-487). Oak Ridge, Tennessee: Oak Ridge National Laboratory.

Singer, J. D., & Willett, J. B. (2003). *Applied longitudinal data analysis: Modeling change and event occurrence.* New York: Oxford University Press, Inc.

Shadish, W. R., Cook, T. D. & Campbell, D. T. (2002). *Experimental and quasi-experimental design for generalized causal inference.* Boston: Houghton Mifflin Company.

Snijders, T. A. B., & Bosker, R. J. (1999). *Multilevel analysis: an introduction to basic and advanced multilevel modeling* (2nd ed.). London, England: Sage Publications, Inc.

Spybrook J., Bloom, H., Congdon, R., Hill, C., Martinez A. & Raudenbush S. (2011). Optimal design plus empirical evidence: documentation for the "optimal design" software. Stern, P. C. (1992). What psychology knows about energy conservation. *American Psychologist*, *47*(10), 1224-1232. http://dx.doi.org/10.1037/0003-066X.47.10.1224

State of Rhode Island, Energy Efficiency and Resource Management Council. (2016). *Annual report*. Retrieved from http://www.rieermc.ri.gov/documents/annual/6_2016%20EERMC%20Annual%20Report.pdf

Stürmer, T., Joshi, M., Glynn, R. J., Avorn, J., Rothman, K. J., & Schneeweiss, S. (2006). A review of the application of propensity score methods yielded increasing used, advantages in specific settings, but not substantially different estimates compared with conventional multivariable methods. *Journal of Clinical Epidemiology*, *59*(5), 437-447. http://dx.doi.org/10.1016/j.jclinepi.2005.07.004

Stürmer, T., Schneeweiss, S., Brookhart, M.A., Rothman, K.J., Avorn, J., and Glynn, R.J. (2005). Analytic strategies to adjust confounding using exposure propensity scores and disease risk scores: nonsteroidal anti-inflammatory drugs and short-term mortality in the elderly. *American Journal of Epidemiology*, *161*(9), 891–898. http://dx.doi.org/10.1093/aje/kwi106

Tonn, B., Carroll, D., Pigg, S., Blasnik, M., Dalhoff, G., Berger, J., & Rose, E. (2014). *Weatherization works – summary of findings from a retrospective evaluation of the U.S. Department of Energy's Weatherization Assistance Program* (Report No. ORNL/TM-2014/338). Oak Ridge, Tennessee: Oak Ridge National Laboratory.

Tonn, B., Schmoyer, R., & Wagner, S. (2003). Weatherizing the homes of low-income home energy assistance program clients: a programmatic assessment. *Energy Policy*, *31*(8), 735–744. doi:10.1016/S0301-4215(02)00124-6

Tyler, R. W., Gagné, R. M., & Scriven, M. (1967). *Perspectives of curriculum evaluation*. Chicago: Rand McNally.

U.S. Department of Energy, Energy Information Administration. (2013). *Annual energy outlook 2013 with projections to 2040* (DOE/EIA-0383). Retrieved from https://www.eia.gov/outlooks/aeo/pdf/0383(2013).pdf

U.S. Department of Energy Office of Project Management Oversight & Assessments. (2016). *Glossary of terms & acronyms.* Retrieved from http://energy.gov/projectmanagement/glossary-terms-acronyms

U.S. Department of State. (2014). *2014 U.S. climate action report to the UN framework Convention on climate change*. Retrieved from http://www.state.gov/e/oes/rls/rpts/car6/index.htm?utm_content=bufferab672

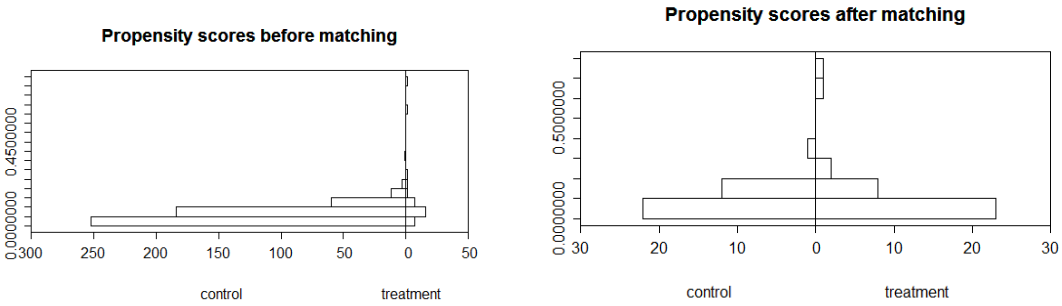
Warren, C. (2003). *Turn on the heat: A program outcome evaluation of the Low-Income Housing Energy Assistance Program (1981-1992)* (Doctoral dissertation). Retrieved from ProQuest Dissertations & Theses Global. (3590583)

Weatherization Assistance for Low-Income Persons, 10 C.F.R. § 440.3. (2001).

Xcel Energy. (2017, January). *Colorado residential electric and natural gas rate schedule summaries*. Retrieved from https://www.xcelenergy.com/staticfiles/xe/Regulatory/COResRates.pdf

Appendix A

June 2013



July 2013

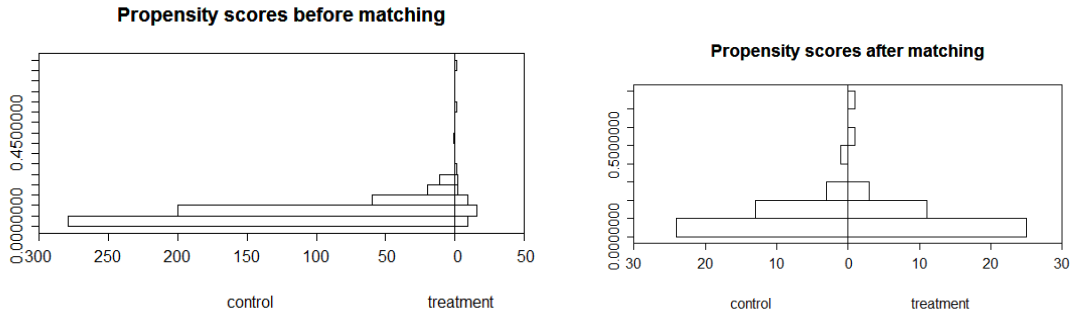


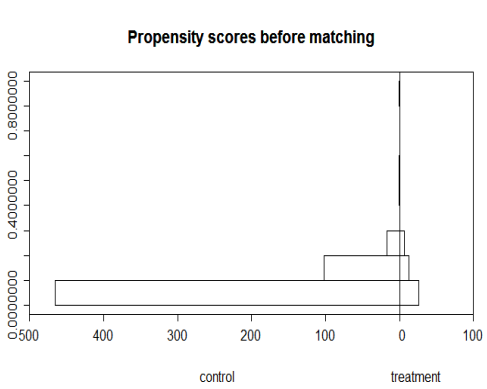
Table A1

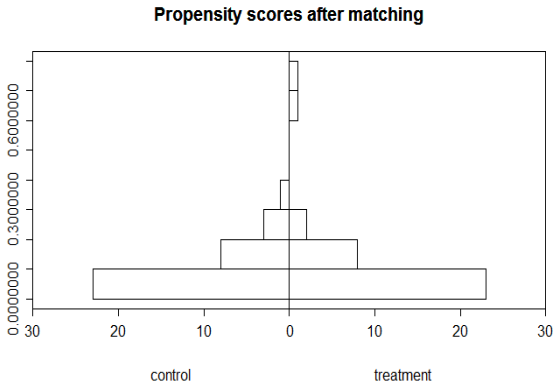
Evaluation of Standardized Differences Pre- and Post- Matching for covariates in August and September of 2013 using Nearest Neighbor 1:1

Time point	August 2013		September 2013	
	Pre – Matching	After NN 1:1	Pre - Matching	After NN 1:1
Age	0.17	-0.09	0.14	-0.14
Primary heating fuel	-0.05	0	0.06	0.12
Square feet	-0.65	0.07	-0.46	-0.05
Water heating fuel	0.01	0	0.04	-0.15

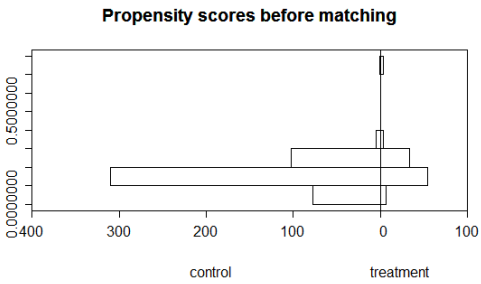
Number of household members	-0.39	0.05	-0.14	0.03
Type of dwelling	-0.09	-0.06	0.20	0.03
Ownership status	-0.25	0	-0.18	0.02
Disability status	0.15	-0.09	0.04	0.07
Sex	-0.24	-0.06	-0.16	-0.05
Hispanic	-0.11	-0.04	0.06	-0.04
Black	0.22	0.05	0.02	-0.03
Unworked income	-0.25	0.10	-0.02	0
Payment method	-0.19	0	-0.16	0

August 2013





September 2013



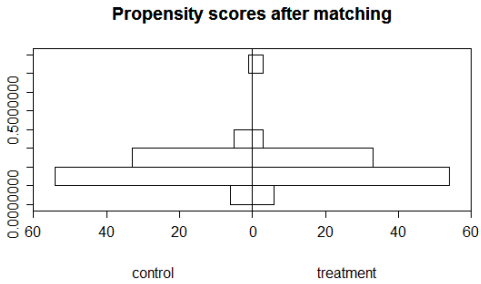
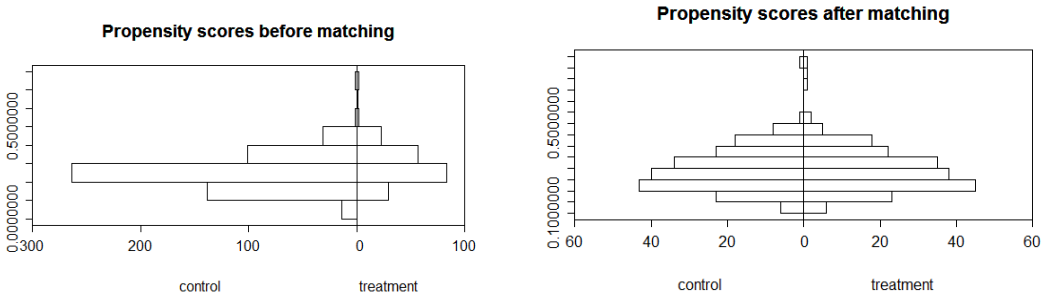


Table A2

Evaluation Standardized Differences Pre- and Post- Matching for covariates in October and November of 2013 using Nearest Neighbor 1:1

Time point	October 2013		November 2013	
	Pre – Matching	After NN 1:1	Pre - Matching	After NN 1:1
Age	0.31	0.05	0.34	0.03
Primary heating fuel	0.05	-0.03	0.03	0.03
Square feet	-0.34	-0.01	-0.32	-0.05
Water heating fuel	0.04	-0.07	0.04	0
Number of household members	-0.14	-0.01	-0.18	-0.05
Type of dwelling	0.15	-0.03	0.01	0.07
Ownership status	-0.26	-0.01	-0.22	0.01
Disability status	0.01	0.06	-0.05	-0.03
Sex	0.01	0.06	-0.05	0
Hispanic	-0.01	0.06	-0.06	-0.02
Black	-0.11	-0.07	-0.11	0.02
Unworked income	-0.06	-0.05	-0.13	-0.03
Payment method	-0.11	0	-0.10	-0.08

October 2013



November 2013

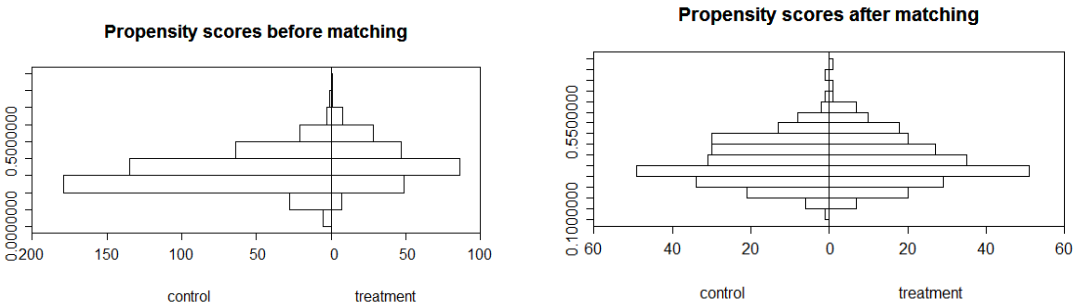
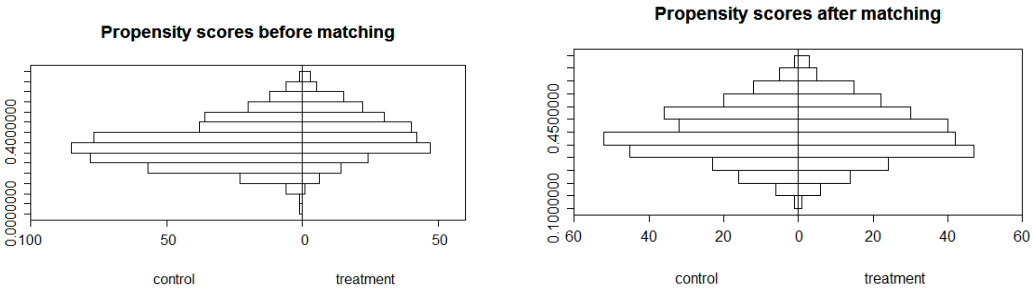


Table A3

Evaluation of Standardized Differences Pre- and Post- Matching for covariates in December of 2013 and January of 2014 using Nearest Neighbor 1:1

Time point	December 2013		January 2014	
	Pre-Matching	After NN 1:1	Pre- Matching	After NN 1:1
Age	0.35	0.01	0.39	0.04
Primary heating fuel	0.08	0.02	0.07	0.02
Square feet	-0.28	-0.04	-0.26	-0.03
Water heating fuel	0.11	0.03	0.05	0.02
Number of household members	-0.15	0.01	-0.20	-0.06
Type of dwelling	-0.03	0	-0.01	0
Ownership status	-0.22	0.01	0.22	-0.05
Disability status	-0.01	-0.09	-0.01	-0.02
Sex	-0.01	-0.02	-0.01	0.08
Hispanic	0.07	-0.02	0.03	-0.06
Black	-0.22	0.04	-0.18	0.04
Unworked income	-0.08	0.03	-0.12	-0.03
Payment method	-0.07	-0.04	-0.07	-0.04

December 2013



January 2014

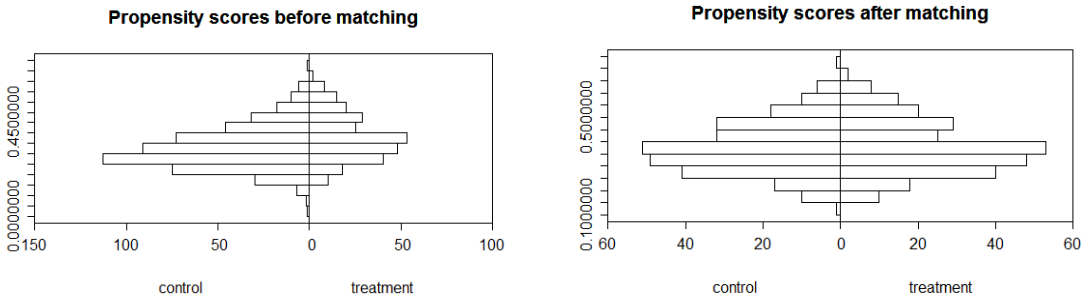


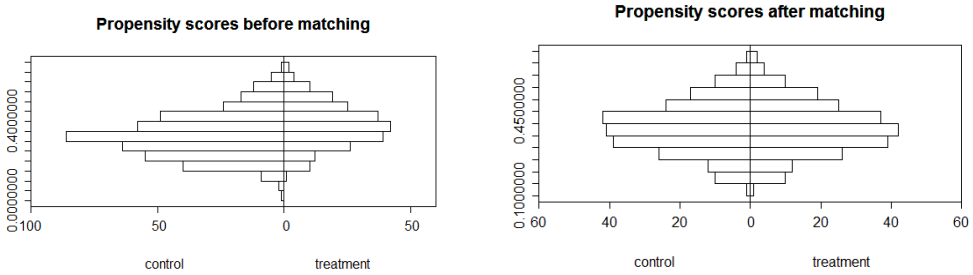
Table A4

Evaluation of Standardized Differences Pre- and Post- Matching for covariates in February and March of 2014 using Nearest Neighbor 1:1

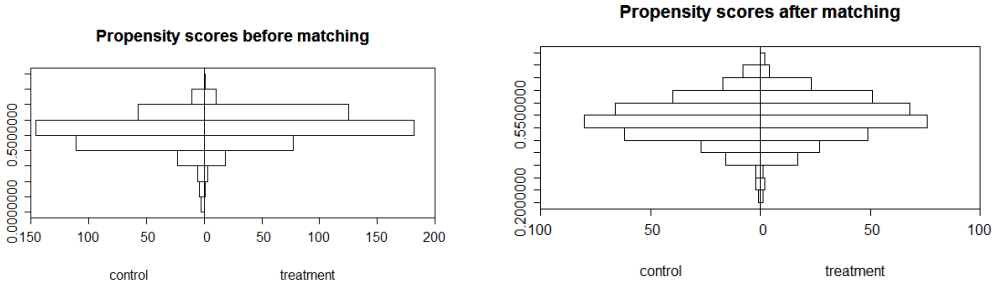
Time point	February 2014		March 2014	
	Pre – Matching	After NN 1:1	Pre - Matching	After NN 1:1
Age	0.37	0.04	0.20	0.07
Primary heating fuel	0.14	0.03	-0.08	-0.03
Square feet	-0.32	-0.04	-0.43	-0.03
Water heating fuel	0.10	0.03	-0.07	0.02
Number of household members	-0.18	-0.05	-0.16	-0.02
Type of dwelling	-0.04	-0.02	0.04	-0.02

Ownership status	-0.18	-0.01	-0.01	0
Disability status	0.03	0	0.04	0.01
Sex	-0.02	0.04	0.06	0.01
Hispanic	0.11	0.03	0.12	0.05
Black	-0.29	0.02	-0.11	-0.03
Unworked income	-0.15	-0.02	-0.17	-0.03
Payment method	-0.04	0	-0.02	0

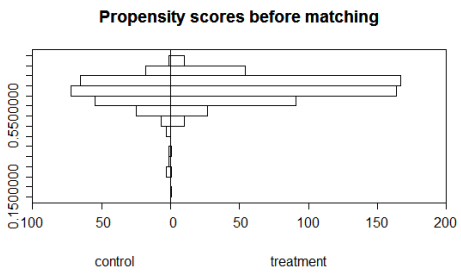
February 2014

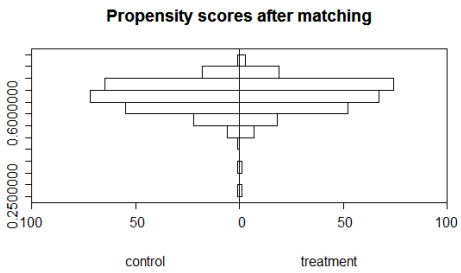


March 2014

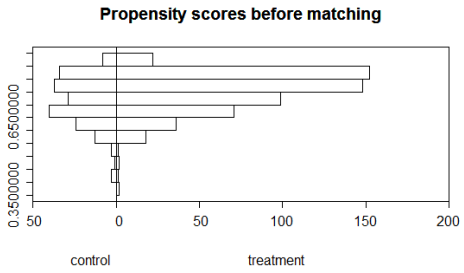


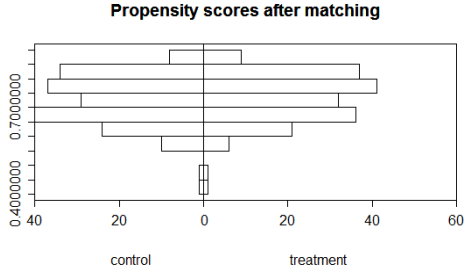
April 2014



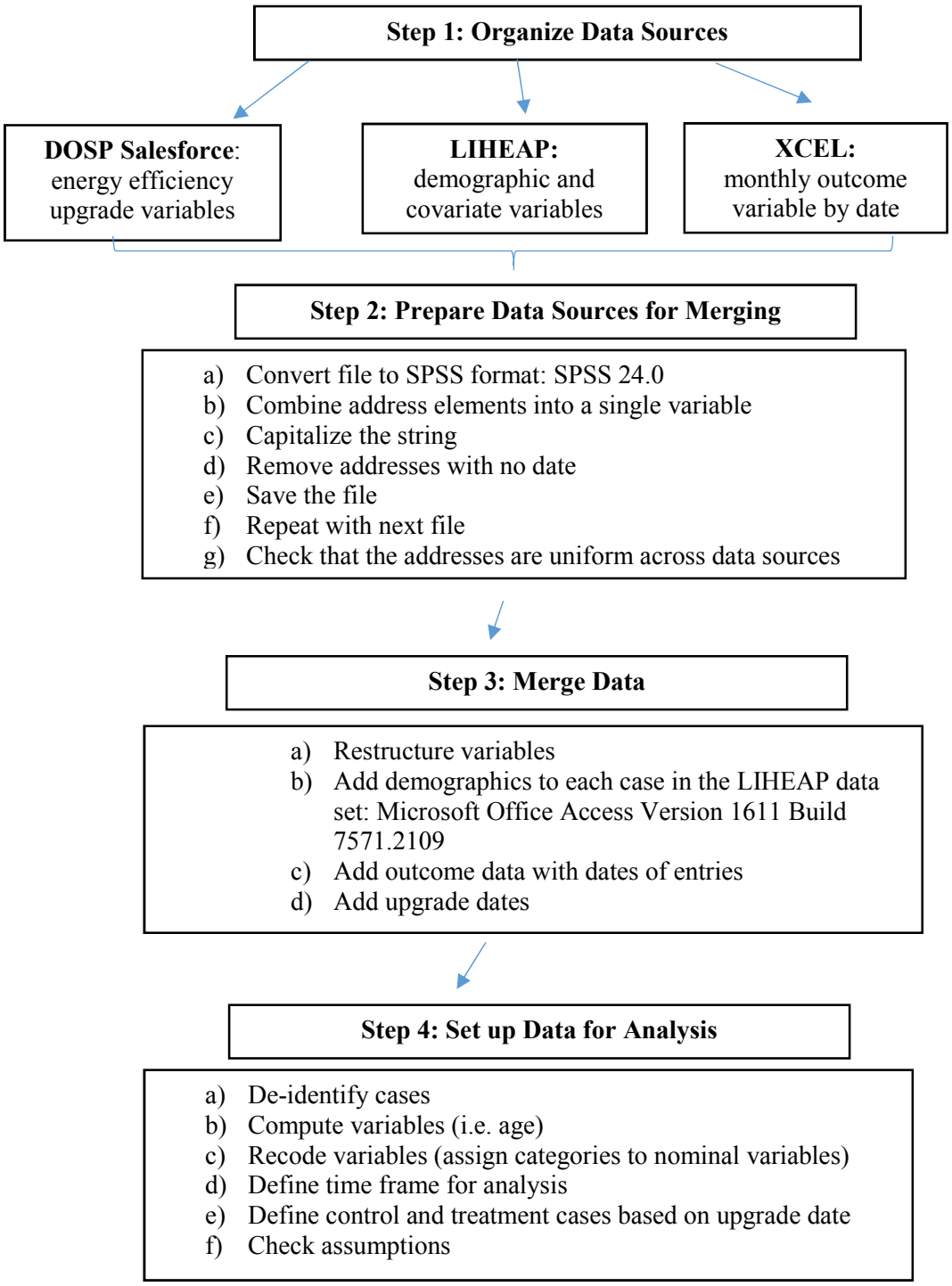


May 2014





Appendix B


Step 1: Organize Data Sources
DOSP Salesforce: energy efficiency upgrade variables
LIHEAP: demographic and covariate variables
XCEL: monthly outcome variable by date
Step 2: Prepare Data Sources for Merging
a) Convert file to SPSS format: SPSS 24.0
b) Combine address elements into a single variable
c) Capitalize the string
d) Remove addresses with no date
e) Save the file
f) Repeat with next file
g) Check that the addresses are uniform across data sources
Step 3: Merge Data
a) Restructure variables
b) Add demographics to each case in the LIHEAP data set: Microsoft Office Access Version 1611 Build 7571.2109
c) Add outcome data with dates of entries
d) Add upgrade dates
Step 4: Set up Data for Analysis
a) De-identify cases
b) Compute variables (i.e. age)
c) Recode variables (assign categories to nominal variables)
d) Define time frame for analysis
e) Define control and treatment cases based on upgrade date
f) Check assumptions