Review of Impact
Evaluation Best Practices
Final Report (R91)

March 30, 2016

Connecticut Energy Efficiency Board

NMR
Group, Inc.

The Cadmus Group, Inc.
An Employee-Owned Company • www.cadmusgroup.com

This page left blank.

Prepared by:
Natalie Bodington
Shannon Donohue
Joseph Iaccarino
Elena Kazarov
Eric Rambo
Scott Reeves
Cadmus

Lisa Wilson-Wright
NMR Group, Inc.

This page left blank.

Table of Contents

List of Tables

List of Figures

Abstract

The purpose of this study (R91) was to survey best practices in impact evaluation, compare methodologies used to estimate savings, and examine the findings of the recently completed Impact Evaluation of the Home Energy Services (HES) and Home Energy Services-Income Eligible (HES-IE) Programs (R16) in light of this review. This best practices review provides an overview of key evaluation protocol and guideline documents and includes an extensive bibliography at the end of this report for reader reference. Due to the large number of customers using fuels such as propane and oil in Connecticut, and the challenges associated with evaluating savings for these fuels, this study also reviewed approaches specific to estimating savings for these delivered fuels. This detailed review of state-of-the-art impact evaluation practices is provided in Section 1 of this report.

In 2014 and 2015, Connecticut conducted an impact evaluation of two flagship residential programs, HES and HES-IE programs (study called R16). The R16 impact evaluation calculated savings and realization rates (i.e., differences between calculated savings based on the planning estimates compared to evaluation results) at a measure level. This study used a multimethod approach to impact evaluation, with methods including billing analysis, building simulation, and engineering algorithms, making R16 an exemplary case study for the R91 best practices review. Among the key findings of this impact evaluation, R16, were divergent realization rates for gas savings attributed to several prominent measures: duct sealing, air sealing, attic insulation, and wall insulation. This report examines how *ex ante* and *ex post* savings calculation approaches may have contributed to these findings (summarized in Table 1 and Table 2). The recommendations in the report suggest updates to the Connecticut Program Savings Documentation (PSD) savings calculation guidelines to better align *ex ante* PSD calculations with the realized savings from the impact evaluation.

Table 1. Gas Realization Rates from R16 and Savings Methodology between PSD and R16.

Type	Measure	Realization Rates (R16)		Method	
		HES	HES-IE	R16	PSD
HVAC	Duct Sealing	42%	16%	Enhanced Simulation	Simple Simulation
Shell	Air Sealing	91%	61%	Billing Analysis	Simple Simulation
Shell	Attic Insulation	76%	129%	Enhanced Simulation	Engineering Algorithm
Shell	Wall Insulation	50%	32%	Enhanced Simulation (HES); Billing Analysis (HES-IE)	Engineering Algorithm

Table 2. Sources of Differences from PSD Formulae vs. R16 Impact Evaluation Results

Sources of Differences from PSD Assumptions and Settings that were Upgraded in R16	Duct Sealing	Air Sealing	Attic Insulation	Wall Insulation
Actual Participant Characteristics		X		
Installation Quality and Measure Persistence		X		
Behavioral and Education Changes		X		X
Weather Data Enhancements		X	X	X
Measure Interactivity	X	X	X	X
HVAC System Efficiency	X	X	X	X
Differentiating Building Types	X		X	X
HDD Adjustment Factor			X	X
Billing Data Calibration	X			
Assumptions and Inputs	X			
Types of Modeling Software	X			
Multiple Model Locations	X			

Executive Summary

The R91 study, conducted by the NMR Group and Cadmus (collectively referred to as the evaluation team), presents a review of best practices in impact evaluation of residential retrofit programs. The study surveys authoritative manuals and protocols for such impact evaluations and presents a detailed overview of the most commonly used methodologies for evaluation of savings. As a case study for this review, R91 examines the R16 impact evaluation carried out by the evaluation team, which employed several different evaluation methodologies in developing both *ex ante* and *ex post* savings values.[1]

In 2014, the Connecticut Energy Efficiency Board (EEB) commissioned the R16 impact evaluation of the Program Year 2011 (PY2011) Home Energy Services (HES) and Home Energy Services-Income Eligible (HES-IE) programs offered by the following Connecticut utilities: Connecticut Light & Power (CL&P), The United Illuminating Company (UI), Connecticut Natural Gas (CNG), Southern Connecticut Gas (SCG), and Yankee Gas Services Company (YGS). This evaluation sought to provide evaluated estimates of energy and demand savings associated with measures installed through these programs.

The R16 impact evaluation calculated savings and realization rates at a measure level using a multimethod approach, including billing analysis, building simulation, and engineering algorithms. The evaluation found that several key measures had divergent realization rates for gas consumption, as shown in Table 3.

Table 3. R16 Impact Evaluation Realization Rates and Savings Estimation Methodologies for Selected Gas Measures

Category	Measure	Realization Rate	
		HES	HES-IE
HVAC	Duct Sealing	42%	16%
Shell	Air Sealing	91%	61%
Shell	Attic Insulation	76%	129%
Shell	Wall Insulation	50%	32%

The R91 best practices study provides an opportunity to better understand key drivers of the differences between evaluation and PSD approaches, with particular attention to best practices in savings estimation methodologies.

This report is framed in two sections, respectively seeking to survey best practices generally and to apply these best practices specifically to the case of the HES and HES-IE PY2011 impact evaluation:

- **Section 1: Best Practices in Impact Evaluation.** The study's best practices review encompasses three topic areas:

---

[1] The Cadmus Group, Inc. and NMR Group, Inc. *Impact Evaluation: Home Energy Services–Income-Eligible and Home Energy Services Programs (R16)*. December 2014. Available online: http://www.energizect.com/your-town/hes-and-hes-ie-impact-evaluation-r16-final-report-12-31-14.

- **Literature Review.** This report provides a literature review that details several commonly referenced manuals and guidelines for impact evaluation. The review synthesizes relevant details from each source. Many of these documents discuss the best practices and applications of evaluation methodologies at length, and readers seeking additional detail are encouraged to consult these sources directly.

- **Methodology-Specific Discussions and Guidelines for Application.** For five common evaluation methodologies, the report presents an overview of each approach including its requirements, limitations, and any emerging applications. The team offers guidance for choosing the most appropriate methodology given different constraints, contrasting each methodology in terms of a range of characteristics and applications.

- **Recommendations for Calculating Oil and Propane Savings.** Because of the particular challenges in calculating savings for bulk fuels such as oil and propane and the prevalence of these fuels in the Northeast, this report provides a thorough review of common practices used to calculate oil and propane savings, offering illustrative case studies and recommendations.

- **Section 2: R16 Case Study—Comparison of Evaluation Approaches.** Following the assessment of common approaches to impact evaluation and their most appropriate applications, the best practices review is applied to the R16 evaluation as a case study. This section includes a discussion of how the approaches used in the R16 evaluation and in the development of savings in the PSD may influence their respective savings calculations. This section provides an overview of the methodologies employed in calculating *ex ante* and *ex post* savings for four measures driving the gas realization rates: attic insulation, wall insulation, duct sealing, and air sealing. Based on this review, the report suggests areas for improvement or further examination.

Section 1: Best Practices in Impact Evaluation

Literature Review

The study's literature review encompassed five commonly-referenced manuals and guidelines that discuss in great detail many of the most common evaluation practices, including their strengths, weaknesses, and best applications. The sources reviewed in full are:

- International Performance Measurement and Verification Protocol (IPMVP)

- Uniform Methods Project (UMP) Protocols

- The State and Local Energy Efficiency Action (SEE Action) Network's Energy Efficiency Program Impact Evaluation Guide

- California Energy Efficiency Evaluation Protocols: Technical, Methodological, and Reporting Requirements for Evaluation Professionals ("Evaluators' Protocols")

- Northwest Power and Conservation Council Regional Technical Forum Roadmap (RTF)

Additional sources are provided in the References section at the end of this report. Readers are encouraged to review these sources directly where additional detail is desired.

Overviews, Constraints, and Development

Following the literature review, the study describes five prevalent approaches to impact evaluation, specifically those most appropriate for residential retrofit programs. Data requirements, constraints, and best applications are noted for each methodology to assist in comparison of each approach's strengths and weaknesses.

Billing Analysis

Billing analysis describes the process by which records of participants' energy usage—typically their utility bills—are compared before and after program participation in order to estimate the savings attributable to program activities. Billing analysis can be used to derive whole-house and, in some cases, measure-specific savings, and reflects participants' behavioral adjustments as well as measure-driven changes in consumption. Weather normalization of customer billing data and use of an appropriate comparison group allow billing analysis to provide high-accuracy results. Billing analysis relies on both utility tracking data and billing data, and requires that the following criteria be met:

- The average reduction in usage must be relatively large (i.e., have a high "signal to noise" ratio) to derive high-precision results through billing analysis.

- Program treatment of the participant group should be relatively consistent in the intensity, type, or magnitude of treatment.

- There must be a sufficiently large sample of participants across which to average consumption data.

- There must be sufficient consumption data available over a long enough period before and after program treatment.

Building Simulation

Building simulations offer a qualified simulation software user the ability to determine the effects of various building retrofit techniques. Building simulations are most appropriately used to determine the energy impacts of weather sensitive measures. A building simulation can be described as a large set of engineering calculations.

Using building simulation, a modeler will either develop a simulation to replicate the conditions observed in a particular building or develop one or more prototypes representative of a population. The outputs of these models—energy usage at varying levels of temporal granularity and specificity of end-use—can then be employed to determine savings for a facility, a measure, or a program. Building simulations allow interactions between different measures to be considered when calculating energy usage patterns, although they rely on numerous modeler assumptions and approximations. By calibrating simulation models to a set of participant billing data, inputs and assumptions can be adapted to provide a relatively accurate representation of the participant or population under consideration.

Equipment Metering

Residential evaluation metering studies are typically reserved for technology-specific energy efficiency programs. Challenges arise when using metering in whole-house program evaluation because envelope, thermostat, and HVAC measures are highly interactive and metering savings at a measure-level may not capture these effects, or may not be able to attribute them to a specific measure. The distribution of measures in the sample of metered sites would have to statistically match the distribution of measures in the program population for whole-house metering to be an appropriate approach. This is typically a logistically demanding and expensive task because metering samples are more difficult to collect than samples that don't require site visits, such as billing samples.

Engineering Algorithms

In select cases in which a measure implemented through an energy-efficiency program is well understood and has minimal interactive effects, an algorithmic approach may appropriately capture the savings derived from installation of this measure. Algorithms based on engineering principles typically employ site-specific data, including details of the measure installed (e.g., quantity), and assumptions about the home, measure, or other interactions occurring. An algorithmic approach is often the least time-intensive method of calculating savings for specific measures, although it is rarely appropriate for programs through which multiple interactive measures may be installed, or where little program- or location-specific data are available.

Multimethod Approaches

Where time, cost, and data constraints allow, using two or more of the methodologies discussed above can mitigate the shortcomings of each, providing a check for consistency of findings and allowing for a greater depth in explanation of drivers of results. Billing analysis and engineering analysis (i.e., building simulation and/or engineering algorithms) are commonly paired, as the former allows for an accurate accounting of reductions in participant consumption, while the latter permits greater scrutiny of savings at the measure level.

Guidance on Application

There is no "one size fits all" to impact evaluation, and each program has its unique requirements and constraints that shape the recommended approach. Table 9 in the main body of the report presents a comparison of the strengths and weaknesses of the five different approaches considered, and the decision trees in Figure 8 through Figure 11 offer recommendations for appropriate evaluation approaches depending on study aims and constraints.

Oil and Propane Savings Calculations

Delivered fuels are common in New England but not as prevalent in other parts of the country, providing fewer models of best practices in evaluating their savings. Moreover, the nature of these fuels—often stored in a tank on site—poses specific difficulties to evaluation. The R91 study specifically examines best practices in assessing these savings, reviewing past evaluations and papers, and providing appropriate recommendations.

Cadmus determined from the literature review that the best practice for evaluating oil and propane program savings is to convert savings values derived using a natural gas billing regression. The underlying assumption for this approach is that per-measure gas savings are statistically equal to per-measure oil or propane savings. Conclusions from evaluations of the U.S. Department of Energy (DOE) National Weatherization Assistance Program (WAP) support this hypothesis. If a regression from the billing analysis is inconclusive, a whole-house gas billing analysis can dictate the whole-house oil or propane savings, as long as the distribution of measures is statistically equal between the oil or propane population and the billing sample of gas consumers.

In cases in which these approaches are statistically inconclusive, the building energy simulation is the next best option. This approach is considered less robust because a billing analysis relies on actual program consumption data, rather than assumptions of savings from modelling simulations.

The least preferred option is the engineering review of algorithms. This study recommends that evaluators should use this approach only when the first two are statistically inconclusive or inappropriate. Other approaches (the metering study and billing analysis using deliverable fuel invoices) are not considered suitable for most evaluations because they require difficult pre-installation operations and are often statistically inconclusive.

Section 2: R16 Case Study—Comparison of Evaluation Approaches

Differing Methodologies

The R91 study examines the methodologies and specific approaches employed both in the R16 impact evaluation and in the development of PSD savings for duct sealing, air sealing, attic insulation, and wall insulation measures. Table 4 presents the different methodologies used to develop *ex ante* and *ex post* gas savings for each of these measures. The study describes each of these approaches in detail in order to facilitate comparison and discussion of differences.

**Table 4. R16 Impact Evaluation and PSD
Savings Estimation Methodologies for Selected Gas Measures**

Category	Measure	HES Evaluation Method	HES-IE Evaluation Method	PSD Method
HVAC	Duct Sealing	Simulation Modeling	Simulation Modeling	Simulation Modeling
Shell	Air Sealing	Billing Analysis (±14%)	Billing Analysis (±31%)	Simulation Modeling
Shell	Attic Insulation	Simulation Modeling	Simulation Modeling	Engineering Algorithm
Shell	Wall Insulation	Simulation Modeling	Billing Analysis (±30%)	Engineering Algorithm

Realization Rate Drivers and Key Differences

The drivers of differences between PSD and R16 savings estimates varied based on the measure in question and the methodology used in either source. Several common themes, indicated below, emerged through this review; additional differences and details are discussed in the body of this report.

- **Site-specific and behavioral factors.** Billing analysis accounts for behavioral factors, such as participant take back, as well as occupancy changes, vacation schedules, participant education, and other similar factors that influence usage. Furthermore, it reflects the quality of measure application, reducing savings where measure savings do not persist or are incompletely administered. While a comparison of actual participant usage before and after program participation can capture these effects, other methodologies used in both the evaluation and the PSD—building simulations and engineering algorithms—presume consistency in occupant behaviors and measure quality.

- **Building simulation input assumptions.** For measures where simulation modeling was used to develop either *ex ante* or *ex post* savings estimates, the assumptions made when constructing a simulated "prototype model" significantly affect the estimated measure savings. The R16 evaluation's input assumptions were shaped through calibration to participant billing data, as well as construction of multiple prototype homes for single family and multifamily homes, and for HES and HES-IE participants.

- **Measure interactivity.** Weather-sensitive measures such as the four considered in this study can be substantially influenced by the concurrent installation of other weather-sensitive measures, especially when envelope and HVAC measures are combined. Billing analysis, by considering aggregate differences in consumption before and after measure implementation, accounts for this interactivity.

- **Geographic specificity of results.** The PSD uses statewide weather profiles to develop savings estimates using both building simulations and engineering algorithms. The statewide estimate of heating degree-days resembles that of Hartford, but not other locations where the evaluation found a large population of participants (i.e., Bridgeport). The evaluation developed building simulations with separate weather profiles for Hartford and Bridgeport participants, and the billing analysis relies on participants' zip codes to determine their nearest weather station and their local weather profile.

- **Robust sources of assumptions.** The PSD building simulations assume that all homes have a natural gas furnace with a total *system efficiency* (combined equipment and distribution efficiency) of 75%, with these savings adapted to other fuel types (e.g., electric, propane, oil) using efficiency assumptions and unit conversions. Similarly, engineering algorithms used to calculate insulation measure savings assume that gas furnaces have a 75% system efficiency. However, system efficiency assumptions are based on PSD developer estimates; given the sensitivity of savings to HVAC system efficiency, this value should be corroborated using recent sources. Federal standards and market assessments indicate that gas furnaces are typically installed at efficiency levels at or above 80% AFUE, but distribution losses reduce the overall system efficiency. Billing analysis captures customers' actual heating equipment type and efficiency, and the evaluation's simulations, having been calibrated to billing data, have also been adjusted to reflect the participant population's heating curves.

Furthermore, PSD insulation measures adjust the heating degree-days (HDD) input to reflect the likelihood that participants do not heat their homes for all hours where the outside temperature

is below 65 °F. However, the adjustment factor currently in use is from the 1989 ASHRAE Fundamentals handbook, and cannot be validated by more recent versions of that source, or by a benchmarking review of other technical reference manuals like the PSD. Providing a robust, current, and Connecticut-specific source to validate this assumption will improve reliability of savings values.

- **Use of year- and program-specific consumption data.** In the R16 evaluation, savings were developed using consumption data specific to the HES and HES-IE programs in PY2011, either directly through the billing analysis or through calibration of the building simulation models. The evaluation results are therefore specific to the program year under consideration, while PSD estimates were developed to be applicable across years and programs.

Recommendations and Conclusions

The R91 study recommends that the following topics be further explored in order to improve alignment between evaluation results and the PSD.

- **Update simulation models for air and duct sealing**. Revise models to use an hourly-iterative simulation software and draw upon participant home characteristics, differentiating between different building, customer, and HVAC types to award the most appropriate savings. Calibrate model prototypes to participant data to ensure that typical consumption patterns of Connecticut customers are reflected in savings computations. In future evaluations, ensure evaluators and PSD developers use an hourly-iterative software package that uses default assumptions and load shapes that are appropriate for residential applications (e.g., BEopt).

- **Differentiate savings values based on population segment.** Certain population segments may not be reflected accurately by the savings developed for an average participant home in the PSD, such as multifamily customers and the lower-income participants in the HES-IE program. Although the air infiltration measure does adapt savings for multifamily customers, the other measures reviewed do not contain a similar adjustment. By adjusting simulation or algorithm inputs and permitting appropriate savings to be awarded specific to these population segments, accuracy of the program-wide *ex ante* savings calculation may be improved.

- **Account for interactivity between HVAC and envelope measures**. Individual measure savings are lowered if installed concurrently; for example, performing duct sealing increases distribution efficiency so that if attic insulation is then installed, heating load drops by a much smaller amount than it would if ducts remained leaky. To account for this interactivity, make an adjustment to reduce savings when multiple shell- or duct-improvement measures are implemented through the program.

- **Consider whether additional weather and location assumptions can improve savings estimates**. The PSD currently uses only a single weather profile to estimate weather patterns that influence savings, which may not reflect the geographic distribution of participants across the state. Areas where a large number of participants are identified (e.g., Bridgeport) have notably lower HDDs than reflected by the statewide average or Hartford weather profiles.

- **Verify that heating HVAC efficiency assumptions remain valid.** Current HVAC system efficiency assumptions rely on estimates that should be validated, given the sensitivity of savings to efficiency values. If system efficiency assumption are found to be low for the participating population, savings may be overestimated. Lower furnace efficiencies require greater HVAC energy consumption to meet winter setpoint temperatures; therefore, measures such as insulation, air sealing, and duct sealing, which reduce heating load, have an amplified effect. Furnace efficiency assumptions influence savings calculated both through building simulation and through the algorithmic approach applied for insulation measures.

- **Assess whether the HDD adjustment factor for insulation measures should be updated.** For attic and wall insulation savings, the current HDD correction factor, which draws from ASHRAE's 1989 handbook, could not be validated with a more current source. An updated value is not provided in more recent versions of this handbook. Provide transparency in what this value seeks to represent.

Introduction

Purpose of this Report

In this study, Cadmus and the NMR Group (collectively referred to as the evaluation team) provide an in-depth review of best practices in common methodologies for residential impact evaluation, as a reference for Connecticut stakeholders. The study also discusses differences in evaluation approaches that may have contributed to divergent *ex ante* and *ex post* savings estimates in a previous study, the R16 impact evaluation of Home Energy Services (HES) and Home Energy Services – Income Eligible (HES-IE) programs.

In 2014, at the behest of Connecticut's Energy Efficiency Board (EEB), the evaluation team conducted an impact evaluation of the Home Energy Services (HES) and Home Energy Services-Income Eligible (HES-IE) programs managed by the Connecticut utilities. This evaluation, the R16 impact evaluation, encompassed programs offered by the following utilities: Connecticut Light and Power (CL&P), United Illuminating Company (UI), Southern Connecticut Gas (SCG), Connecticut Natural Gas (CNG), and Yankee Gas (Yankee). R16 investigated the electric, gas, oil, and propane savings achieved in Program Year 2011 (PY2011) through the HES and HES-IE programs. The HES and HES-IE programs target residential customers living in single-family houses or multifamily buildings, with income eligibility for the HES-IE program set at 60 percent of Connecticut's median gross annual income.

During the evaluation, the team calculated the energy savings achieved through the program, both at the whole-house level and at the measure level, comparing findings with the reported savings calculated using Connecticut's Program Savings Document (PSD). In several cases—most notably for shell measures and duct sealing—the evaluation identified meaningful differences between the evaluated (*ex post*) and claimed (*ex ante*) savings. It was suspected that these discrepancies stemmed, in whole or in part, from differences between the methodologies used to calculate savings by the R16 impact evaluation and the PSD.

Key Impact Metrics

Impact evaluations have a variety of aims depending on the program and measure types evaluated and on the client's specific needs and concerns; nevertheless, several key metrics are commonly reported to indicate program performance and to assess the alignment between claimed and evaluated savings. Metrics commonly discussed in impact evaluation include:

- **Baseline Consumption.** Baseline consumption, ideally, is the counterfactual participant consumption: the consumption that would have occurred in the absence of a program. For measures that were operational at the time of replacement, the baseline consumption is that of the existing measure. For measures replaced at the end of their life (either not working or soon to fail), two approaches are commonly used to establish the baseline:
    - The average efficiency of measures sold in a market at the time of replacement or installation.

- The minimum code-based or standard-based efficiency of a measure sold at the time of replacement or installation.

While the former definition may better capture the meaning of the counterfactual, the efficiency represented by the market baseline can be determined only through often complex and expensive research and may be subject to considerable uncertainty and disagreement. A codes and standards baseline usually represents a more certain counterfactual value, though it may not capture the concept of what would have happened as closely.

- **Claimed Savings.** The claimed, or *ex ante*, savings are the energy and/or demand savings reported by program administrators prior to evaluation. For prescriptive programs, savings are often deemed and presented in tables or in simple engineering algorithms in a technical reference manual (TRM) such as Connecticut's PSD. Custom programs often require additional data tracking and the analysis of engineering work papers used by the implementer to estimate savings.

- **EM&V Adjustments to Savings.** The evaluated, or *ex post*, savings are the energy and/or demand savings derived following an evaluator's review. This report discusses at length several methodologies evaluators may employ to derive *ex post* savings. In some cases, the evaluator may simply seek to verify the savings calculated by the program administrator, while in other cases a more thorough accounting of realized savings is desired. There are three general kinds of evaluation adjustments made to claimed savings.

  - **Database Adjustments** are made to the tracking data produced by program administrators due to input errors, out of range or missing values, or duplicate entries.

  - **Verified Installation Adjustments** reflect changes made to claimed savings based on research conducted to determine whether the quantity and type of measures claimed by program administrators are in place.

  - **Methodology-Based Adjustments** are based on research into the amount of savings that are achieved by installed measures. These are the primary focus of Section 1 of this report and include the following, among others:
    - Billing Analysis
    - Building Simulation
    - Equipment Metering
    - Engineering Algorithms
    - Multimethod Approaches

- **Gross Savings.** These are the savings of all measures installed through the program, whether *ex ante* or *ex post*. The meaning of "installed through the program" is typically that the participants received an incentive for an installed measure or another benefit from the program.

- **Gross Realized Savings Ratio.** The gross realized savings ratio, shortened to "the realization rate," is the ratio of *ex post* gross savings to *ex ante* gross savings. A ratio greater than one indicates the evaluated savings are higher than those initially claimed by a program

administrator, while a ratio below one indicates lower evaluated savings than those claimed by a program administrator.

- **Net Savings.** Net savings are savings that can be attributed to the program, i.e., savings that would not have happened without the program. Net savings include a reduction from gross savings for measures installed through the program that would have been installed even in the absence of a program. Customers who receive a rebate for a measure but would have purchased the measure even without a rebate are known as freeriders and their savings are netted out. Net savings also include additions to gross savings for measures that have been installed *because of* the program (i.e., spillover) but not installed *through* the program.

- **Net-to-Gross Ratio:** The net-to-gross ratio is the ratio of net savings to gross savings--conventionally referred to as NTG. It is often used as an indicator of the quality of program design. A well designed program effectively changes decision-making rather than rewarding people for decisions they would have made without the program.

Report Organization

This report is organized into two sections: the first addresses impact evaluation best practices and the second uses the R16 impact evaluation and the savings in the Connecticut PSD as a case study.

Section 1: Best Practices in Impact Evaluation

The study's best practices review encompasses three topic areas:

- **Literature Review.** This report provides a literature review that details several commonly referenced manuals and guidelines for impact evaluation. The review synthesizes relevant details from each source. Many of these documents discuss the best practices and applications of evaluation methodologies at length, and readers seeking additional detail are encouraged to directly consult these sources.

- **Methodology-Specific Discussions and Guidelines for Application.** For five common evaluation methodologies, the team presents an overview of each approach including its requirements, limitations, and any emerging applications. The team offers guidance for choosing the most appropriate methodology given different constraints, contrasting each methodology in terms of a range of characteristics and applications.

- **Recommendations for Calculating Oil and Propane Savings.** Because of the particular challenges in calculating savings for bulk fuels such as oil and propane and the prevalence of these fuels in the Northeast, this report provides a thorough review of common practices used to calculate oil and propane savings, offering illustrative case studies and recommendations.

Section 2: R16 Case Study—Comparison of Evaluation Approaches

Following the assessment of common approaches to impact evaluation and their most appropriate applications, the best practices review is applied to the R16 evaluation as a case study. This section includes a discussion of how the approaches used in the R16 evaluation and in the development of savings in the PSD may influence their respective savings calculations. Section 2 provides an overview of

the methodologies employed in calculating *ex ante* and *ex post* savings for four measures driving the gas realization rates: attic insulation, wall insulation, duct sealing, and air sealing. Based on this review, the report suggests areas for improvement or further examination.

Section 1: Best Practices in Impact Evaluation

Overview

Scope of Review

This report considers evaluation methodologies appropriate for calculating first-year gross energy savings associated with energy efficiency programs targeting residential retrofits. The evaluation practices discussed are intended to cover the range of residential measures in such programs, including building envelope, HVAC, lighting, hot water, and appliance measures. This focus is in line with the EEB's special interest in the recent R16 evaluation of the HES and HES-IE whole-house retrofit programs (which include an initial whole home audit and installation of multiple measures noted above). This report provides general guidance on expanding the insights to upstream programs, demand response programs, behavior programs, new construction programs, and other residential program types that require more specialized and specific approaches.

Impact Evaluation Approaches Considered

A sample of the most prevalent evaluation approaches and methodologies were selected for the best practices review, investigating the following:

- Billing Analysis
- Building Simulation
- Equipment Metering
- Engineering Algorithms
- Multimethod Approaches

The following sections provide a detailed discussion of each of these approaches, with references to appropriate guidelines and manuals to support the review.

Literature Review and Research Sources

This report, in its study of impact evaluation best practices, conducted a literature review of widely used and authoritative evaluation protocols and manuals. This section includes an overview of research sources that the team found useful in assessing the above methodologies and a detailed discussion of the scope and recommendations of each of the following sources:

- International Performance Measurement and Verification Protocol (IPMVP)
- Uniform Methods Project (UMP) Protocols
- Northwest Power and Conservation Council Regional Technical Forum (RTF) Roadmap
- The State and Local Energy Efficiency Action (SEE Action) Network's Energy Efficiency Program Impact Evaluation Guide
- California Energy Efficiency Evaluation Protocols: Technical, Methodological, and Reporting Requirements for Evaluation Professionals ("Evaluators' Protocols")

Table 5. Overview of Evaluation Protocol Sources

Guideline or Protocol	Date of Last Revision	Methodologies Addressed			
		Billing Analysis/ Statistical Modeling	Building Simulation	Equipment Metering	Engineering Algorithms
IPMVP*	2012	X	X	X	–
UMP Protocol**	2013	X	–	X	–
Northwest Power and Conservation Council RTF Roadmap***	2014	X	–	X	X
Energy Efficiency Program Impact Evaluation Guide****	2012	X	X	X	X
California Energy Efficiency Evaluation Protocols: Evaluators' Protocols*****	2006	X	X	X	X

* Efficiency Valuation Organization (EVO). International Performance Measurement & Verification Protocol, Volume I: Concepts and Options for Determining Savings. January 2012. Available online: http://www.evo-world.org/ipmvp.php
** National Renewable Energy Laboratory, U.S. Department of Energy (NREL). *The Uniform Methods Project: Methods for Determining Energy Efficiency Savings for Specific Measures.* April 2013. Available online: http://www.nrel.gov/extranet/ump.
*** Northwest Regional Technical Forum (RTF). *Roadmap for the Assessment of Energy Efficiency Measure*. June 2014. Available online: http://www.nwcouncil.org/energy/rtf/Default.htm.
**** State and Local Energy Efficiency Action Network (SEE Action). *Energy Efficiency Program Impact Evaluation Guide*. Prepared by Steven R. Schiller, Schiller Consulting, Inc. December 2012. Available online: http://www.seeaction.energy.gov.
***** California Public Utilities Commission (CPUC). *California Energy Efficiency Evaluation Protocols: Technical, Methodological, and Reporting Requirements for Evaluation Professionals.* Prepared by TecMarket Works. April 2006. Available online: http://www.calmac.org/publications/EvaluatorsProtocols_Final_AdoptedviaRuling_06-19-2006.pdf.

The following appendices contain useful tables and charts from these sources that can assist in the selection of the best approach for a given program:

- Appendix A: IPMVP M&V Option Selection Flowchart

- Appendix B: Suggested IPMVP M&V Options for Different Projects

- Appendix C: SEE Action: Applications for Each IPMVP M&V Option

- Appendix D: UMP Recommended Consumption Data Analysis FormAppendix D: UMP Recommended Consumption Data Analysis Form

- Appendix E: RTF Roadmap for Assessment of Energy Efficiency Measures: Selecting a Method for Savings Estimation

International Performance Measurement & Verification Protocol

The Efficiency Valuation Organization (EVO) prepares the IPMVP to provide guidelines, definitions, and an overview of best practice techniques for verifying the results of energy efficiency projects in commercial and industrial facilities. The IPMVP Committee, consisting of global industry experts, develops the protocol.

EVO released *IPMVP Volume I: Concepts and Options for Determining Energy and Water Savings*[2] in January 2012. The document outlines four options for determining energy savings, Options A–D, and recommends Options C or D when estimating savings at the whole-building level. Brief descriptions of the four options follow:

- **IPMVP Option A, Retrofit Isolation with Key Parameter Measurement:** Option A is a retrofit isolation technique in which the evaluator determines savings by measuring the key performance parameters that influence the energy use of energy conservation measures of interest. The technique uses engineering models to calculate energy use as a function of several inputs. Option A applies to system- or project-level evaluations when the evaluator can calculate savings at the end use, such as with lighting or ventilation systems.

- **IPMVP Option B, Retrofit Isolation with All Parameter Measurement:** Option B is similar to Option A in its focus on systems or projects and its use of engineering models. Unlike Option A, Option B measures all parameters that influence the energy and demand savings of an energy conservation measure, such as hours of use for lighting, meaning that it requires more rigorous measurements of equipment characteristics and performance factors.

- **IPMVP Option C, Whole Facility:** In Option C, program evaluators measure the energy consumption of the entire facility. Option C calculates savings through an analysis of utility meter data gathered before and after program participation, making appropriate adjustments with single comparison or regression analysis techniques. Typical applications for Option C are programs that encompass a variety of measures and produce large savings compared to random energy variations at the whole-building level.

- **IPMVP R:** Option D also calculates the savings for an entire facility, but uses computer simulation software to predict facility energy use, calibrated with hourly or monthly billing data. This option is suitable for users who want to estimate savings from individual measures or for energy management systems that lack baseline energy data.

The IPMVP description of the appropriate use of each option is detailed and includes best applications guidelines for each option, in addition to an option selection guide and a flowchart to aid in choosing a logical approach for the program being evaluated (see Appendix A: IPMVP M&V Option Selection Flowchart and Appendix B: Suggested IPMVP M&V Options for Different Projects).

The IPMVP further addresses common measurement and verification (M&V) issues affecting Options A–D: the role of uncertainty, the differences between observed and true energy use, and baseline adjustments to account for unexpected or one-time changes in conditions within the systems under

---

[2] Efficiency Valuation Organization (EVO). International Performance Measurement & Verification Protocol, Volume I: Concepts and Options for Determining Savings. January 2012. Available online: http://www.evo-world.org/ipmvp.php

evaluation. This protocol also provides general guidelines for balancing cost and uncertainty, taking into account variations in energy use and the value of the energy conservation measures.

Uniform Methods Project

The UMP[3] is an initiative of the National Renewable Energy Laboratory (NREL), funded by the Office of Electricity Delivery and Energy Reliability and the Office of Energy Efficiency and Renewable Energy, which draws on the expertise of a wide range of industry experts. In April 2013, NREL produced protocols for seven energy efficiency measures, both residential and commercial, designed to provide guidance for specific energy efficiency measures, program types, and evaluation activities. An additional five crosscutting protocols were issued in 2013, and a further nine protocols were published in 2014. Table 6 lists the residential and crosscutting protocols issued through UMP in 2013 and 2014.

Table 6. Uniform Methods Project Protocols Published in 2013 and 2014

Residential Protocols	Crosscutting Protocols
• Residential Lighting • Residential Refrigerator Recycling • Residential Furnaces and Boilers • Residential Behavior • Residential Whole-Building Retrofit	• Assessing Persistence and Other Evaluation Issues • Metering • Peak Demand and Time-Differentiated Energy Savings • Sample Design • Survey Design and Implementation for Estimating Gross Savings • Estimating Net Savings: Methods and Practice

The measure-specific protocols, including lighting, refrigerator recycling, and furnaces/boilers, address calculations and evaluation recommendations appropriate to these measures in isolation. For example, *Chapter 6: Residential Lighting Evaluation Protocol* discusses an algorithmic approach to calculating savings for lighting measures, accompanied by a discussion of measurement and verification (M&V) approaches, secondary sources, and other methodologies that may be used to derive algorithm inputs.

Chapter 8: Whole-Building Retrofit with Consumption Data Analysis Evaluation Protocol provides a recommended approach to calculate savings for a whole-house program in which multiple measures may be installed, similar to Connecticut's HES and HES-IE programs. The protocol offers an overview of consumption data analysis, or billing analysis, as the recommended method for calculating savings for a whole-building retrofit. This UMP protocol recommends the approach laid out in IPMVP Option C for Whole Facilities, provided that the savings are large enough to be observed in consumption data and the billing data are sufficient. The UMP protocol favors this approach because of its focus on whole-house performance and its handling of interactions between multiple measures. The approach offers two analytical methodologies, depending on the consumption data and the comparison group available:

---

[3] National Renewable Energy Laboratory, U.S. Department of Energy (NREL). The Uniform Methods Project: Methods for Determining Energy Efficiency Savings for Specific Measures. April 2013. Available online: http://www.nrel.gov/extranet/ump.

two-stage regression analysis and pooled regression analysis.[4] Appendix D: UMP Recommended Consumption Data Analysis Form Figure 21 provides a table with recommendations for the appropriate modeling approach based on program characteristics.

Chapter 9: Metering Cross-Cutting Protocols identifies considerations and best practices for collecting meter data on a variety of equipment types for residential and commercial applications. Although limited in its discussion of the appropriate scenarios in which these data should be used, this protocol provides a thorough outline of different measurement methodologies and metering devices appropriate for different measured parameters.

Northwest Regional Technical Forum Roadmap for the Assessment of Energy Efficiency Measures

The RTF is an advisory committee established in 1999 to develop standards to verify and evaluate conservation savings. The RTF released the *Roadmap for the Assessment of Energy Efficiency Measures* in June 2014, which consists of instructions for assessing energy efficiency measures. The *Guidelines for the Estimation of Energy Savings* specifically address energy savings.

The RTF guidelines include four savings estimation methods:[5]

- **Unit Energy Savings (UES).** Consistent savings are awarded for each measure unit installed.
- **Standard Protocol.** Savings are variable, but can be consistently calculated using a prescribed and widely applicable methodology or protocol.
- **Custom Protocol.** Savings are variable and require a site-specific data collection plan and analysis approach.
- **Program Impact Evaluation.** Savings are calculated at the aggregate program level rather than by participant or by measure.

For the last of these methods, the RTF provides general guidance on estimating savings from a period of program operation. Measures should be categorized into UES, Standard Protocol, and Custom Protocol, and their respective protocols should be followed, even within a program evaluation approach. The overview of methods specific to overall program evaluation operates at a high level, and it is accompanied by comprehensive guidance on data collection methods and protocols for measure-specific savings estimation models. As shown in Figure 22 of Appendix D: UMP Recommended Consumption Data Analysis Form, the RTF provides a decision tree to assist evaluators in selecting a

---

[4] For detailed definitions of two-staged and pooled approaches, see UMP documentation. NREL. *The Uniform Methods Project: Methods for Determining Energy Efficiency Savings for Specific Measures*. April 2013. pp. 8–22. Available online: http://www.nrel.gov/extranet/ump.

[5] For detailed definitions of estimation methods, see RTF documentation. RTF. *Roadmap for the Assessment of Energy Efficiency Measure. Guidelines for the Estimation of Energy Savings.* June 2014. p. 2–3. Available online: http://www.nwcouncil.org/energy/rtf/Default.htm.

model (e.g., engineering or statistical). The RTF also provides a checklist of required knowledge and skills for program impact evaluations and factors to consider during portfolio assessment.

The RTF guidelines extensively discuss the UES method, which is appropriate for measures in which an evaluator can reliably estimate per unit savings through the measure's lifetime. Program savings equal the sum of the measure count multiplied by the UES value. For measures where the baseline is considered to be the pre-participation condition, a comparison of participant consumption pre- and post-participation could be employed using the following techniques:

- Regression models involving groups of participants with monthly pre- and post-installation billing and weather data;
- Site-specific statistical models to estimate pre- and post-energy use, with the difference representing savings; or
- Calibrated site-specific engineering models to estimate pre- and post-energy use with the difference representing savings.

For measures with a current practice baseline, the study design could employ either site-specific or statistical approaches. The site-specific approach uses only post-participant measurements and compares the average efficient-case energy use with one representing current practice. The statistical approach estimates regression models involving participants and nonparticipants. Using a true experimental design is ideal for reducing bias, but other quasi-experimental designs may be considered.

Energy Efficiency Program Impact Evaluation Guide

The SEE Action Network consists of eight working groups that promote improvements for the design and implementation of state and local energy efficiency policies and programs. SEE Action's evaluation, measurement, and verification (EM&V) group develops the *Energy Efficiency Program Impact Evaluation Guide.* The 2012 version of this guide describes terminology and approaches used to evaluate energy and demand savings from residential, commercial, and industrial energy efficiency programs and portfolios, spanning single system to whole facility applications.

This guide recommends referring to the Department of Energy's UMP for more detailed impact evaluation information, and defers to the IPMVP guidelines (2012) for calculating M&V savings, describing the same options as those offered in the IPMVP, and includes summaries, comparisons, and decision trees for selecting IPMVP options. The guide also suggests additional resources to complement the IPMVP options. For example, for whole-facility applications, the guide recommends using the U.S. Environmental Protection Agency (EPA) Portfolio Manager[6] to analyze a building's utility billing meter data, highlighting the Portfolio Manager's framework and metric for tracking, measuring, and monitoring whole-building energy use. This methodology can be used in conjunction with IPMVP Option C. The evaluator can calculate savings at the building level and can account for differences between fuel

---

[6]   https://portfoliomanager.energystar.gov/pm/login.html

types by combining multiple meters. The guide also provides additional information that is useful in selecting building energy simulation programs for Option D.

In addition to the IPMVP Options, the SEE Action guide presents two large-scale consumption data analysis (billing analysis) methods, randomized controlled trials (RCTs) and quasi-experimental methods (QEMs), which can be used to calculate savings for energy efficiency programs by measuring the differences in energy use between participants and nonparticipants. The guide recommends using these methods for programs in which many participants have common characteristics (e.g., a low-income program) or for evaluations of programs targeting residential behavior, whole-house retrofits, or weatherization.

Using these approaches, the evaluator assigns a study population to a treatment or control group and gathers energy-use data (from meter or billing data) for all facilities to estimate savings. The evaluator examines differences or reductions in energy use between the groups to estimate savings. The guide indicates that RCT methods are appropriate when the control group can be randomly assigned, whereas in QEMs the control group is nonrandomized. The *Energy Efficiency Program Impact Evaluation Guide* recommends using the former because it reduces selection bias and yields more reliable estimates of energy savings.

The section on impact evaluation considerations aids in determining baselines and demand savings, calculating the persistence of savings, addressing the uncertainty of savings estimates, setting evaluation budgets, and establishing evaluation principles and ethics. This section also references helpful resources, including standard industry practices for defining baselines for various categories of programs.

California Energy Efficiency Evaluation Protocols: Technical, Methodological, and Reporting Requirements for Evaluation Professionals

In April 2006, the TecMarket Works Protocol Project evaluation team prepared The California Energy Efficiency Evaluation Protocols: Technical, Methodological, and Reporting Requirements for Evaluation Professionals ("Evaluators' Protocols") for the California Public Utilities Commission (CPUC). The protocols include the *Direct Impact Evaluation Protocols*, which outline approaches to estimating energy and demand savings for impact evaluations of programs that directly achieve these savings. The *Direct Impact Evaluation Protocols* encompass three sections: Gross Energy Impact, Gross Demand Impact, and Participant Net Impact. For each section, the document specifies two or three rigor levels, comparisons of these methods, and approaches to achieve each rigor level.

At the basic level, the Gross Energy Impact Protocol recommends using a simple engineering model (SEM) with M&V equal to IPMVP Option A (the retrofit isolation technique) or a normalized annual consumption (NAC) method. The NAC method uses pre- and post-program participation consumption data from utility bills normalized with weather data. At the enhanced level, the Gross Energy Impact Protocol recommends using the following:

- A regression analysis of consumption information from utility bills;

- A building energy simulation model calibrated in accordance with IPMVP Option D or a process-engineering model;

- A retrofit isolation engineering model as described in IPMVP Option B; or

- An experimental design to obtain reliable net energy savings based upon differences in energy consumption between treatment and nontreatment groups derived from consumption data.

The protocol provides examples of different M&V approaches and detailed guidance on the skills required to conduct impact evaluations, overall results reporting, and sampling strategies.

Methodology Overviews

This section includes a brief discussion of each methodology. Readers should refer to an appropriate sources in the Research Sources section for more in-depth discussion (see Table 5). A summary and comparison of these methodologies is presented in Table 9.

Billing Analysis

Billing analysis uses consumption data collected by utilities for the purpose of billing their customers (i.e., billing data) to estimate program impacts. At its simplest, the program impact computed using this approach is the average change in consumption across a set of participants when comparing a period before program treatment to a period afterward. Under the right circumstances, billing analysis provides solid empirical evidence of a program's impact. It is often employed where multiple measures are installed in the same building, a situation that creates potential interactions that are difficult to capture on a measure-by-measure basis. (This is the case with the HES and HES-IE Programs.) Billing analysis provides a whole-building estimate of program savings.

Weather Normalization

In the most typical cases for residential programs, the energy savings from program treatment are at least partially dependent on weather conditions, especially temperature. Since those conditions vary from year to year, and the analysis aims to provide a generalized estimate of savings, billing analysis attempts to control for the effect of weather on the results. For example, if an unusually cool summer preceded treatment, and an unusually warm summer followed treatment, the impact of efficient cooling equipment would be understated: the reduction in consumption would appear to be less because the efficient equipment had to work harder in the post-treatment year than it would have in the pre-treatment year.

For this reason, billing analysis is rarely a simple subtraction of post-treatment average consumption from pre-treatment average consumption. To account for differences in temperature, as well as an array of other factors, billing analysis is conducted as a regression model where the dependent variable is energy consumption over a specific time period and the independent variables are indicators of

temperature—heating and cooling degree-days.[7] This approach "normalizes" the findings, controlling for temperature differences in the period before and after treatment. The program effect can either be an indicator variable (i.e. 0/1 dummy variable) within the model—indicating whether the time period occurs prior to or succeeding treatment, or separate pre- and post-treatment models can be run. If separate models are run, the program impact is the difference in predicted consumption using the same average temperature values in each model.

Heating degree-days (HDD) and cooling degree-days (CDD) are measures of the difference between outdoor air temperature and a baseline temperature. Heating and cooling degree-days can be summed over time periods of different length to obtain a meaningful indicator of the heating and cooling requirements of that period. The baseline is a temperature at which no heating or cooling is required. A single baseline value can be chosen, somewhat arbitrarily, for all buildings, and often a value of 65° F is used. However, the baseline temperature is likely to vary over a range and may be different for each building, and may differ between heating and cooling seasons. Statistical methods can be used to identify the most appropriate baseline temperature for each building.

Use of a Comparison Group
Weather normalization adjusts for one source of error in comparing pre-treatment consumption with post-treatment consumption; however, it is possible that other differences between the pre- and post-periods, unrelated to the program treatment, also influence results. This could include changes in the cost of energy, general changes in economic conditions, and other factors. To account for such influences (note that it is not necessary to identify what the factors are), the regression can be applied to a comparison group. The impact of the program is then interpreted as the *difference in the differences* in consumption between the program participants and the comparison group in the pre-treatment period and the post-treatment period.

One factor that is incorporated into the analysis by use of a comparison group is naturally occurring changes in energy efficiency. If, during the period of the analysis, households that are not participating in a program are naturally engaging in behaviors or installing measures that reduce consumption, the use of a comparison group in a billing analysis *nets out* those changes.

However, difficulties arise in establishing an appropriate comparison group, that is, a group of customers in which the *only difference* between that group and the program participants is the fact that the comparison group did not participate in the program. Indeed, only under conditions of a randomized control test (RCT), where during program implementation households are randomly assigned to treatment or non-treatment, is there strong assurance that the comparison group is only different in this one regard. Nevertheless, some ingenious approaches have been developed to obtain reasonable comparison groups. For instance, households who will subsequently participate in the program can be

---

[7] Annual degree-days represent the difference between the average daily temperature and a reference temperature, typically 65 °F, summed over all days in a year, with units of °F-days. Degree-days are separated into *heating degree-days* (HDD), when the daily temperature is below the reference temperature, and *cooling degree-days* (CDD), when the daily temperature is higher than the reference temperature.

used as a comparison for earlier participants. Or, a large sample of non-participants can be selected and matched to the participant sample through known characteristics such as square footage or home vintage. For more discussion of comparison groups, the reader is referred to the UMP.

Modelling Approaches

There is significant variety in the way billing analysis regression modelling can be approached. Many of the refinements are beyond the scope of this discussion; however, some broad differences can be mentioned. For instance, one approach is to model each program participant in a separate regression model, predicting a weather-normalized program impact and then averaging the savings across model results. Another approach is to include all participants within a single model that accounts for differences between households—i.e., a fixed effects model—and produces a cross-sectional average directly. It is prudent to estimate savings in more than one way and retain the estimate that has the greatest overall predictive power.

Where there are sufficient data, and especially enough variability in the kinds of measures that have been installed, rather than simply estimating whole-building impacts a measure-level billing analysis can be estimated. In one variety of this approach, indicator variables for each efficient measure type are included in the regression and the separate impact of each can be estimated. A refinement of this approach, referred to as statistically-adjusted engineering (SAE) regression, uses savings estimates as the measure parameters instead of indicator variables. The model produces realization rates on the estimated savings values.

Building Simulation

Building simulations offer a qualified simulation software user the ability to determine the effects of various building retrofit techniques. Building simulations are most appropriately used to determine the energy impacts of weather sensitive measures. Weather sensitive measures influence the energy consumption of heating and cooling systems; examples of such measures include insulation and air sealing. The energy consumption of an HVAC system is dependent on a broad array of factors including: insulation, air leakage, internal gains, temperature, humidity, wind-speed, and solar gain. Accurately accounting for the many possible factors increases the complexity of calculations beyond simple algorithms, which often necessitates a building simulation model. A building simulation can be described as a large set of engineering calculations.

Using building simulation, a modeler will either develop a simulation to replicate the conditions observed in a particular building or develop one or more prototypes representative of a population. The outputs of these models—energy usage at varying levels of temporal granularity and specificity of end-use—can then be employed to determine savings for a facility, a measure, or a program. Building simulations allow interactions between different measures to be considered when calculating energy usage patterns, although they rely on numerous modeler assumptions and approximations.

The quality of building simulation results depends on several parameters detailed in Table 7.

Table 7. Building Simulation Quality Dependencies

Dependency	Detail
Quality of the chosen simulation software	• Supported software (the developer publishes regular updates) • Consistent results, given similar inputs
User's understanding of the software	• Users receive proper training on simulation software and building science • Users understand each input and its effect on results
Quality and quantity of data available	• Measured data points (model inputs) are more accurate than assumed values • Every data point (model input) is measured (ideally)
Assumptions used by the modeler	• Greater accuracy in specific assumptions may be needed depending on the measures and/or effects being tested • Assumptions on heating and cooling efficiency create considerable uncertainty in results
Modeler's approach to calibration using energy consumption data	• Aggregate calibration or single home calibration • Weather-normalized calibration

Model Inputs and Assumptions

The most complex energy simulations account for nearly every major source of energy use in a home, from lighting and appliance loads to occupants opening windows in the summer for natural ventilation. The amount of data needed depends on the simulation tool used for analysis. Many tools use reasonable default assumptions to calculate details such as internal gains from lighting and appliances, and others require these data to be specified. The most important aspects are the efficiencies of HVAC equipment and the specific measures added to the home.

One of the primary inputs defined in an energy simulation is the weather profile under which the home is assumed to operate. If the user is concerned with how the home will perform during an "average year," the user will use a typical meteorological year (TMY) weather set. TMY weather estimates what the weather conditions are like in a typical year for a chosen location based on previous years' weather data.[8] This is most useful for forecasting purposes: estimating how much energy one can expect to save as a result of program participation in a future, typical year.

Alternatively, the user can specify actual meteorological year (AMY) weather data. AMY weather is the actual observed weather for a specific year. This is most useful for determining program impacts in a

---

[8]   For example, TMY3 datasets use an average from 1991 through 2005, while TMY2 uses 1961 through 1990.

specific past year, rather than estimating savings for prospective or "typical" application, and is therefore a best practice for impact evaluations of discrete program years.

Modeling Scenarios and Sensitivity Testing

To determine the savings attributable to one or more measures, prototype models are developed for a variety of scenarios. Typically, a baseline model is developed that consists of the characteristics of a standard, untreated home. Several baseline models may be built when there is interest in determining savings for a variety of home and equipment types.[9] The baseline prototype is built for a given home configuration, and the desired parameter or parameters (e.g., attic insulation levels) are varied between a pre- and a post-treatment state, with energy consumption calculated for both scenarios. When different degrees of treatment are possible, multiple pre- and post-treatment scenarios may be studied.

An important component of this process is sensitivity testing, in which key inputs are varied to determine how strongly the resulting savings are affected by small changes to certain characteristics. When savings are found to be strongly influenced by an assumption or input, the modeler may develop separate prototypes to account for variation in this value or seek additional validation of the input.

Simulation Software Packages

A key to choosing the most appropriate simulation software is understanding the "simulation engine" behind the graphical user interface (GUI), or "shell." The shell of a software program is what the user sees and interacts with to build the simulation. Calculations are then performed by the simulation engine (e.g., DOE-2, Energy Plus, BLAST, and the California Simulation Engine). Many software packages share a common simulation engine.

Residential energy simulation software packages generally fall into one of two categories: hourly iterative or degree-day. Hourly iterative simulations use an approach similar to finite element analysis and weather forecasting models in which the user programs a set of starting conditions or inputs, which are then recalculated for each hour of the year based on profiles of weather conditions and occupant behaviors. Degree-day simulations are best described as advanced engineering algorithms. There are typically no iterative steps in the simulation process; rather, the simulation is a "once through" calculation that uses physics equations and annualized weather datasets to predict energy usage.

Hourly Iterative

Simulation experts consider hourly iterative techniques to be best practice, and these techniques are the basis of the IEA BESTEST[10] (International Energy Agency Building Energy Simulation Test) comparative analysis. When new simulation software is developed, it is often tested for accuracy against

---

[9]   For example, multistory and single-story homes, multifamily and single-family homes, and many more.

[10]   Judkoff, R., and J. Nymark. 1998. "The BESTEST Method for Evaluating and Diagnosing Building Energy Software." The American Council for an Energy-Efficient Economy (ACEEE). Available Online: http://aceee.org/files/proceedings/1998/data/papers/0515.PDF

a suite of hourly simulation models in the IEA BESTEST procedure. Examples of residential hourly simulation software packages include BEOpt, Energy Gauge, and TREAT.[11]

The primary advantage of hourly simulations is that they most closely reflect real-world conditions. In an hourly simulation, the engine uses physics-based equations to calculate thermal losses of the building to the environment using observational weather data. Thermal gains to the building from internal heat loads such as lighting, solar gain, and appliances are concurrently calculated. Based on those gains and losses, the engine also calculates the required energy the building needs to add or dissipate to maintain the temperature of a space. This process is shown below in Figure 1. The best simulation engines simultaneously calculate temperatures in diverse regions of the model: inside the building's conditioned area, in the attic, in the foundation or crawlspace, in the soil, and in the garage. These temperatures feed into the next hourly calculation, and the process repeats. Through this iterative procedure, effects that are unique to certain climates and housing stocks can be identified.

Figure 1. Iterative Calculation
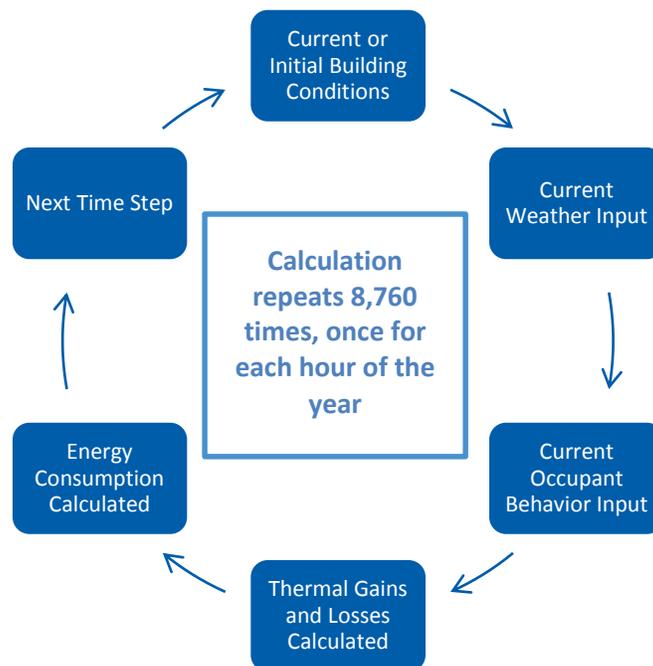
The primary drawbacks of hourly simulations are the computing power and time needed to perform the increasingly complex simulations. Modern computers can perform hourly simulations in approximately 2–30 minutes per simulation, depending on the number of variables defined in the simulation and the computing power available. This length of time is not prohibitive for a small number of models, but can add up when dozens or hundreds of simulations need to be run. For example, an energy auditor may want to test a home with five measure configurations using a typical laptop computer. The measure

---

[11]   International Building Performance Simulation Association. *BEST Directory 2015.* Available Online: http://www.buildingenergysoftwaretools.com

configurations, plus the baseline home, define six individual models that the computer must run separately. If each model were to take five minutes, each home would require 30 minutes to calculate consumption for the entire set. Compared to degree-day models, this is extremely slow.

Degree-Day

Degree-day models use a once-through approach that estimates home energy consumption in a similar way to engineering algorithms. The complexity of these models can range from a simple spreadsheet to an entire software package. The most common degree-day model is REM/*Rate*, used by home energy rating systems nationwide. Degree-day models calculate results very quickly by simplifying energy usage to an annual or monthly basis. This allows the user to run dozens to hundreds of simulations in minutes. These types of simulations are very useful for forecasting energy savings before any work has been completed. An energy professional can quickly compare several sets of energy efficiency measures to determine the most cost-effective sets of measures. Figure 2 illustrates the key data elements required for degree-day modeling using the once-through calculation approach.

Figure 2. Degree-Day Modeling – Once-Through Calculation
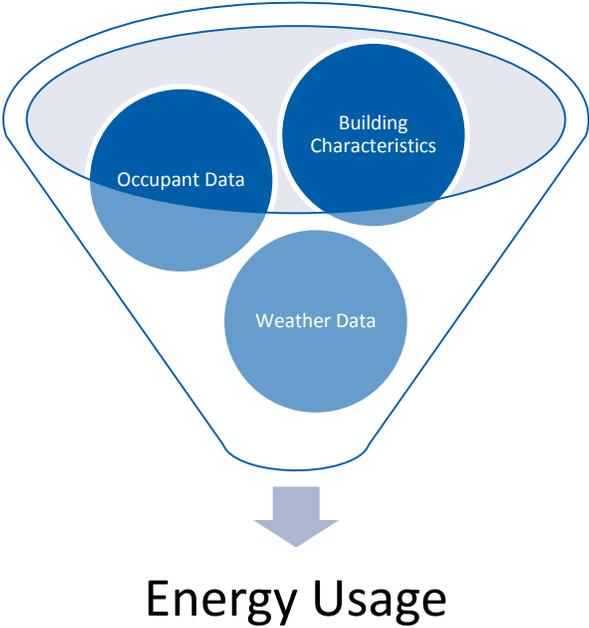
Most degree-day models, including REM/*Rate*, are analyzed using IEA BESTTEST methods to verify that the model gives reasonable results compared to more complex hourly simulations.

Calibration

Billing calibration using the IPMVP Option D method involves verifying that the simulated energy usage matches weather-normalized billing data. The user manually adjusts input assumptions that have low certainty within an appropriate range for each parameter to match billed energy consumption patterns to modeled energy consumption. Common input parameters with low certainty are thermostat settings,

occupancy schedule, plug loads, natural ventilation (opening windows), mechanical ventilation (fans), and thermal mass (important in dry climates with high daily temperature differences).

Verifying that calibration has been achieved can be a relatively subjective process. A common method is to compare regression analysis outputs (i.e., stream weather-normalized billing data) to modeled data. A perfect fit will have a slope of one, indicating that the same consumption is measured using model results and billing data, and a y-intercept of zero, signifying that a model with zero consumption corresponds to a home whose billing data show no consumption. Figure 3 illustrates an example of good calibration for a furnace-heated model. The regression of modeled results has a slope of nearly one, and an intercept at the origin.

Figure 3. Calibration Curve Fitting Example
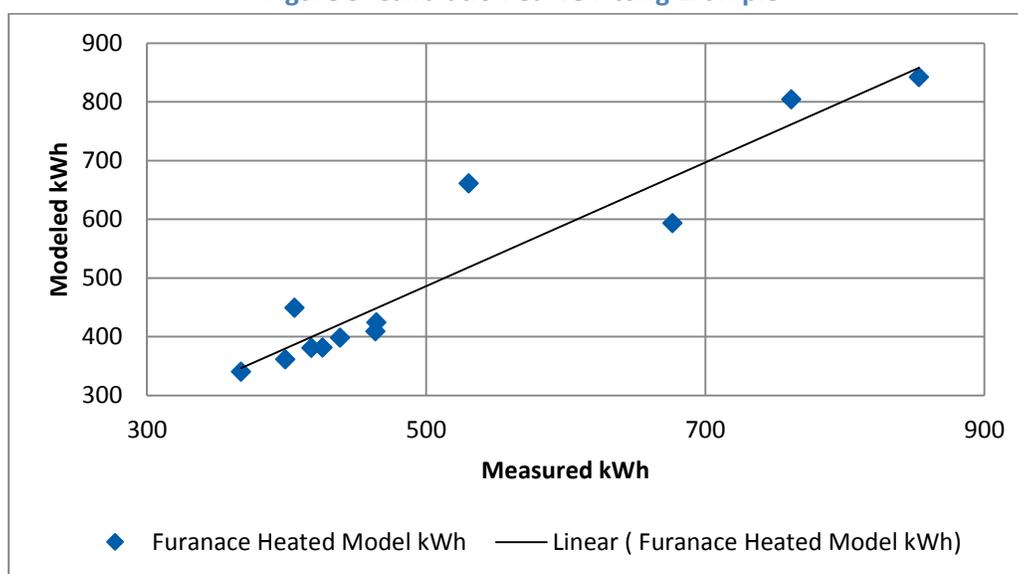
Many energy simulation tools include a utility bill calibration feature. Utility bill calibration is an exceptionally powerful feature that is useful for producing very accurate results by aligning the modeled building energy consumption with actual customer billing data. Utility billing data allows the model to correct for the most important unknowns in energy simulations: the weather-sensitive home usage patterns.

REM/*Rate* has a similar feature called billing disaggregation; with this tool, the user can import the home's bills into the software to be disaggregated into each end use. This tool uses a variable degree-day regression analysis[12] and will normalize consumption to TMY weather conditions. Although the tool does not calibrate the model automatically, it provides feedback to the user that can be used in the calibration process. The user can then calibrate the model by varying inputs such as thermostat settings.
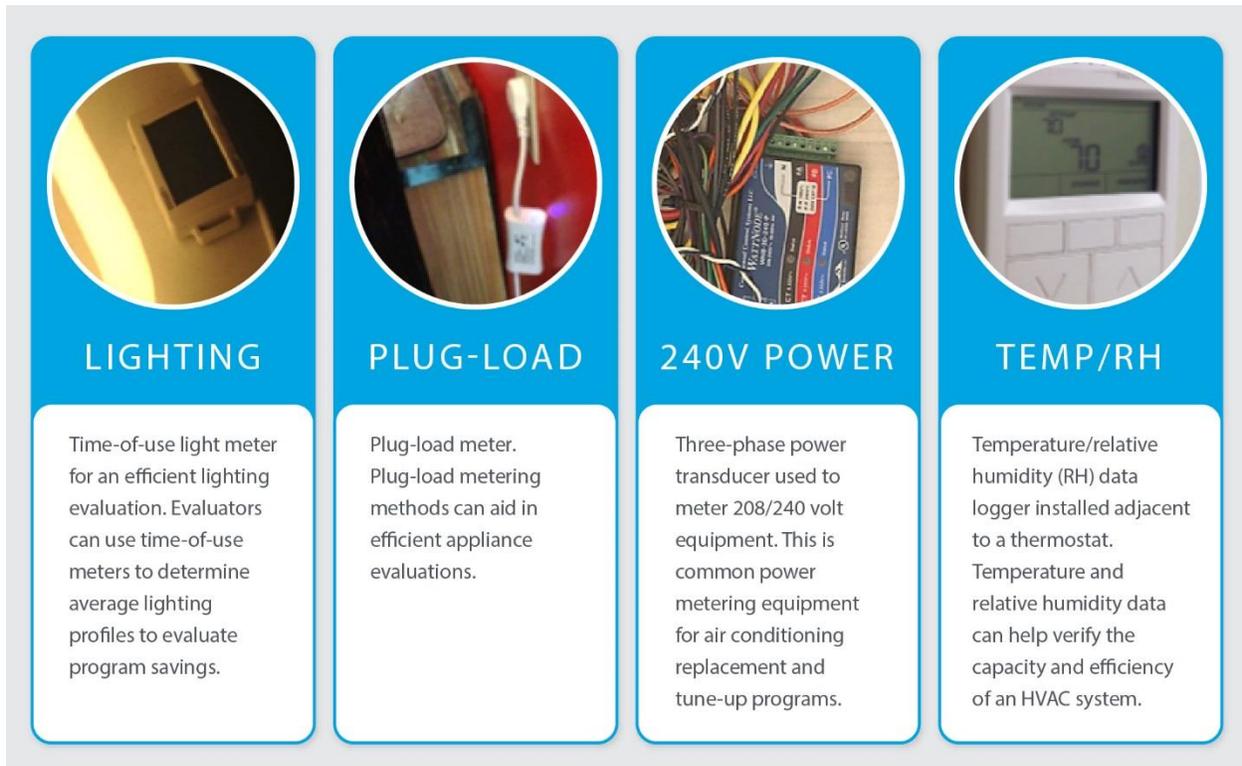
---

[12]   A variable degree-day regression is a type of multivariate regression correlating HDD, CDD, and energy usage.

Equipment Metering

Residential evaluation metering studies are typically reserved for technology-specific energy efficiency programs. Examples include efficient lighting, efficient appliances, smart thermostats, and cooling and heating equipment replacement programs. Figure 4 displays some examples of field equipment for metering these end uses.

Figure 4. Field Metering Technologies

LIGHTING	PLUG-LOAD	240V POWER	TEMP/RH
Time-of-use light meter for an efficient lighting evaluation. Evaluators can use time-of-use meters to determine average lighting profiles to evaluate program savings.	Plug-load meter. Plug-load metering methods can aid in efficient appliance evaluations.	Three-phase power transducer used to meter 208/240 volt equipment. This is common power metering equipment for air conditioning replacement and tune-up programs.	Temperature/relative humidity (RH) data logger installed adjacent to a thermostat. Temperature and relative humidity data can help verify the capacity and efficiency of an HVAC system.

Challenges arise when using metering in whole-house program evaluation because envelope, thermostat, and HVAC measures are highly interactive and metering savings at a measure-level may not capture these effects, or may not be able to attribute them to a specific measure. Metering the HVAC system (see example data in Figure 5) for a home that has received any combination of thermostat, HVAC, or envelope measures may result in accurate total savings; however, an evaluator has no way of attributing these savings to individual measures since this process can only correlate directly observable values, such as compressor power draw, humidity, and temperature, without identifying the measures causing changes in energy usage patterns. The distribution of measures in the sample of metered sites would have to statistically match the distribution of measures in the program population for whole-house metering to be an appropriate approach. This is typically a logistically demanding and expensive task because metering samples are more difficult to collect than samples that don't require site visits, such as billing samples.

Figure 5. Raw Data Example Readout from a Metering Study*
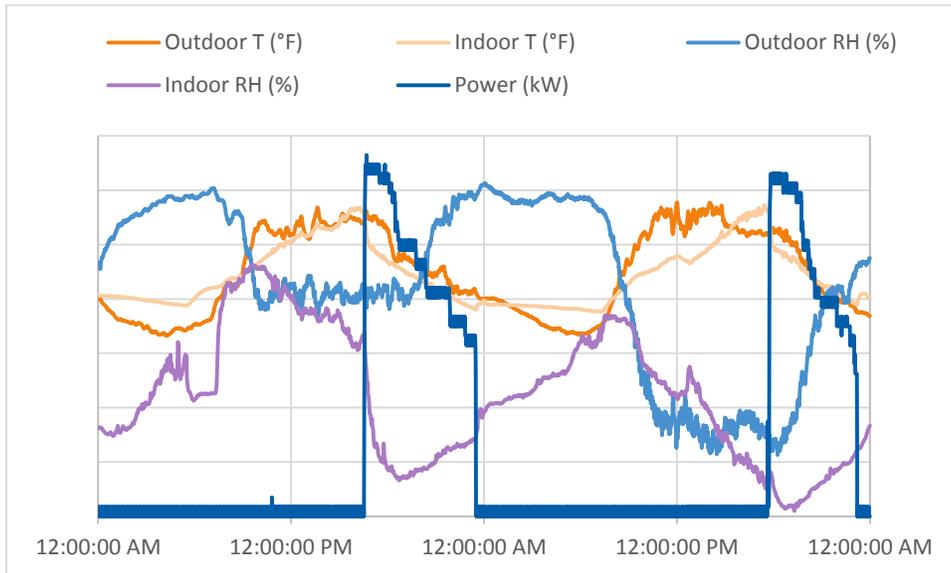
*Plot shows two days of data from an air conditioning evaluation.

Engineering Algorithms

In select cases in which a measure implemented through an energy-efficiency program is well understood and does not affect consumption elsewhere in a home (i.e., there are minimal interactive effects), an algorithmic approach may appropriately capture the savings derived from installation of this measure. For example, many appliances fall into this category; annual energy usage of dishwashers and clothes washer-dryers can be derived from efficiency characteristics and basic assumptions of annual usage patterns, providing a straightforward methodology for calculating the savings achieved by replacing an older unit or installing in a newly-constructed home.

Algorithms based on engineering principles typically employ site-specific data, including details of the measure installed (e.g., quantity), and assumptions about the home, measure, or other interactions occurring. When assumed parameter values are used in the calculation, this approach may correspond with IPMVP Option A: Partially Measured Retrofit Isolation (see Literature Review and Research Sources, International Performance Measurement & Verification Protocol, above).[13] IPMVP Option B: Retrofit Isolation is similar, but parameters are not permitted to be specified using this approach and must be measured on-site.

Algorithmic approaches are advantageous in that a variety of inputs can be used to calculate savings, making the savings awarded easily adaptable to specific participants' home and measure data. While billing analysis and equipment metering reflect the conditions of the homes analyzed, and separate building simulations must be run in order to examine different measure or home characteristics,

---

[13]   Efficiency Valuation Organization (EVO). International Performance Measurement & Verification Protocol, Volume I: Concepts and Options for Determining Savings. January 2012. Available online: http://www.evo-world.org/ipmvp.php

engineering algorithms may be more adaptable to variation in inputs based on measure or home characteristics. However, as the number of inputs to an algorithm increases, so too does the number of datapoints that must be tracked by contractors and program implementers, which may become prohibitively burdensome. In these cases, broader assumptions based on alternate data sources are typically employed instead.

An algorithmic approach is often the least time-intensive method of calculating savings for specific measures, although it is rarely appropriate for programs through which multiple interactive measures may be installed, or where little program- or location-specific data are available. The approach often requires specific participant data to be tracked for the baseline and post-treatment conditions, and may depend on assumptions with varying degrees of suitability. Algorithms may be best employed when inputs and assumptions can be corroborated through other methodologies, for example, using hours-of-use or runtime from metering or billing analysis, or adjusting for interactive factors calculated with a building simulation.

Simple Verification: Deemed Savings Review

In some cases, an evaluation may employ engineering algorithms that align with "deemed" algorithms commonly used to calculate *ex ante* savings. SEE Action's *Energy Efficiency Program Impact Evaluation Guide* (2012)[14] discussed above details the benefits and considerations for using a deemed savings or algorithmic approach, noting that evaluating savings using an approved algorithm or set of values, particularly aligning with those in a jurisdiction's TRM, is quite prevalent,[15] can simplify an evaluation, and will typically reduce costs. However, accuracy of results depends greatly upon the validity and source of assumptions, as well as the adaptability of the calculation to different applications of the measure.

---

[14]   State and Local Energy Efficiency Action Network (SEE Action). Energy Efficiency Program Impact Evaluation Guide. Prepared by Steven R. Schiller, Schiller Consulting, Inc. December 2012. Available online: http://www.seeaction.energy.gov.

[15]   "We found that nearly all states (36 states, 86%) use some type of deemed values in the evaluation framework. In terms of what types of values are "deemed," we found 35 states (97% of those responding to this question) deem savings amounts for particular measures….

We also inquired about the source of the deemed values used by the states. It appears that there is a lot of 'borrowing' going on within the industry. Twenty-six states (70%) cite the use of sources or databases from other states. In nine states, the utilities develop and file certain key deemed values, and in two states, the Commission is responsible for developing the deemed values. In most states (28 states, 80%), the results of their own in-state evaluations are used to modify and update deemed values over time." (pg. 34)

Kushler, M., S. Nowak, and P. Witte. (February 2012). *A National Survey of State Policies and Practices for the Evaluation of Ratepayer- Funded Energy Efficiency Programs*. American Council for an Energy- Efficient Economy (ACEEE). Report Number U122. www.aceee.org/ research-report/u122.

Multimethod Approaches

Where time, cost, and data constraints allow, using two or more of the methodologies discussed above can mitigate the shortcomings of each, providing a check for consistency of findings and allowing for a greater depth in explanation of drivers of results.

Billing analysis and engineering analysis (i.e., building simulation and/or engineering algorithms) are commonly paired, as the former allows for an accurate accounting of reductions in participant consumption, while the latter permits greater scrutiny of savings at the measure level. Calibration of building simulation models, discussed in further detail above, is sometimes considered a multimethod evaluation approach due to the required analysis of load data.

In their paper *Dynamic Duo: How Combining Billing Analysis and Engineering Simulation Methods Improves Evaluation Quality and Understanding*,[16] Crossman et al. note the complementary nature of the advantages and disadvantages of these approaches, as shown in Figure 6.

Figure 6. Advantages and Disadvantages of Billing Analysis and Building Simulation

	Billing Analysis	Engineering Analysis
Advantages	Captures actual changes in usage due to efficiency upgrades and behavioral change Accounts for interactive savings effects between the measures. Preferred approach used in previous Massachusetts evaluations	Directly tied to data on type and quantity of measures installed Can be used for measures of any size, participation level and fuel type
Disadvantages	Requires relatively large impacts and/or sample sizes May not provide sufficient relative precision for smaller measures Requires monthly billing data: not applicable for oil heating customers	Limited by quality and quantity of audit data available Does not capture behavioral effects

Source: Crossman, K., Tabor, L., Perussi, M., and D. Basak. Dynamic Duo: How Combining Billing Analysis and Engineering Simulation Methods Improves Evaluation Quality and Understanding. Proceedings of the 2013 International Energy Program Evaluation Conference, Chicago.

---

[16]   Crossman, K., L. Tabor, M. Perussi, and D. Basak. Dynamic Duo: How Combining Billing Analysis and Engineering Simulation Methods Improves Evaluation Quality and Understanding. Proceedings of the 2013 International Energy Program Evaluation Conference, Chicago.

Billing analysis can reflect behavioral changes, such as customer "take back,"[17] that result from program participation, while these effects are not observed through an engineering analysis. Building simulations, on the other hand, offer an opportunity to investigate the impact on savings of a wide variety of inputs, allowing measure-level savings to be inspected for a variety of home and equipment configurations. This paper recommends using codified criteria in determining to apply one set of savings rather than the other in order to remove a potential area of subjectivity in this approach.

Aligning the results of a billing analysis and building simulation, as was done in both the R16 study and the aforementioned Massachusetts HES evaluation, adds confidence to the findings of the evaluation and enables evaluators to achieve a deeper understanding of the drivers of these results. Building simulations that have been appropriately calibrated to billing data, and that are able to predict similar levels of savings, may be used to test how these savings vary under different conditions and with different baselines. High-granularity, or measure-specific, results from billing analysis typically require substantial sample sizes to obtain results for each combination of home type, HVAC equipment type, and/or measure investigated, which may not always be available; a combined approach therefore provides a reliable alternative. Moreover, measures with a low "signal-to-noise" ratio, where savings are difficult to detect through a standard billing analysis, can be investigated using engineering simulations or algorithms.

Data and Program Requirements

This section describes limitations or considerations that might make a methodology more or less feasible for a given program based on data availability or program structure.

Billing Analysis

Billing analysis relies on two primary data sources: customer billing data and program tracking data. Tracking data are needed in order to verify the measures installed and treatment dates. Because it relies on averaging differences in consumption across households, and because there is substantial natural variation in consumption unrelated to program effects, billing analysis provides the most precise and reliable results when the following four criteria are addressed:

- **Impact Magnitude.** The combined savings of installed measures should constitute a large proportion of total household consumption. A common rule of thumb is that an average reduction of 10% of participant consumption or more can be discerned with reasonable precision through billing analysis.

- **Treatment Variability.** There should be limited variability in the intensity, type, or magnitude of treatment. For instance, participants receiving only direct install measures with low savings

---

[17] Participant take back, also known as snap back or rebound effect, refers to changes in customer behavior prompting an increase usage as a result of participating in energy-efficiency programs. For example, participants who install low-flow showerheads may choose to take longer showers, and those with more efficient HVAC equipment may change their setpoints to increase comfort levels, in each case, rationalizing increased usage based on the assumption their equipment is now operating more efficiently.

should not be included in an analysis with participants receiving major building envelope or HVAC upgrades.

- **Sample Size.** There must be a sufficiently large sample of participants across which to average consumption data. Billing analyses of residential retrofit programs using a sample of fewer than 50 participants typically provide results with poor precision.

- **Data Timespan.** There must be sufficient consumption data available over a long enough period before and after program treatment. The majority of consumption data records should report consumption at no less granularity than a monthly level, with relatively few estimated or imputed readings, and should extend for 12 months before and 12 months after treatment.

The requirement of a substantial period of post-treatment billing data means that billing analysis is typically conducted with program participants not in the current evaluation period but rather in a *prior* program year. The results of the analysis are typically applied as a realization rate on claimed savings rather than per participant, building-level savings. In other words, rather than assigning a given amount of savings to each customer to arrive at an aggregate of, for example, 1,500 kWh in first-year energy savings for the program, the analysis would instead estimate that, for example, 96% of claimed savings had been achieved. This is a subtle but important distinction because it acknowledges the aggregate nature of billing analysis results, shifting the focus of the evaluation from a specific set of participants to the methodology used to estimate savings. Where that methodology has been applied consistently, and where other program changes have not intervened, the evaluation of participants from one time period can be applied to another.

Building Simulation

Building simulations generally fall into two categories: those that model a specific building and those that model a representative prototype or set of prototypes. For the former, accurate building simulations require a full audit of the building or buildings in question before the retrofit work is completed, and accurate details on the retrofit work performed must be provided. For the latter, the average participant home must be characterized for the scenarios to be considered (e.g., an average multifamily home or an average single-family detached home).

Average participant characteristics are optimally drawn from participant data, but may be supplemented by general customer data, or data from secondary resources such as the Residential Energy Consumption Survey[18] (RECS) or the American Community Survey (ACS).[19] Care must be taken to refrain from biasing assumptions about home and measure characteristics; for example, participants who receive additional attic insulation through the program may have a lower baseline level of attic

---

[18] United States Energy Information Administration. *Residential Energy Consumption Survey (RECS).* http://www.eia.gov/consumption/residential/.

[19] United States Census Bureau. *American Community Survey (ACS).* Available online: https://www.census.gov/programs-surveys/acs/.

insulation than the average participant who does not receive this measure. In this case, relying on data from participants who received attic insulation to inform the assumed R-value of a typical participant home could result in an underestimation of attic insulation levels. Use of typical customer population data may also bias assumptions in cases where participants are not representative of the population at large. For programs in which certain segments of the population are targeted, such as income-qualified offerings, this is particularly critical.

Table 8 outlines data that are key to a simulation-based evaluation for a specific building; additional data may be useful, but are not essential.

Table 8. Data Requirements for Simulation of a Specific Building

Key Data (Required)	Useful Data (Optional)
• Size of home (square feet) • Home status: primary, vacation, or secondary residence • Number of occupants • Thermostat settings and usage • Heating and cooling equipment types and efficiencies • Heating and cooling distribution and efficiency characteristics (duct/pipe insulation and duct leakage testing) • Insulation levels applied to ceiling/attic, walls, and floor/foundation/crawlspace • Window types and glazed area • Home air leakage (blower door test) • Water heating equipment type and efficiency	• Appliances, included fuel types and efficiencies • Lighting types and prevalence • Exact dimensions and layout of home (floorplan) • Thermal mass • Door insulation and area • Infrared inspection of home • Detailed HVAC data: blower motor type, expansion valve type, special controls

The amount of data needed also depends on the simulation tool used for analysis. Many tools have reasonable default assumptions to calculate details such as internal gains due to lighting and appliances; others require that these assumptions be specified. The most important details are the heating and cooling efficiencies and details on the measures added to the home. To calculate accurate energy savings for measures that affect heating and cooling end uses, energy modeling is also sensitive to a building's balance-point temperature or reference temperature.

The balance-point temperature is the outdoor temperature at which the home does not require heating or cooling. This temperature depends on two major factors: how well-insulated and sealed the home is, and how much heat is generated inside the home. A very well-insulated home with large appliances and many lighting fixtures will have a low balance-point temperature, whereas a poorly insulated home with few appliances and lighting fixtures will have a higher balance-point temperature. Balance point is important because it defines the switchover point between heating and cooling, from which building simulation software determines the runtime of a home's HVAC equipment. Collecting accurate data on the structure of the building and the internal heat gains refines the balance point of the building.

Engineering Algorithms

A benefit of an algorithmic approach to calculating savings is the limited amount of data that needs to be collected or requested, with most inputs readily available in program tracking data. The data required to be tracked for an algorithmic approach to calculating savings depends on the measure in question and the level of detail in the program tracking data. Site-specific data—such as square footage, HVAC type and efficiency, setpoint temperatures, and HVAC runtime—can improve the accuracy of the calculations, and details of the measure(s) installed through the program and the baseline condition are particularly critical. An algorithmic approach is most likely to yield appropriate savings when assumptions are minimized and participant data can be used to provide customized savings calculations within the program.

When customer-specific data are not available, assumptions must be made based on alternative sources. As noted above in Methodology Overviews, deemed savings and algorithms are often drawn from other jurisdictions and/or from previous evaluations. Caution must be taken in these cases to ensure that the assumed values are relevant to the region, customer segment, and program type being evaluated. Sources from which assumptions may be drawn include:

- Recent program-specific or location-specific primary data collection, such as metering;
- Program evaluations or studies;
- Location-specific and segment-specific customer data; and
- Public databases (e.g., RECS, ACS).

Algorithms may also depend on physical assumptions about weather and water main temperatures. Although location-specific studies are generally preferred, these details can typically be found through publicly available resources such as the National Oceanic and Atmospheric Agency (NOAA), but must be regularly updated as new data become available. Other assumptions about typical equipment sizing, efficiency, and operation can often be found in evaluations and industry white papers; industry manuals or databases (e.g., ASHRAE, Manual J, and the Air Conditioning, Heating, and Refrigeration Institute); and may also be informed by federal standards, state standards, and ENERGY STAR® specifications.

In many cases, the broad and simplified nature of these assumptions yields results that differ from savings measured using other methodologies, such as billing analysis, metering, or building simulation. In some instances, adjustment factors are included in engineering algorithms to reflect typical customer behaviors that influence savings. For instance, because HVAC system operation will depend upon an occupant's setpoint, cost sensitivity, and vacation schedule, adjustment factors may be used to associate a physical calculation of heating and cooling hours, such as heating or cooling degree-days, with typical HVAC runtimes or equivalent full-load hours. Care must be taken to ensure that these adjustment factors are based on a robust and recent study of a relevant population.

Multimethod Approaches

Data requirements for multimethod approaches depend upon the nature of the approaches employed; however, both billing data and site-specific data are typically required, as outlined in the Billing Analysis and Building Simulation sections above. Nevertheless, the use of corroborating analyses can sometimes compensate for deficient data; for example, calibrating simulation models to participant data can yield reliable results even where participant characteristic data may be sparse, as the calibration process can allow evaluators to tune assumptions to the program population.

Innovations and Emerging Practices

Advanced Load Disaggregation

As electrical monitoring technology becomes more widespread and less expensive, the volume of data from electrical meters and load monitoring devices increases exponentially. According to an IBM whitepaper, smart meters typically sample power usage every 15 minutes,[20] and more advanced metering systems sample power usage into the kilohertz range (thousands of samples per second).[21] Typical billing analysis based on monthly billing data can be disaggregated into heating usage, cooling usage, and baseload when at least twelve months of data are available. Smart meter data allow further disaggregation into more precise approximations of heating and cooling and baseload; however, this analysis is limited to identifying large loads operating on regular intervals such as air conditioners and electric dryers. Sampling power usage into the kilohertz range allows load disaggregation to determine which appliances are consuming energy at a given time.[22] As shown in Figure 7, certain types of appliances have unique signatures in the power line when they turn on and draw power. This type of analysis is known specifically as nonintrusive load monitoring.

---

[20] IBM Whitepaper. "Managing big data for smart grids and smart meters." Available online: http://www-935.ibm.com/services/multimedia/Managing_big_data_for_smart_grids_and_smart_meters.pdf.

[21] Laughman et al. "Power Signature Analysis." *ICEE Power and Energy*. April 2003. Available online: http://www.cse.psu.edu/~sem284/cse598e-f11/papers/laughman.pdf.

[22] Gupta, S. EIA Energy Conference 2014 ElectriSense Presentation. Available online: http://www.eia.gov/conference/2014/pdf/presentations/gupta.pdf.

Figure 7. High-Frequency Load Disaggregation
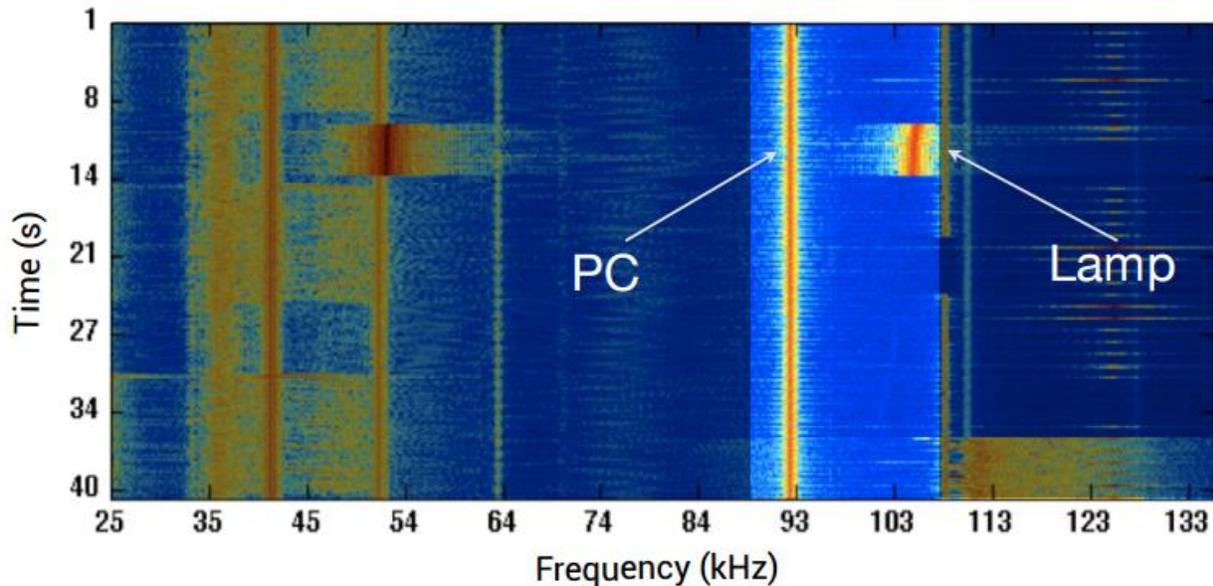
Source: EIA Energy Conference 2014 ElectriSense Presentation

One of the most exciting aspects of nonintrusive load monitoring also gives rise to social concerns that may create barriers for future adoption. Being able to identify exactly which appliances a customer is using at a given time raises concerns about privacy and security.[23] However, from an evaluation standpoint, this granularity offers a variety of new opportunities: large datasets of energy usage profiles may be developed, informing the creation of precise load shapes for appliances and lighting without costly field studies. For example, if a utility program installs high-efficiency lighting fixtures in a home, nonintrusive load monitoring would allow those load reductions to be directly measured. This increases confidence in the results of an evaluation because the impacts of the program can be demonstrated through direct measurement. Efforts are currently underway to codify the security requirements for collection of data from advanced metering infrastructure (AMI).[24]

Specialized Program Approaches

As noted, several types of non-retrofit-style programs merit more specialized and specific approaches for evaluation, including behavior, demand response, new construction, and upstream programs (e.g.,

---

[23]   Murrill, B., E. Liu, and R. Thompson II. Smart Meter Data: Privacy and Cybersecurity. Congressional Research Service Report for Congress. Prepared for Members and Committees of Congress. February 2012. Available online: https://www.fas.org/sgp/crs/misc/R42338.pdf.

[24]   Brown et al. AMI System Security Requirements v1.01. December 2008. Prepared by the AMI-SEC Task Force. Available online: http://energy.gov/sites/prod/files/oeprod/DocumentsandMedia/14-AMI_System_Security_Requirements_updated.pdf

lighting). While these programs were not the focus of this research, the following paragraphs provide a high-level overview of impact evaluation approaches for them.

Both behavioral and demand response programs have relatively few existing evaluation protocols. Behavior-based programs involve providing customer feedback around energy use (e.g., mailers, in-home displays), along with energy education and incentivizing energy reduction, and may involve strategies including competition between participants. Demand response programs can be characterized by pricing (e.g., time-of-use) or control (e.g., direct load control) programs aimed at reducing seasonal or daily peak energy demand. Best practices for these programs often focus on true experimental designs, such as randomized control trials or randomized encouragement designs, and involve regression analysis of energy use for periods before, during, and sometimes even after the program period. These analyses typically rely on large sample sizes, required to discern relatively small impacts in energy (e.g., 1–3% savings) or hourly demand impacts occurring during program events. See the following sources for additional information for more details: UMP protocol for residential behavioral programs,[25] SEE Action protocols for residential behavior-based energy efficiency programs,[26] CPUC's Load Impact Estimate for Demand Response: Protocols and Regulatory Guidance, [27] and Measurement and Verification for Demand Response.[28]

Upstream programs are geared toward providing incentives at the retailer or manufacturer level, rather than to end-use customers. This way, for example, utilities may provide incentives to lighting manufacturers or retailers to stock and sell energy-efficiency bulbs at a lower price point to encourage

---

[25] Stewart, J., and A. Todd. 2015. "Chapter 17: Residential Behavior Protocol." The Uniform Methods Project: Methods for Determining Energy Efficiency Savings for Specific Measures. Available online: http://energy.gov/sites/prod/files/2015/02/f19/UMPChapter17-residential-behavior.pdf

[26] State and Local Energy Efficiency Action Network (SEE Action). *Evaluation, Measurement, and Verification (EM&V) of Residential Behavior-Based Energy Efficiency Programs: Issues and Recommendations*. Prepared by Annika Todd, Elizabeth Stuart, and Charles Goldman of Lawrence Berkeley National Laboratory, and Steven R. Schiller, Schiller Consulting, Inc. May 2012. Available online: https://www4.eere.energy.gov/seeaction/system/files/documents/emv_behaviorbased_eeprograms.pdf

[27] California Public Utilities Commission, Rulemaking 07-01-041. *Decision Adopting Protocols for Estimating Demand Response Load Impacts. January 2007*. Available online: http://www.calmac.org/events/FinalDecisionLoadImpact.pdf http://www.calmac.org/events/FinalDecision_AttachementA.pdf

[28] Goldberg, M., and G. Agnew. February 2013. "Measurement and Verification for Demand Response." Prepared for the National Forum on the National Action Plan on Demand Response: Measurement and Verification Working Group. Available online: https://emp.lbl.gov/sites/all/files/napdr-measurement-and-verification_0.pdf

customer purchases without the additional burden on customers to fill out rebate forms or turn in coupons. The challenge for these programs is that participants often are unaware that they took part in a program. Therefore, evaluators either find technology-specific ways of identifying participants, or they assess impact more at a market level, with making post-analysis adjustments for the program. Otherwise, evaluation protocols for upstream programs may be similar to general protocols for a specific technology. An UMP protocol provides guidance on assessing various inputs used in the savings equation (e.g., hours of use, baseline wattage).[29] Techniques for assessing these inputs cover a broad range of primary and secondary research, including metering to assess average hours of use and phone and in-home audits to assess in-service rates.

Residential new construction evaluations typically focus on using calibrated simulations to model efficiency conditions relative to a baseline of new home construction. The baseline is derived from existing codes and standards, sometimes supplemented with on-site assessments of actual baseline conditions and their level of adherence to codes and standards. The question of what is the baseline against which to compare program homes is usually the most critical component of new construction evaluation. The baseline can be assessed by characterizing the average stage of efficiency in the market or by code compliance studies, the latter of which involves assessing applied energy codes for a sample of new construction projects. The evaluation then uses those estimated baselines as inputs for simulation modeling and compares consumption of those homes with program homes, to determine program impacts. Alternatively, regression-based consumption analysis can be used and applied to treatment and comparison groups, assessing participants' post-treatment use compared to pre-use, which implicitly assuming a market baseline. Commonly referenced protocols include IPMVP Option D, which most often assumes a code baseline.[30]

Conclusions and Additional Guidance

Methodology Strengths and Weaknesses

Depending on the aims of the evaluation and the constraints—in time, cost, data, or some combination thereof—one of the evaluation methodologies detailed above may best suit the needs of the evaluation. Table 9 provides a side-by-side comparison of each of these five methodologies based on the accuracy of findings, precision of results, data requirements, sample size required, evaluation costs, and whether bulk fuels such as oil and propane can be evaluated. Evaluation types that are most appropriate for each approach and those for which the methodology is not suitable are described.

---

[29] Dimetrosky, S., K. Parkinson, and N. Lieb. 2014. "Chapter 6: Residential Lighting Evaluation Protocol." The Uniform Methods Project: Methods for Determining Energy Efficiency Savings for Specific Measures. Available Online: http://www.nrel.gov/extranet/ump/pdfs/20140514_ump_res_lighting_draft.pdf

[30] Efficiency Valuation Organization (EVO). *International Performance Measurement & Verification Protocol, Volume I: Concepts and Options for Determining Savings*. January 2012. Available online: http://www.evo-world.org/ipmvp.php

Table 9. Method Comparison Matrix

	Billing Analysis	Building Simulation	Equipment Metering	Engineering Algorithms	Multimethod
Accuracy	Particularly when an appropriate comparison group is used, results are highly accurate as they directly reflect changes in participant usage.	Moderately accurate, increased accuracy with billing calibration.	Results in high accuracy data for measure-level savings.	Relies on assumed conditions; typically does not provide a robust accounting of interactive factors.	Very high accuracy due to validation of engineering results with measured changes in consumption.
	High	Medium to High[31]	High	Low	High
Precision	Precision of savings estimates depend on the size of the participant sample.	Produces high-precision results, software dependent, including for measures with low levels of savings.	Measurement intervals allow for high data resolution.	N/A	Dependent upon specific methodology and measures.
	Varies	High	High	Low	Varies

---

[31]   Simulation results have increased accuracy when coupled with billing calibration.

	Billing Analysis	Building Simulation	Equipment Metering	Engineering Algorithms	Multimethod
Data Gathering Requirements	Billing analysis requires approximately one year of billing data both before and after program activity for the participant group and, where relevant, a comparison group; tracking data to determine the date and magnitude of program participation are also required. These data are typically available directly from utilities.	Requires collection of detailed home or population characteristics to inform model inputs. When used without billing calibration, more data should be gathered to produce accurate results.	Requires participant contact information and detailed program tracking data on equipment. Need to measure both pre- and post-participation, and depending on the equipment in question, estimates can be improved using a comparison group.	Data on home characteristics and the installed measure are typically required. Additional inputs may improve accuracy of results.	Requires both billing data for participants and control groups and site- and measure-specific data required for development of simulation models. However, shortcomings in one may be compensated for by the other.
	Medium	High	High	Low	High
Sample Size	Savings can be evaluated at once for all participants for whom sufficient billing and tracking data are available.	Requires low to medium sample size.	Requires relatively low sample size for gross savings calculations.	Savings may be easily and quickly calculated for all participants for whom the required home and measure characteristic data are provided. Calculations reflect participant-specific values rather than averages.	Ideally will include billing data for a census and a low to medium sample size for simulation. However, shortcomings in one may be compensated for by the other.
	Census	Medium	Low	Census	Varies

	Billing Analysis	Building Simulation	Equipment Metering	Engineering Algorithms	Multimethod
Evaluation Cost	Data cleaning and analysis can be time-intensive, but data are typically available without significant cost incurred in collection.	Data collection through site visits and/or building plan review is the most significant cost.	Logistically difficult to plan and schedule site visits. High labor and equipment costs are possible.	Costs are generally low because of limited data requirements, typical availability of these data, and the systematic nature of computation. Additional costs may be incurred if evaluators must develop input assumptions using other methodologies.	Due to two-stage evaluation, costs tend to be high dependent on sample size for simulation process.
	Medium	Medium	High	Low	Medium/High
Bulk Fuel[32]	No	Yes	Yes	Yes	Yes

---

[32] Referring to the ability of the method to calculate energy savings for bulk delivered fuels such as oil and propane.

	Billing Analysis	Building Simulation	Equipment Metering	Engineering Algorithms	Multimethod
Best Applications	Reflection of overall program savings, including behavior changes.	Measure-specific savings for interactive measures, pilot programs, and planning purposes.	Measure-specific savings, plug-load measures, and non-interactive measures.	Budget-limited savings verification, measure-level savings for non-interactive measures, and planning purposes.	Accurate reflection of program and measure-specific savings, including behavioral considerations, interactive effects, and low-savings measures. Appropriate for program planning purposes.
Less Suitable Applications	Measure-specific savings for plug load, low-consumption measures, nascent or pilot programs	Programs with a highly variable population of participant homes and measures.	Weatherization or whole-house retrofit programs where interactive effects are prevalent and a diverse range of measures are implemented.	Accurate savings estimations, whole-house programs, interactive measures.	Pilot programs, low-budget evaluations, simple verification, nascent or pilot programs.

Guidance for Application

As outlined in the Methodology Strengths and Weaknesses section above, numerous considerations may motivate the choice of one impact evaluation methodology over another. Considerations that must be taken into account include the following:

- Evaluation objectives,
- Program and measure types,
- Data availability,
- Evaluation costs, and
- Time allotted for evaluation.

For four prevalent evaluation aims, this report provides recommendations using the decision trees in Figure 8 through Figure 11, which indicate appropriate methodologies to employ depending on the constraints faced by the evaluation, as well as the program and measure types. Constraints and characteristics considered are described in Table 10.

Table 10. Evaluation Characteristics and Constraints

| Evaluation Aim	Program or Measure Type	Cost	Time Allotted	Data Required
• Accurate Overall Savings • Measure-Specific Realization Rates • Measure-Specific Assumptions • Simple Verification	• Whole House • Measure-Specific – Interactive* Measures • Measure-Specific – Non-Interactive* Measures	• Low ($30k) • Medium ($100k) • High (>$100k)	• Short (<3 months) • Medium (3-12 months) • Long (>12 months)	• Program Tracking Data • Customer or Participant Contact Information • Billing Data • Measure Details • Site Visits (Primary)
* Interactive measures refer to those measures whose savings may be affected by the concurrent application of other measures. Weather-sensitive measures are typically interactive, but other measures, such as lighting, also have interactive effects. Programs through which a single measure type is installed may not be concerned with interactivity, providing greater flexibility in the choice of evaluation approach.				

Figure 8. Evaluation Methodology Decision Tree: Accurate Overall Savings
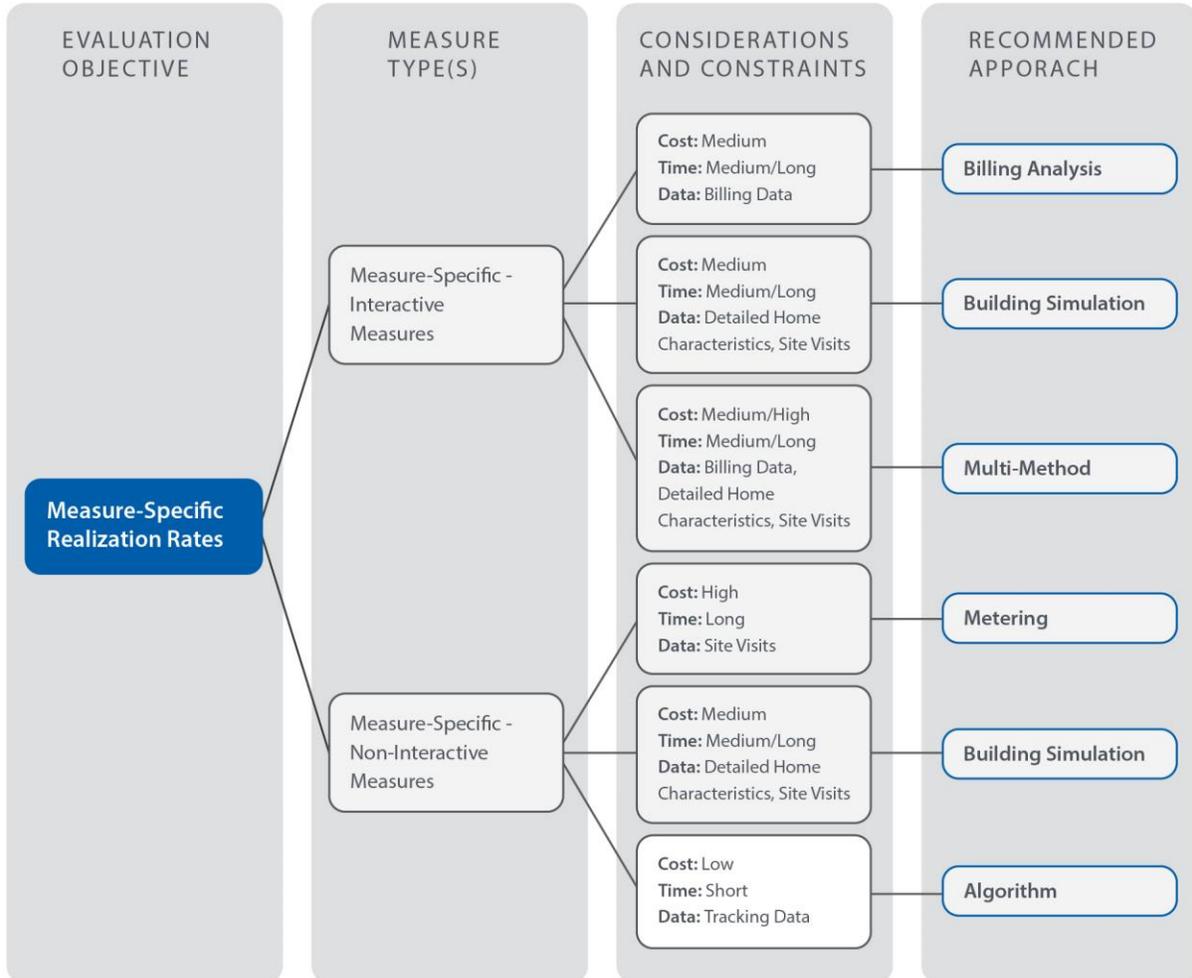
Figure 9. Evaluation Methodology Decision Tree: Measure-Specific Realization Rates
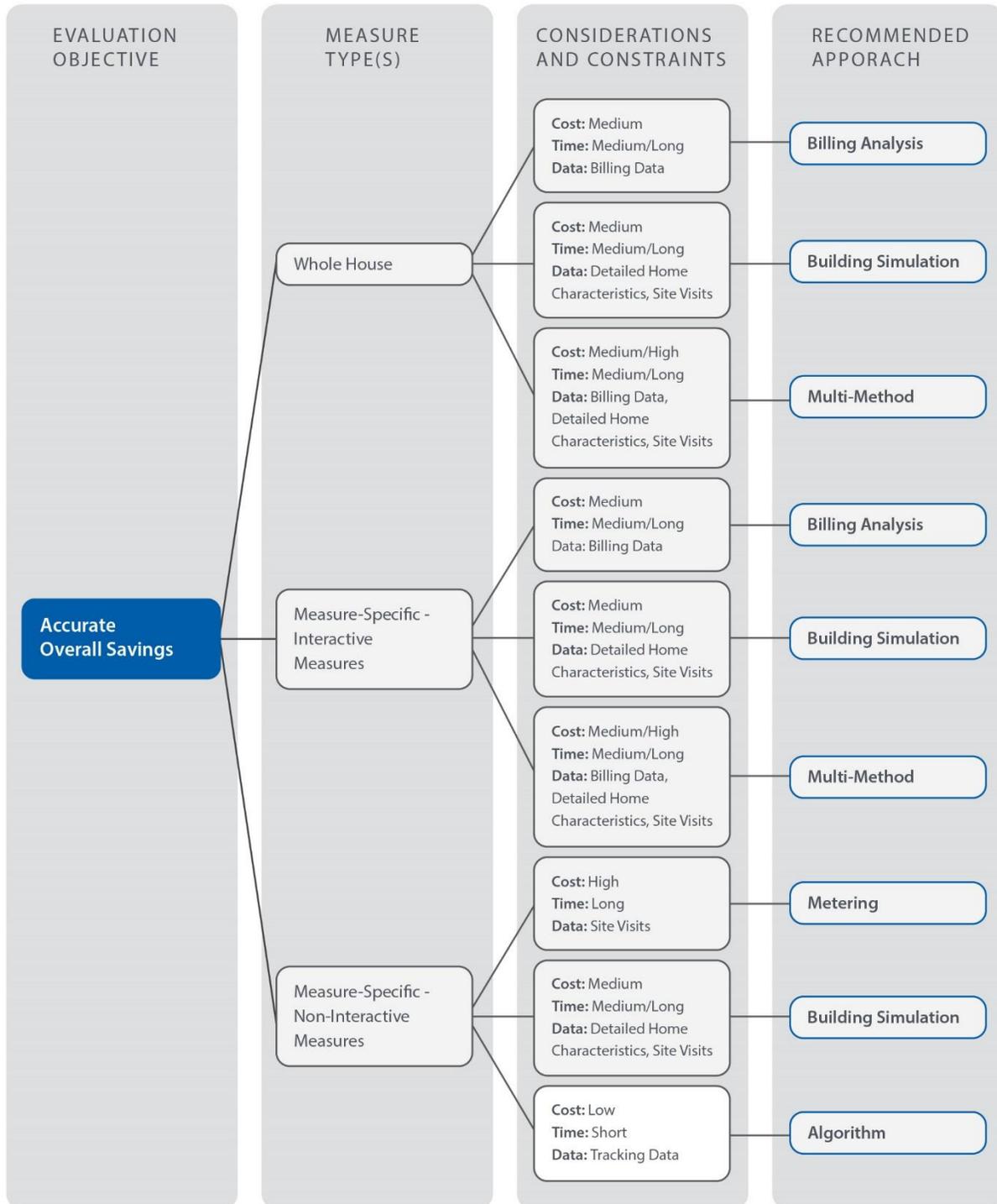
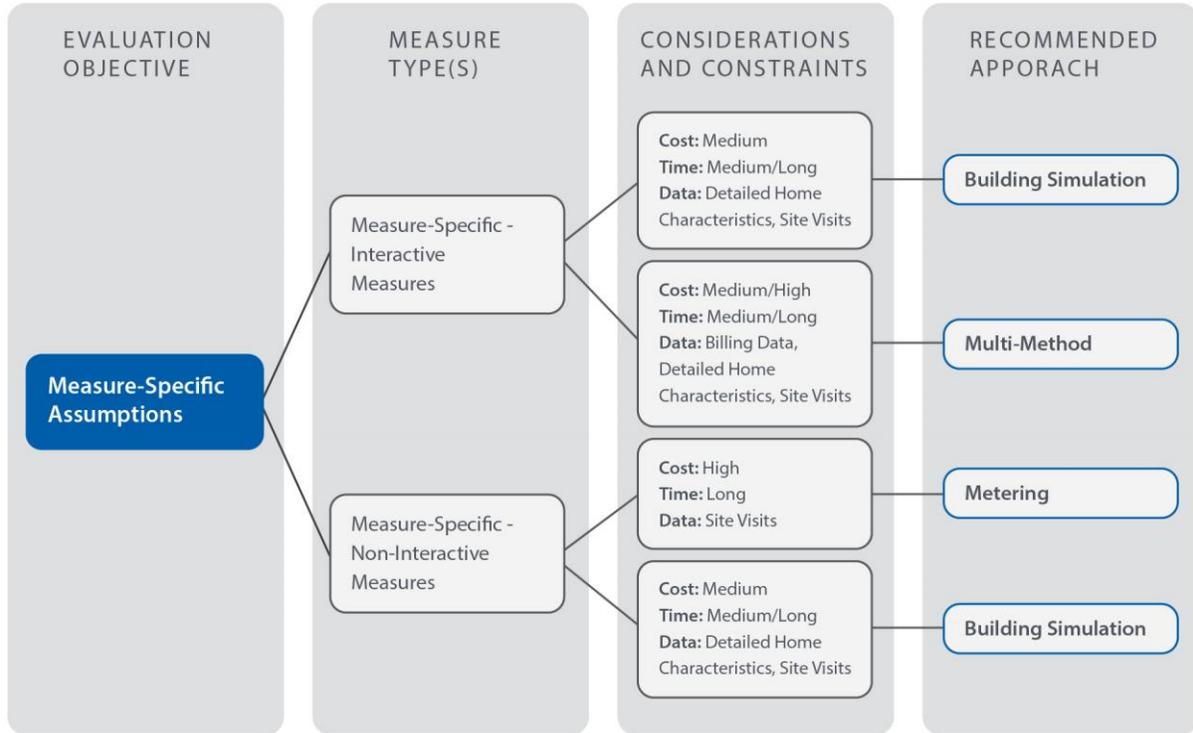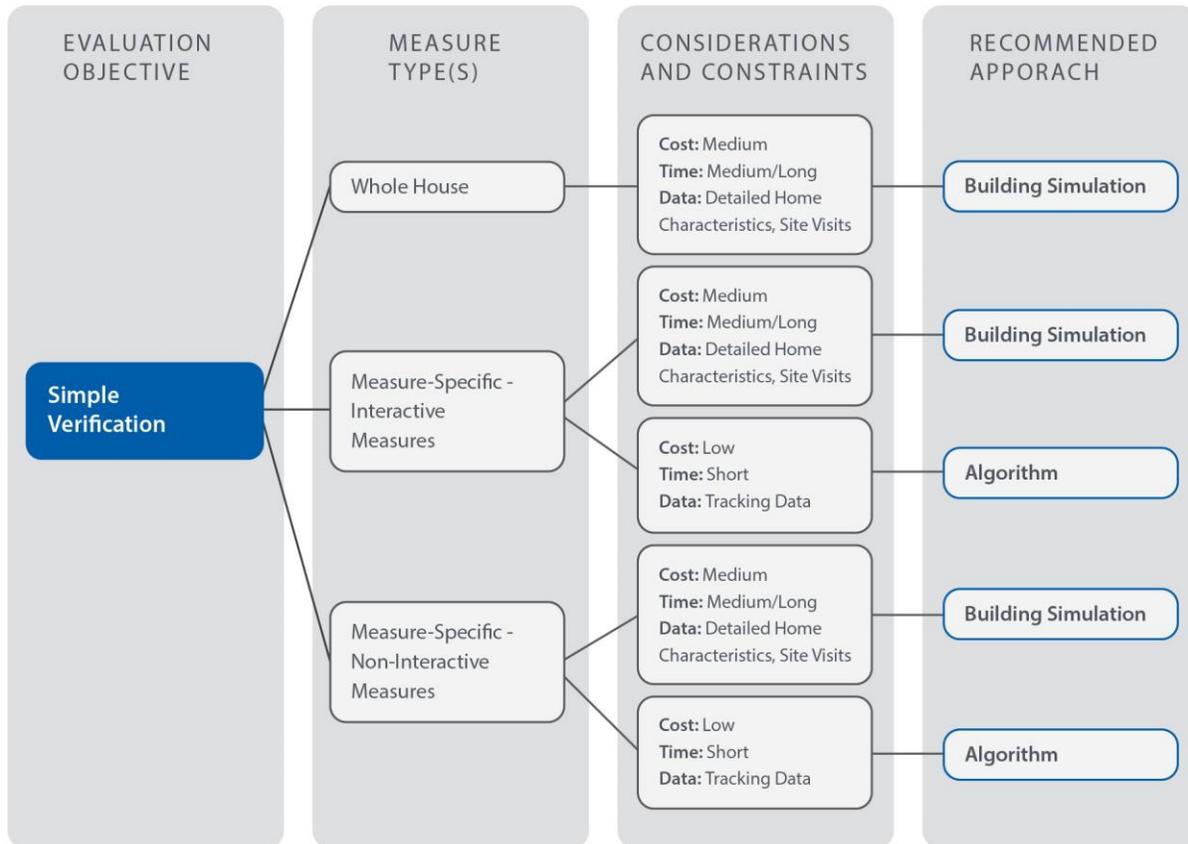**Figure 10. Evaluation Methodology Decision Tree: Measure-Specific Assumptions**

Figure 11. Evaluation Methodology Decision Tree: Simple Verification

Special Topic: Oil and Propane Savings

In certain cold climate regions of the country, primarily in the Northeast, deliverable fuels such as #2 oil and propane make up a substantial market share of residential heating sources. Because of this, whole-house programs that affect heating savings typically include *ex ante* savings for deliverable fuels. The inherent difficulty in evaluating these savings arises from the fact that unlike natural gas, there is no predictable or consistent billing for the fuel, making billing analyses difficult and inaccurate. This study applied the following to determine typical and best practices for evaluating oil and propane savings:

- **A review of whole-house impact evaluation literature supplemental to the R16 report.** The R91 study reviewed recent literature (published within the last four years) including five utility impact evaluation reports and the latest impact evaluation of the National Weatherization Assistance Program (WAP) by the U.S. Department of Energy (DOE). This review included older WAP literature, including an impact evaluation from 2003 and an evaluation plan from 2006.
- **A review of five technical reference manuals (TRMs) from the northeast region, including the Connecticut PSD, that address oil and propane savings algorithms.** The TRM review provided further clarification on acceptable and best practices in treatment of oil and propane savings.

Literature Review of Past Evaluations

The approaches for evaluating oil and propane savings examined in the oil and propane impact evaluation review include:

- Engineering algorithms
- Billing analysis using deliverable fuel invoices
- Building energy simulation
- Metering study
- Natural gas billing regression

The following sections discuss each approach and provide conclusions regarding best practices.

Engineering Algorithms

Engineering reviews of savings algorithms are a common and simple approach for delivered fuel savings evaluations; they are, however, typically used when more robust approaches are unavailable or inappropriate. The sources of the algorithms are usually from an applicable TRM, which may or may not have specific algorithms for delivered fuels. The common method in this case is to apply algorithms for natural gas and convert Btu savings to gallons of oil or propane using known conversion factors, as shown in Figure 12.

Figure 12. Illustrated Approach for Oil/Propane Savings using Algorithms

The 2012 home performance impact evaluation in Delaware[33] used the algorithms in the Mid-Atlantic TRM and the New York standard approach document[34] for calculating gross impacts of all measure types. The Mid-Atlantic TRM does not address deliverable fuels, and the New York standard approach applies the same Btu savings from natural gas algorithms to oil and propane. Similarly, the 2012

---

[33] Opinion Dynamics Corporation. *Delaware Department of Natural Resources and Environmental Control SEU and State Energy Efficiency Programs EM&V Report.* July 2012. Available online: http://www.dnrec.delaware.gov/energy/information/otherinfo/Documents/EM-and-V-guidance-documents/DNREC%20EMV%20REPT%20_FinalRev_2013.pdf

[34] New York State Department of Public Service. *New York Standard Approach for Estimating Energy Savings from Energy Efficiency Programs—Residential, Multi-family, and Commercial/Industrial Measures: Version 2*. December 10, 2014. Available online: http://www3.dps.ny.gov/W/PSCWeb.nsf/ca7cd46b41e6d01f0525685800545955/06f2fee55575bd8a852576e4006f9af7/$FILE/TRM%20Version%202%20December%2010,%202014.pdf

Massachusetts low-income single family[35] and HES[36] whole-house evaluations also incorporated savings algorithms for evaluating deliverable fuel impacts. However, the Massachusetts evaluations only used the algorithm approach when regression modelling and building energy simulation methods were deemed inappropriate or inconclusive. Instead of relying solely on TRM algorithms, the Massachusetts evaluations pooled sources from federal standards, several TRMs, and the billing data from gas participants to inform inputs in the algorithms when necessary (such as baseline heating efficiency).

From these reviews, this study concludes that although natural gas engineering algorithms and assumptions are an acceptable proxy for estimating oil and propane savings, it is important to use project-specific or equipment-specific inputs when available.

Billing Analysis Using Deliverable Fuel Invoices

As mentioned above, it is inherently difficult to provide a billing analysis based on oil or propane billing and invoices because it is unclear from an invoice whether the system's storage tank was completely filled and how full the tank was at the beginning and end of a heating season. These were the issues encountered in the 2011 impact evaluation report for New Hampshire home performance in the ENERGY STAR program.[37] In addition the invoice ambiguity, the evaluation team did not receive enough participant billing records to make any analysis statistically significant.

In an attempt to resolve these issues for the 2006 WAP, the DOE included a provision in their evaluation plan.[38] To allow for an accurate billing analysis, they planned on arranging a sample of homeowners to ensure that their fuel suppliers filled their tank completely at the beginning and end of both the heating season and the measure installation period. The fuel supplier also had to keep accurate records of the amount of fuel delivered and to fill the tank completely each time they made deliveries. DOE did not end

---

[35]   The Cadmus Group, Inc. *Low Income Single Family Program Impact Evaluation*. June 2012. Available online: http://ma-eeac.org/wordpress/wp-content/uploads/Low-Income-Single-Family-Program-Impact-Evaluation_Part-of-the-Massachusetts-Residential-Retrofit-Low-Income-Program-Area-Evaluation.pdf

[36]   The Cadmus Group, Inc. *Home Energy Services Impact Evaluation*. August 2012. Available online: http://ma-eeac.org/wordpress/wp-content/uploads/Home-Energy-Services-Impact-Evaluation-Report_Part-of-the-Massachusetts-2011-Residential-Retrofit-and-Low-Income-Program-Area-Evaluation.pdf

[37]   The Cadmus Group, Inc. and NMR Group, Inc. *Impact Evaluation: New Hampshire Home Performance with ENERGY STAR® Program*. June 13, 2011. Available online: http://www.puc.state.nh.us/electric/Monitoring%20and%20Evaluation%20Reports/124%20NH%20HPwES%20Impact%20Evaluation%20Report%20June%2013%202011.pdf

[38]   Oak Ridge National Laboratory. *National Evaluation of the Weatherization Assistance Program: Preliminary Evaluation Plan for Program Year 2006*. February, 2006. Available online: http://weatherization.ornl.gov/Retrospectivepdfs/ORNL_CON-498.pdf

up evaluating the 2006 program, and the program year 2007 evaluation[39] did not implement the plan. A plan such as this is the only way to ensure an accurate billing analysis for deliverable fuels, but it is challenging from an evaluation standpoint due to planning and quality control logistics.

Building Energy Simulation

For oil and propane simulation modeling, the common practice is to leverage a simulation modeling software package with natural gas billing data. For the two Massachusetts 2012 whole-house impact evaluations,[40,41] the evaluation team performed a natural gas billing analysis for interactive measures (envelope insulation and air sealing) for the gas participants in the program. From this billing analysis, the team determined an average Btu baseline consumption as a proxy for the oil and propane modelling baseline annual use.

Figure 13. Illustrated Approach for Oil/Propane Savings using Simulation Modeling
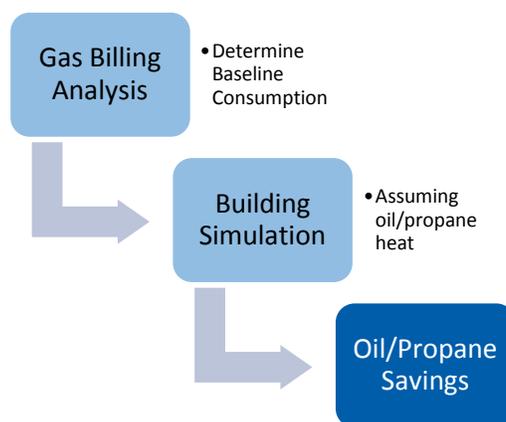
This approach is only valid when billing data from a similar population of gas participants is available. Typically, propane or oil programs are integrated with gas participants because they are run by gas and/or electric utilities. As such, this approach will usually be viable as long as the sample size of natural gas participants billing data is statistically significant in terms of representing the population of oil and propane consumers.

---

[39] Oak Ridge National Laboratory. *National Weatherization Assistance Program Impact Evaluation: Energy Impacts for Single Family Homes.* September 2014. Available online: http://weatherization.ornl.gov/Retrospectivepdfs/ORNL_TM-2015_13.pdf

[40] The Cadmus Group, Inc. *Low Income Single Family Program Impact Evaluation*. June 2012. Available online: http://ma-eeac.org/wordpress/wp-content/uploads/Low-Income-Single-Family-Program-Impact-Evaluation_Part-of-the-Massachusetts-Residential-Retrofit-Low-Income-Program-Area-Evaluation.pdf

[41] The Cadmus Group, Inc. *Home Energy Services Impact Evaluation*. August 2012. Available online: http://ma-eeac.org/wordpress/wp-content/uploads/Home-Energy-Services-Impact-Evaluation-Report_Part-of-the-Massachusetts-2011-Residential-Retrofit-and-Low-Income-Program-Area-Evaluation.pdf

Metering Study

Given the appropriate timeline and budget, a metering study is typically a suitable method for most evaluations, regardless of fuel type. However, unless pre- and post-treatment metering is performed, whole-house metering applications are inherently limited because of the interactive nature of most measures. The DOE's 2014 evaluation[42] of the WAP program implemented a pre- and post-metering study of 120 oil-heated participant homes during a single heating season to evaluate savings. In this study, the evaluation team metered only the runtime of the heating systems, so measure-level savings could not be measured. The evaluation ultimately used another approach to determine gross savings from the program, using the metering study solely to inform conclusions about other approaches (discussed further in the Natural Gas Billing Regression section). This was partially a result of no propane sites having been metered, but also because it is difficult to create a metering sample with a similar distribution of measures to the entire population.

Natural Gas Billing Regression

Similar to the simulation modeling approach, a population of program participants using gas heat can provide billing analysis savings that directly define the savings for the population using oil and propane. The 2014 DOE WAP evaluation[43] experimentally supported the hypothesis that measure savings for natural gas customers equal those of oil- and propane-heated homes. Similar to the 2012 Massachusetts evaluations,[44,45] the WAP evaluation used a billing analysis to determine savings for participants with gas heat. The WAP evaluators used a regression model from the analysis to determine measure-level savings for the population using gas heat. As described in the previous section, the evaluation team also performed a metering study of 120 homes using oil heat, applying the gas billing analysis regression model to each home (incorporating each home's unique distribution of measures). The results from the regression approach differed from the metering results by less than 3%. The evaluation team deemed the billing analysis regression approach as essentially the same as directly metering the savings (Table 11). The team then applied the billing analysis regression to evaluate all of the program's deliverable fuel savings.

Table 11. 2014 DOE WAP Evaluation Results Regarding Whole-House Savings

Approach	Gross MMBtu Savings per Home	Difference
Metering of Oil Participants	22.0 (± 5.0)	-0.6 (± 5.3)
Gas Billing Analysis Regression	22.6 (± 1.8)	

---

42   Oak Ridge National Laboratory. *National Weatherization Assistance Program Impact Evaluation: Energy Impacts for Single Family Homes.* September 2014. Available online: http://weatherization.ornl.gov/Retrospectivepdfs/ORNL_TM-2015_13.pdf

43   Ibid.

44   The Cadmus Group, Inc. *Low Income Single Family Program Impact Evaluation*. June 2012. Available online: http://ma-eeac.org/wordpress/wp-content/uploads/Low-Income-Single-Family-Program-Impact-Evaluation_Part-of-the-Massachusetts-Residential-Retrofit-Low-Income-Program-Area-Evaluation.pdf

45   The Cadmus Group, Inc. *Home Energy Services Impact Evaluation*. August 2012. Available online: http://ma-eeac.org/wordpress/wp-content/uploads/Home-Energy-Services-Impact-Evaluation-Report_Part-of-the-Massachusetts-2011-Residential-Retrofit-and-Low-Income-Program-Area-Evaluation.pdf

The 2014 WAP evaluation outlines important conclusions about the similarities between gas and deliverable fuel savings. The study shows that, given a similar distribution of measures, a natural gas billing analysis can directly indicate the per-home savings for participants using oil and propane. If the distribution of measures between gas and deliverable fuel participants are not statistically similar, a regression model approach for the gas sample is necessary to determine per-measure savings that can then be applied to corresponding measures for deliverable fuels.

Literature Review of Technical Reference Manuals

Several evaluation approaches rely on TRM algorithms for oil and propane savings, whether the evaluation is a direct engineering review or a conversion of natural gas savings to deliverable fuel savings. Cadmus performed a review of several TRMs that address deliverable fuel savings to determine any discrepancies in methodologies.

Cadmus found that the common methodology for deliverable fuel savings in the TRMs is to use the same algorithms for natural gas savings, but to implement equipment efficiency differences based on fuel. Depending on the measure, the equipment is either a domestic hot water system or boiler/furnace heating system. The Maine,[46] Vermont,[47] and New York[48] TRMs all assume no efficiency difference between natural gas and deliverable fuel equipment for measures that include fossil fuel savings. The only TRM reviewed that defines explicit differences in efficiencies by fuel type is the Connecticut PSD.[49] For most measures in the PSD (all domestic hot water and building envelope measures), no difference is assumed; the PSD only assumes differences in equipment efficiencies for heating system upgrades. See Figure 14 for a breakdown of baseline efficiency assumptions by heating technology (boiler vs. furnace), system install date, and fuel type.

---

[46] Efficiency Maine Trust. *Residential Technical Reference Manual: Version 2016.1.* July 1, 2015. Available online: http://www.efficiencymaine.com/docs/EMT-TRM_Residential_v2016_1.pdf

[47] Efficiency Vermont. *Technical Reference User Manual (TRM): Measure Savings Algorithms and Cost Assumptions.* March 16, 2015. Available online: http://psb.vermont.gov/sites/psb/files/docketsandprojects/electric/majorpendingproceedings/TRM%20User%20Manual%20No.%202015-87C.pdf

[48] New York State Department of Public Service. *New York Standard Approach for Estimating Energy Savings from Energy Efficiency Programs – Residential, Multi-family, and Commercial/Industrial Measures: Version 2*. December 10, 2014. Available online: http://www3.dps.ny.gov/W/PSCWeb.nsf/ca7cd46b41e6d01f0525685800545955/06f2fee55575bd8a852576e4006f9af7/$FILE/TRM%20Version%202%20December%2010,%202014.pdf

[49] The United Illuminating Company, and Connecticut Light & Power Company. *Connecticut Program Savings Document*. November 5, 2014. Available online: http://www.dpuc.state.ct.us/DEEPEnergy.nsf/c6c6d525f7cdd1168525797d0047c5bf/8525797c00471adb85257db600666127/$FILE/2015%20Program%20Savings%20Documentation%20.pdf

Figure 14. CT PSD Baseline Heating Efficiencies by Technology, Vintage, and Fuel Type
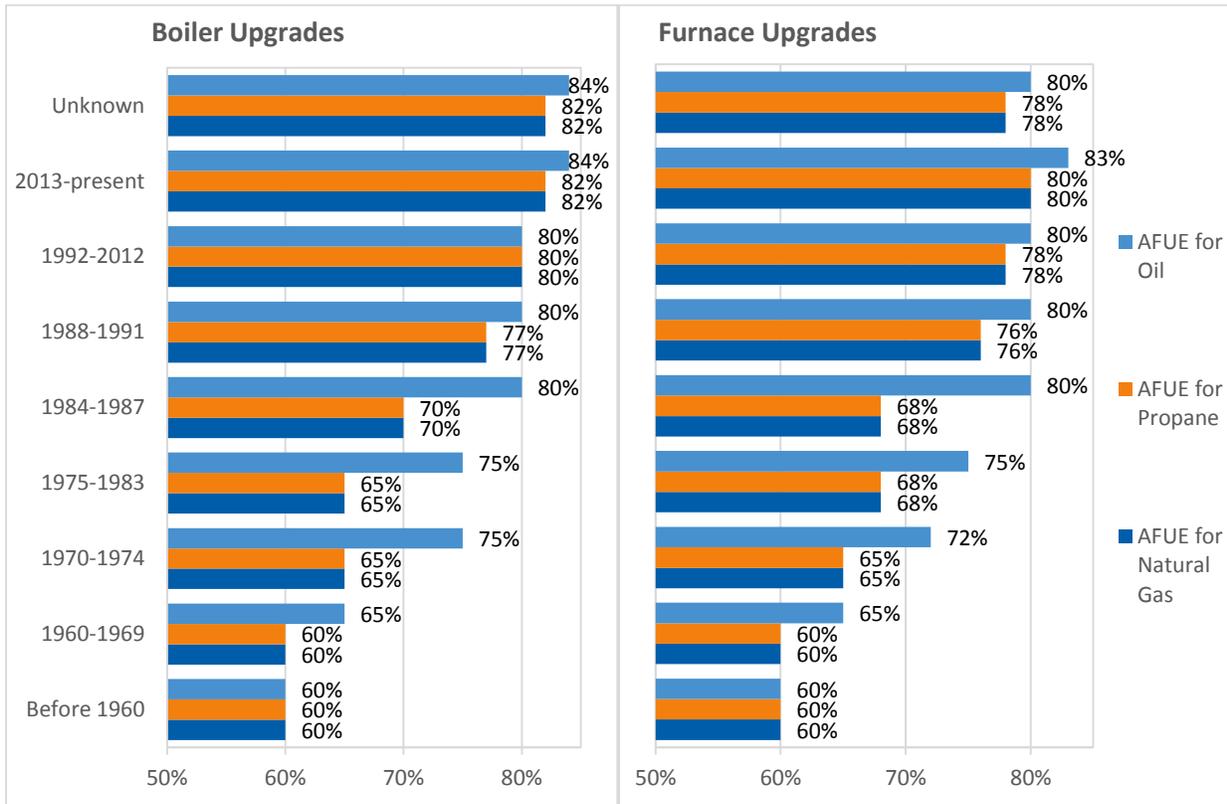
The source of these efficiency values is a combination of ENERGY STAR savings calculators and the current federal standards for minimum efficiencies. These sources only include differences between gas and oil systems, and the PSD assumes natural gas and propane share the same baseline efficiency standards. As seen in Figure 14, the age of the system strongly influences baseline efficiency assumptions: large discrepancies exist between oil and propane or gas systems that were installed between 1970 and 1987.

However, the approach of using age to determine baseline efficiency is only appropriate for early retirement programs. For lost opportunity savings, an evaluator would use the current federal standard efficiency (2013–present values in Figure 14). Even for early retirement calculations, it is uncommon to use efficiencies prior to the 1992–2012 phase because the PSD heating system's effective useful life assumptions are between 15 and 20 years.

Nevertheless, all the TRMs are fairly consistent with the federal minimum standards for baseline equipment efficiency, which indicate small efficiency differences between gas and oil equipment (no standards for propane exist). Furthermore, the 2014 WAP evaluation determined that Btu savings from gas measures were essentially equal to those of equivalent deliverable-fuel measures. This supports the conclusion that differences in equipment efficiencies by fuel type are negligible and deliverable fuel savings are essentially equal to Btu savings from gas measures.

Conclusions and Recommendations

Cadmus determined from the literature review that the best practice for evaluating oil and propane program savings is to use a natural gas billing regression. The underlying assumption for this approach is that per-measure gas savings are statistically equal to per-measure oil or propane savings. The DOE WAP conclusions support this hypothesis. If a regression from the billing analysis is inconclusive, a whole-house gas billing analysis can dictate the whole-house oil or propane savings, as long as the distribution of measures is statistically equal between the oil or propane population and the billing sample of gas consumers.

In cases in which these approaches are statistically inconclusive, the approach outlined in the Building Energy Simulation section is the next best option. This approach is considered less robust because a billing analysis relies on actual program consumption data, rather than assumptions of savings from modelling simulations. The building simulation approach does rely on billing data for gas consumers, but it is more likely to generate more statistically significant results than a billing analysis because only pre-installation data are analyzed instead of pre and post data.

The least preferred option is the engineering review of algorithms. This study recommends that evaluators should use this approach only when the first two are statistically inconclusive or inappropriate. Other approaches (the metering study and billing analysis using deliverable fuel invoices) are not considered suitable for most evaluations because they require difficult pre-installation operations and are often statistically inconclusive.

Section 2: R16 Case Study—Comparison of Evaluation Approaches

Overview

In 2014, the evaluation team conducted an impact evaluation of the Connecticut HES and HES-IE programs. The impact evaluation used a multimethod approach to derive both whole-house and measure-specific savings, the latter of which were compared against savings reported in utility program tracking systems derived using Connecticut's PSD. Through this process, the team identified differences between the gas savings calculated for duct sealing, air infiltration, attic insulation, and wall insulation measures using the Connecticut PSD and those calculated using the evaluation methodologies. For each of these measures, Table 12 presents the evaluation methodologies employed, the PSD methodology used, the *ex ante* and evaluated per-unit savings, and the measure-specific realization rates.

Table 12. R16 Impact Analysis Results for Gas Measures with Significant Variation in Realization Rates

Category	Measure	Reported *Ex Ante* Savings (CCF / Household) (A)	Gross Evaluated Savings (CCF/ Household) (B)	Realization Rate (B/A)	Evaluation Method	PSD Method
HES						
HVAC	Duct Sealing	45	19	42%	Simulation Modeling	Simulation Modeling
Shell	Air Sealing	62	57	91%	Billing Analysis (±14%)	Simulation Modeling
Shell	Attic Insulation	179	135	76%	Simulation Modeling	Engineering Algorithm
Shell	Wall Insulation	449	224	50%	Simulation Modeling	Engineering Algorithm
HES-IE						
HVAC	Duct Sealing	174	28	16%	Simulation Modeling	Simulation Modeling
Shell	Air Sealing	59	36	61%	Billing Analysis (±31%)	Simulation Modeling
Shell	Attic Insulation	152	197	129%	Simulation Modeling	Engineering Algorithm
Shell	Wall Insulation	304	96	32%	Billing Analysis (±30%)	Engineering Algorithm

By using the *ex ante* savings reported by utilities to calculate the R16 realization rates, there is an implicit assumption that the *ex ante* calculations correctly applied the Connecticut PSD algorithms and that measures were appropriately implemented and characterized. Additional analysis of these factors could further clarify variations in realization rates.

Specifically, an additional study (R151)[50] has recently been completed in Connecticut investigating on-site factors that could influence realized savings and their persistence, such as quality of measure installation. The findings from this study are noted below as they relate to the results of the R16 impact evaluation; however, since only billing analysis would capture the effects of non-ideal measure implementation, R151's findings pertain only to the R16 realization rates for air sealing and wall insulation through the HES-IE program.

In this report, the team uses the R16 impact evaluation and the Connecticut PSD as a case study for the above best practices review, examining the different methodologies used to calculate savings and analyzing how these differing approaches may produce differences at the measure level.

R16 Multimethod Approach Overview

The R16 impact evaluation developed gross per-unit savings used as *ex post* estimates for each HES and HES-IE measure, in addition to calculating whole-house savings, using a multimethod approach. The combination of analytical approaches employed included: (1) billing analysis, (2) calibrated simulation modeling, and (3) engineering algorithms.

Brief descriptions of R16's application of these approaches follow. For greater detail, refer to the R16 report:[51]

- **Billing Analysis.** Fixed-effects savings regression models were developed to estimate measure-level savings for measures installed through the HES and HES-IE programs. Weather-normalized models were developed that incorporated detailed measure information from utility tracking data. For the billing analysis, the study used a comparison group composed of future HES and HES-IE participants to test for exogenous effects (e.g., macroeconomic factors) that might have affected energy consumption between the pre- and post-periods. The adjusted gross savings for each measure were calculated as shown in Equation 1.

$$Adjusted\ Gross\ Savings$$
$$= Pre\ Usage_{Participant} \times \left(\frac{\Delta Usage_{Participant}}{Pre\ Usage_{Participant}} - \frac{\Delta Usage_{Non\text{-}Participant}}{Pre\ Usage_{Non\text{-}Participant}} \right)$$

Equation 1

- **Building Simulation.** For program measures known to generate interactive effects (e.g., those increasing or decreasing the energy consumption of another end use, such as insulation), the

---

[50] NMR Group, Inc. Connecticut HES Air Sealing, Duct Sealing, and Insulation Practices Report (R151) *Draft*. December 2015. Available online: http://www.energizect.com/sites/default/files/CT%20HES%20Air%20Sealing%2C%20Duct%20Sealing%2C%20and%20Insulation%20Practices%20%28R151%29_review%20draft_12.23.15.docx.

[51] The Cadmus Group, Inc. and NMR Group, Inc. *Impact Evaluation: Home Energy Services–Income-Eligible and Home Energy Services Programs (R16)*. December 2014. Available online: http://www.energizect.com/your-town/hes-and-hes-ie-impact-evaluation-r16-final-report-12-31-14.

evaluation estimated savings using eQuest, a DOE-2-based simulation model, calibrated using the average pre-program energy consumption of HES and HES-IE participants.

- **Engineering Algorithms.** For measures not typically subject to interactive effects, the evaluation estimated savings using standard industry engineering algorithms.

The R16 study relied on measure- and fuel-specific savings estimated from billing analysis when these met a precision threshold set at ±35% or less at the 90% confidence level.[52] When measures fell outside of this threshold, savings were derived using simulation modeling or engineering algorithms.[53] The former methodology was used for measures when significant interactive effects were expected, and the latter methodologies were used for non-interactive measures.

The R16 evaluation used this process to develop measure-specific electric, gas, oil, and propane savings for measures administered through the HES and HES-IE programs. Oil and propane savings were evaluated by converting natural gas billing regression results to fuel-specific savings, as recommended above in the Special Topic: Oil and Propane Savings discussion of best practices. This analysis resulted in diverse measure- and fuel-specific realization rates. As noted in Table 12 above, the four measures of interest were either HVAC or shell measures, which require consideration of interactive effects. Therefore, either billing analysis or building simulation was used to evaluate these measures.

PSD Methodology Overview

The PSD used two methodologies to develop savings for the measures in question, as shown in Table 12 above: building simulations for air-sealing and duct-sealing measures, and engineering algorithms for attic and wall insulation measures.

The air-sealing measure was first developed in 2006 using an engineering algorithm approach. With the advent of the HES program, a building simulation approach was adopted in 2008 using REM/*Rate* software.[54] This model was used to calculate the savings achieved through the program for air- and duct-sealing measures for each CFM_{50} and CFM_{25} reduction, respectively.[55] Based on industry

---

[52] Although acceptable, these levels fall short of 90/10 confidence and precision for most specific measures, reflecting the relatively low incidence of these measures in the participant population. The evaluation team, however, sought to achieve 90/10 for specific measures whenever possible (and at the program levels). The report shows results with up to a ±35% sampling error to provide the most information possible for assessing program impacts and for future program planning. The reader, however, should recognize that results with a sampling error greater than ±10%–20% do not adhere to standard statistical conventions for acceptable levels of precision. In short, it is very possible that another study with small sample sizes or, preferably, larger ones would produce different conclusions about savings from measures with high sampling errors.

[53] In several instances in which tracking data did not provide sufficient measure details, the evaluation team accepted reported ex ante savings estimates without further evaluation adjustments.

[54] Modeling was performed by the C&LM Planning team, Northeast Utilities.

[55] CFM_{50} and CFM_{25} indicate airflow in cubic feet per minute at pressures of 50 Pa and 25 Pa, respectively.

experience, the developer built a prototypical model for a typical Connecticut single-family home and calculated energy savings per reduction in air or duct leakage. These energy savings, developed for a natural gas furnace, were converted to savings appropriate for a range of HVAC equipment types based on assumptions of unit efficiency. The developer reported that an examination of the results for a variety of leakage reductions revealed a relatively linear relationship between energy savings and a reduction in infiltration or duct leakage, allowing for easy application of these savings.

Attic and wall insulation measures employ a parallel flow heat-exchange calculation, as shown in Equation 4. In this calculation, savings are proportional to the difference in the inverse of the pre-existing and post-treatment R-values,[56] adjusted to account for typical attic and wall construction or framing. These calculations were developed to broadly mirror REM/*Rate*'s internal calculation process for these measures, with some modifications to inputs and assumptions.

R16 Approach and Results

For duct sealing and attic insulation installed through both programs and wall insulation installed through the HES program, the evaluation derived savings using simulation models. Wall insulation installed through the HES-IE program and air-sealing measures were evaluated using billing analysis.

R16 Simulation

The R16 evaluation for duct sealing, attic insulation, and some wall insulation measures relied on both billing analysis and energy simulation results. While several other measures, such as air sealing, were evaluated using a weather-normalized billing analysis, the sample size for homes receiving duct sealing and other weather-sensitive measures was insufficient to develop results within the required precision range. Specifically, the analysis used energy models to calculate the percentage of savings for each weather-sensitive measure. These percentages were then applied to the pre-period weather-sensitive usage for each model to calculate evaluated energy savings.

The R16 impact evaluation developed calibrated models using eQuest, a DOE-2 engine building simulation software that provides hourly outputs. To prevent identified tracking data discrepancies from producing erroneous savings values, the modeling approach applied savings as a percentage of weather-sensitive load rather than awarding savings per measure unit (e.g., per square foot of insulation).

Several prototypical models were developed to provide savings for different scenarios. Prototypical homes were developed for participants in each of the following categories:

- HES and HES-IE participants;

---

[56]   The inverse R-value, or U-value, indicates the rate at which heat is transmitted across a barrier, dependent on barrier area and the temperature difference between both sides. In conjunction with the insulated area and assumed temperature differential, the difference between pre- and post-treatment U-values therefore gives the reduction in heat transmission. The PSD calculation translates this reduction to lower heating and cooling consumption based on unit and distributional efficiency.

- Participants with gas heat and electric heat;
- Participants in single-family and multifamily units; and
- Homes in two weather locations, Hartford and Bridgeport.[57]

This approach resulted in 16 prototypical homes for which energy consumption was observed before and after measure installation. Pre- and post-treatment measure characteristics (e.g., insulation R-values) were based on the average values reported in the program-tracking data for recipients of that measure.[58]

Through the calibration process, the evaluation team adjusted assumptions for the pre-retrofit model to align consumption with billing data disaggregated to show heating, cooling, and other (baseload) consumption values prior to treatment through the program. The models were updated with appropriate assumption values for parameters including HVAC efficiencies, thermostat settings, and end-use schedules. The evaluation team aligned the modeled heating, cooling, and baseload electrical consumption to within 1% of the monthly and annual billing data values.

After the calibration of each baseline prototype home was completed, each measure was separately applied to the model to determine consumption in the improved scenario. The percent savings for each weather-sensitive measure were then applied to the pre-period weather-sensitive usage for each model to calculate evaluated energy savings. The evaluation team then compared the savings calculated through modeling to the billing data used in the calibration (disaggregated to examine only weather-dependent loads) to determine the percentage of energy savings attributable to improvements in each measure in each configuration.

Duct Sealing

Duct sealing savings were first calculated on a per-CFM_{25} basis based on pre- and post-leakage values derived from participant data.[59] The per-CFM savings were then applied to the average leakage reduction from participant tracking data to arrive at measure savings, and compared against weather-dependent load from billing data to determine the percent savings attributable to these measures. Two variations on the duct sealing models were created: one assuming that the ducts were located in an unconditioned basement, the other in an unconditioned attic. The average savings from these two scenarios were used to calculate measure-level percent savings.

---

[57] Although separate savings values were generated based on weather profiles in both locations, models were not separately calibrated on the basis of location.

[58] Multi-family homes were assumed to be single-story, 930 square-foot units for both HES and HES-IE programs. Single-family homes in the HES program were assumed to be two-stories with 1,984 conditioned square feet, while single-family participants in HES-IE were assumed to have 1,500 square feet of conditioned area.

[59] For duct leakage, pre-treatment leakage was assumed to be 522 CFM_{25} and post-treatment leakage was assumed to be 393 CFM_{25} for the HES program, and 786 CFM_{25} and 625 CFM_{25} for the HES-IE program, respectively.

Attic and Wall Insulation

The levels of insulation installed in the modeled homes before and after treatment were calculated from participant data based on program type, building type, and heating fuel. Values were derived through component U-value calculations, weighted by installed surface areas. Appropriate pre-installation levels for wall insulation and attic insulation were determined through reviews of the program data; final input values are shown in Table 13.

Table 13. R16 Modeled Insulation Levels

	HES		HES-IE	
	Pre	Post	Pre	Post
Attic	R-5	R-32	R-3	R-37
Wall	R-0.5	R-15.5	n/a	n/a

R16 Billing Analysis

The R16 study used combined fixed-effects regression models—i.e., a model in which all participants are included, with household differences accounted for using explanatory variable in the model—to estimate savings associated with both program- and measure-level impacts. To estimate measure-level effects, the model included indicator (i.e., binary) variables to denote projects with a given measure.[60] The study relied on participant usage data before and after program participation (provided by CL&P and UI for January 2010 through October 2013) in concert with program data that tracked the measures and installation periods for each participant. Measures with a high frequency of installation that were installed independently of other measures and that demonstrated a sufficiently large level of savings tended to produce measure-specific savings with precision estimates that met the study threshold.

PSD Approach and Results

To calculate savings for air infiltration reduction and duct-sealing measures, the PSD awards savings based on the leakage reduction multiplied by a model-derived factor, the REM value. Wall and attic insulation, conversely, rely on engineering algorithms; default assumptions around wall and attic framing and construction were similar to those in REM/*Rate*, but the assumptions did not necessarily strictly align.

Air Sealing and Duct Sealing

The PSD algorithm for estimating air-sealing and duct-sealing savings is proportional to the variables outlined in Equation 2. The equation assumes that REM energy savings for each type of heating system are proportional to the reduction in duct leakage at a 25 Pa test pressure.

---

[60]   This contrasts with an SAE-based approach, which relies on *ex ante* savings as a model input.

$$Energy\ Savings\ \propto REM \times \Delta CFM_{25}$$

Where:

|		
REM	=	Modeled savings using REM/*Rate* per CFM_{25} airflow change and fuel type
ΔCFM_{25}	=	Reduction in duct system leakage using the CFM_{25} leakage to outside airflow test

The primary basis for the PSD savings algorithm is the REM value developed using REM/*Rate*™ V12.99 energy simulations. The REM values were calculated using prototype home simulations. According to conversations with the modeler, the prototypical home characteristics were based on engineering judgement and industry experience; the typical participant home was assumed to be approximately 1,700 square feet with an unconditioned basement and a 75% system efficiency gas furnace (includes equipment and distribution losses). To calculate REM values for a variety of heating equipment types, energy savings from the model with the gas furnace were converted to the appropriate units and modified based on assumed equipment efficiencies.

The thermostat setting was assumed to be near 68 °F in the winter and 77–78 °F in the summer. Other building characteristics, including the number of stories, could not be confirmed. The modeler calculated building energy consumption for a range of leakage levels for air and duct sealing, reporting that savings varied linearly with leakage decreases, thus justifying the relationship shown in Equation 2. The R91 study verified that modeled air sealing savings vary linearly with leakage reduction using a sample REM/*Rate* model, as shown in Figure 15.

Figure 15. Linear Relationship between Gas Consumption to Air Leakage[61]

---
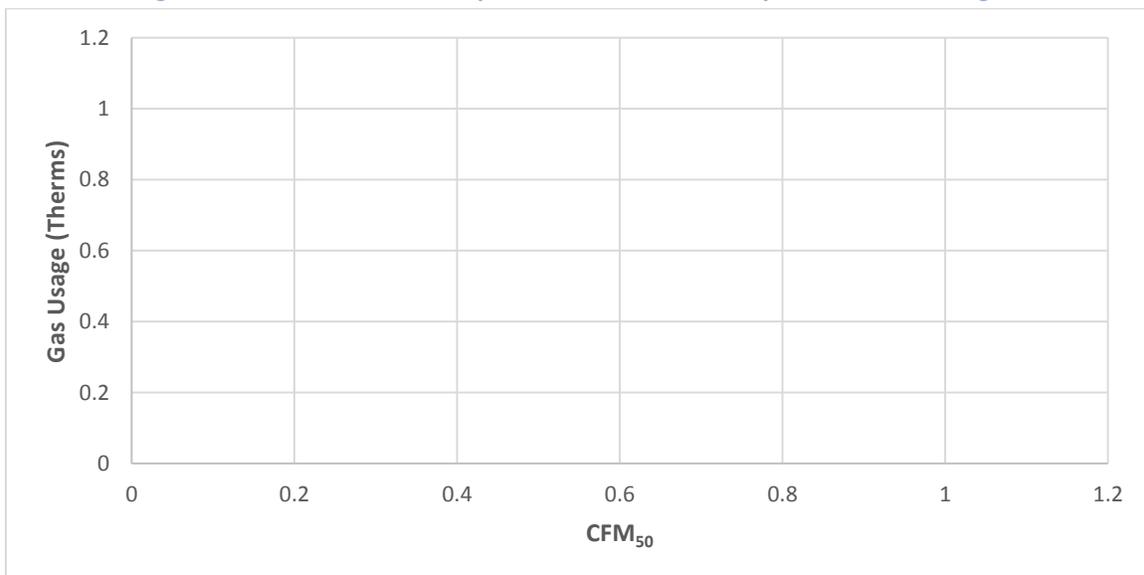
61   Tested by the R91 study using REM/*Rate* V14.

Duct Sealing

Duct-sealing fossil fuel savings values for each system type are shown in Table 14.

Table 14. Duct Sealing Fossil Fuel Energy Savings REM Factors

	Heating Energy (MMBtu)	Oil (Gallons) REM$_{Oil}$	Natural Gas (Ccf) REM$_{NG}$	Propane (Gallons) REM$_{Propane}$
Savings per CFM$_{25}$ Reduction	0.035	0.252	0.340	0.383

Air Sealing

The air-sealing measure was developed in a similar manner to the duct-sealing measure; however, air-sealing savings are dependent on a third parameter, the blower door CFM reduction factor:

$$Energy\ Savings\ \propto REM \times \Delta CFM_{50} \times BF$$

Where:

REM	=	Modeled savings using REM/*Rate* per CFM$_{50}$ airflow change and fuel type
ΔCFM$_{50}$	=	Reduction in home leakage using the CFM$_{50}$ blower door leakage test
BF	=	Blower door CFM reduction factor for multifamily homes

The BF value is applied to account for an unguarded blower door test common in multifamily dwellings. In an unguarded test, the home being tested shares its walls, attic, or floor with another dwelling. Because the air leakage between two dwelling units typically does not result in energy transfer, the test values are corrected using the BF value. The BF value depends on the location of ductwork, the door area, the area of shared surfaces, the envelope perimeter, and the dwelling's age.

Air-sealing fossil fuel savings values for each system type are shown in Table 15.

Table 15. Air Sealing Fossil Fuel Energy Savings REM Factors

	Heating Energy (MMBtu)	Oil (Gallons) REM$_{Oil}$	Natural Gas (Ccf) REM$_{NG}$	Propane (Gallons) REM$_{Propane}$
Savings per CFM$_{50}$ Reduction	0.012	0.087	0.117	0.131

Attic Insulation

The PSD algorithm for estimating attic insulation heating savings is proportional to the variables outlined below.

$$Energy\ Savings_{heating} \propto \left(\frac{1}{R_{existing}} - \frac{1}{R_{new}} \right) \times \frac{HDD \times A}{Eff\ (\%)}$$

Where:

$R_{existing}$	=	R-value of existing attic insulation
R_{new}	=	Total R-value of attic insulation after upgrade
HDD	=	Heating degree-days based on weather location
A	=	Area of attic receiving upgrade
Eff (%)	=	Heating equipment efficiency

The PSD applies adjustment factors to the R-values in the algorithms to account for framing-factor and building material R-values. The framing factor takes into account the structural components of the attic floor that deduct from the gross area insulated. The PSD also applies a correction factor to HDDs to account for the effects of solar and internal heat gains.

The only site-specific measure inputs in the algorithms are the R-values and attic area. The PSD outlines discrete assumptions for the rest of the variables listed below.

- HDD = 5,885 °F-day per year (Connecticut state average)

- Fossil fuel heating equipment efficiency = 75% (regardless of fuel type)

- Electric resistance heating efficiency = 100%

- Heat pump heating efficiency (COP) = 2

Wall Insulation

The PSD algorithms for energy and demand savings for wall insulation measures mirror those for attic insulation (see above section, PSD Approach and Results, Attic Insulation). The wall-specific attributes of the algorithms are listed below.

- The R-value adjustment factors are specific to a 2x4 wall framing and cavity structure

- Energy savings are multiplied by a grade factor (GF).

The GF is less than one for wall structures that are partially below grade and serves as a conservative approach for energy savings; however, it is only applicable if program tracking is detailed enough to provide information about subgrade portions of walls. Also, wall insulation measures will typically result in the GF equaling one because most programs classify subgrade wall insulation as foundation or basement insulation. All other inputs and assumptions are the same as the attic insulation measure in the PSD.

Realization Rate Drivers and Key Differences

Duct Sealing

The duct sealing measure is the sole measure for which the R16 approach to calculating per-measure savings resembles the approach used to develop these savings for the Connecticut PSD, with both employing simulation models to determine measure-level savings. However, notable differences were identified in the approach used during the modeling process, and key assumptions are likely to differ. The original REM/*Rate* files used for the PSD were unavailable, so it is not possible to specify all

differences in inputs and assumptions. Nevertheless, the following are probable drivers of different savings amounts:

- **Modeling software.** Whereas the R16 evaluation used eQuest, an hourly iterative modeling software, for its evaluation, the PSD developer used REM/*Rate*, a degree-day-based modeling software. As discussed above in the Simulation Software Packages section, degree-day based modeling software uses a once-through calculation approach that quickly estimates annual or monthly energy usage. The speed of these software packages allow a greater number of home and measure configurations[62] to be assessed in a short timespan. However, hourly iterative modeling, while more time-intensive, may provide more accurate results for evaluation purposes.

- **Differentiating building type.** The R16 evaluation constructed separate prototype models for single-family and multifamily homes, as well as for participants in the HES program and the HES-IE program. Each prototypical model was calibrated to billing data within the same building and participant category, allowing evaluators to award savings appropriate to each housing type and program. Because homes in each of these categories may have notable differences in house characteristics, HVAC equipment, and operating patterns, this differentiation provides granularity in savings. Table 16 shows the percentage of the gas heating reduction calculated using the R16 evaluation's simulation models. In particular, differentiating savings for single-family and multi-family applications provides markedly different savings estimates.

Table 16. R16 Percentage of Gas Heating Reduction from Simulation Models

Program	Building Type*	Heating Fuel	Wall Insulation % Gas Heating Reduction	Attic Insulation % Gas Heating Reduction	Duct Sealing % Gas Heating Reduction
HES	SF	Gas	31%	13%	2%
HES-IE	SF	Gas	31%	14%	3%
HES	MF	Gas	18%	29%	5%
HES-IE	MF	Gas	17%	33%	5%

* SF = Single Family, MF = Multifamily

- **Input and assumption sources.** The inputs used for the simulation from which the PSD duct-sealing savings are based on modeler experience with Connecticut building stock, and are therefore specific to the time at which the models were developed in 2008. Absent the original models, a thorough comparison of inputs cannot be performed; however, assumptions of home characteristics and equipment may have changed in the intervening years. Furthermore, it cannot be determined whether the PSD inputs are based on knowledge of typical Connecticut

---

[62] Home and measure configurations refer to the potential combinations of installed measures, HVAC systems, home type (e.g., multifamily, single-family, mobile home), home size (e.g., number of stories, square footage), unconditioned spaces (e.g., attics, attached garages, basements), and a multitude of other factors that influence a home's energy consumption.

homes or typical program participant homes, which may influence savings estimates if the average participant differs from the average Connecticuter.

- **HVAC system-specific modeling.** The savings presented in the PSD were calculated assuming that the installed equipment type was a natural gas furnace, and the calculation does not account for any difference in operation between equipment types. According to 2009 RECS data, 30% of New England homes (except for those in Massachusetts)[63] have gas furnaces, with 53% using oil or propane and 7% using electric furnaces (the remaining 10% use either wood or kerosene). Modeling specific HVAC equipment types provides a more customized attribution of savings and allows for any difference in system operation between equipment types to be reflected in savings estimates.

- **Billing data calibration.** As part of the R16 evaluation, the evaluation team calibrated the eight separate building prototypes (home characteristics were assumed to not differ between Bridgeport and Hartford) to participant billing data, allowing for adjustment of parameter assumptions such as heating and cooling setpoints to resemble participant load and behaviors. Calibration can improve the accuracy of simulation estimates, especially when evaluating savings for a specific program year for which billing data are available.

Air Sealing

Despite using a distinct methodology for estimating evaluated savings (i.e., billing analysis) compared to the approach used to develop the PSD estimate, there are several key drivers that may have contributed to deviations in the realization rate for air sealing:

- **On-site factors related to installation or persistence**. As billing analysis uses actual participant consumption data, the savings evaluated using this methodology can be affected by factors that limited the full potential of expected savings. These factors may include quality installation of the measures by the contractor (e.g., complete home sealing with high-quality materials), or persistence of the measure installed as intended (e.g., no home remodels or material degradation or failure).

  At the time of the R16 study, the impact of these effects could not be assessed without additional research activities, such as contacting homeowners, performing site visits, or accompanying installation contractors to monitor their installation techniques. However, the subsequent R151 evaluation researched the effect of some of these on-site factors as they relate to envelope and HVAC measures. In its review of air infiltration reduction measures in Connecticut, R151 found "air-sealing work that was of acceptable technical quality but was incomplete." However, the savings left on the table by incomplete treatment would be reflected in the higher leakage measured in post-treatment blower door tests, and therefore would not contribute to the observed realization rates. Based on R151's assessment, the adequate quality of work and expected persistence of savings within the first the first year after participation— the period upon which the R16 billing analysis is based—would not have resulted in a

---

[63]   Connecticut, Maine, New Hampshire, Rhode Island, and Vermont.

substantial decline in evaluated savings relative to those calculated using the PSD methodology. Nevertheless, the R151 study found that inconsistent testing procedures may have resulted in an overestimation of leakage reduction, a likely contributor to the realization rates for this measure. The report notes that current program materials and guidelines around test procedures allow the potential for vendors to treat conditioned or partially-conditioned basements as unconditioned spaces outside the building's thermal envelope. Leakage measured in this way is likely to be understated, and infiltration between the basement and outside would not be accounted for. For additional detail, the R151 report's Appendix B.3 presents observations regarding this issue and recommendations for improvement.

- **Behavioral or occupancy changes**. Similarly to the on-site installation issue noted above, billing analysis accounts for changes in participant behavior or occupancy between the periods before and after participation that may contribute to an increase or decrease in household energy consumption, such as extended vacations or addition of a household member. The use of comparison group is intended to control for some of the natural changes in household occupancy or behavior, as similar adjustments could be expected in both groups. However, program participants will also receive energy education during the audit administered through the HES and HES-IE programs, which may result in behavioral changes that can be considered attributable to the program. Participant take back—the potential for participants to increase their usage based on the assumption their equipment is now operating more efficiently—is also reflected in savings derived through billing analysis.

- **Alignment of PSD model assumptions with the population**. In using actual participant usage data for the analysis sample, implicitly the equipment and conditions of participant homes are taken into account through billing analysis. Similar assumptions may be made through an engineering-based approach, as in the PSD, but there is the potential for those assumptions to differ from the actual participant population. This is particularly relevant for gas customers, for whom the PSD assume a heating HVAC system efficiency of 75% AFUE in its simulation models. The efficiency of gas furnaces on the market has typically exceeded the federal standard of 78% AFUE, which increased to 80% AFUE in November 2015.[64] A 2011 market assessment by the Energy Efficiency and Renewable Energy Office (EERE)[65] found the typical AFUE for non-weatherized furnaces[66] to be 80% AFUE and above (see Figure 16). Distribution losses will

---

[64] U.S. Department of Energy, Energy Efficiency and Renewable Energy. "Residential Furnaces." Available online: https://www1.eere.energy.gov/buildings/appliance_standards/product.aspx/productid/72.
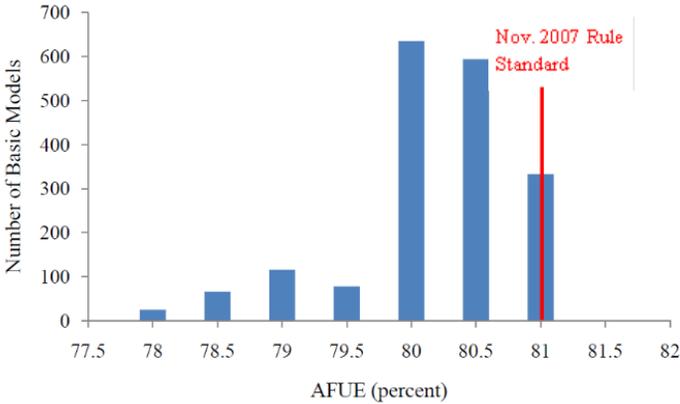
[65] 2011 Federal Register Technical Support Document: Energy Efficiency Program for Consumer Products: Residential Central Air Conditioners, Heat Pumps, and Furnaces. "Chapter 3. Market and Technology Assessment ." Available online: http://www.regulations.gov/contentStreamer?documentId=EERE-2011-BT-STD-0011-0012&attachmentNumber=4&disposition=attachment&contentType=pdf

[66] Non-weatherized furnaces—furnaces designed for indoor installation—are typical in the residential sector.

reduce the overall system efficiency, but a robust and recent source is needed to determine the accuracy of the PSD's 75% system efficiency estimate.

Figure 16. EERE Market Assessment of Non-Weatherized Furnace Efficiencies

Source: 2011 Federal Register Technical Support Document: Energy Efficiency Program for Consumer Products: Residential Central Air Conditioners, Heat Pumps, and Furnaces. "Chapter 3. Market and Technology Assessment." Available online: *http://www.regulations.gov/contentStreamer?documentId=EERE-2011-BT-STD-0011-0012&attachmentNumber=4&disposition=attachment&contentType=pdf*

Different assumptions around furnace and distribution efficiency will impact estimates of air sealing savings, which are contingent upon HVAC efficiency and reductions in the level of equipment operation attributed to reduced home leakage.

- **Home configuration**. While billing analysis accounts for the actual home configuration underlying participants' usage, simulation models require an assumption around home configuration for developing prototypes to calculate energy savings. As discussed, the models used to develop the PSD estimate of air sealing savings used a single building prototype; therefore, unless that prototype perfectly represents the participant population, billing analysis results will deviate from those generated by the prototype home simulation. This is particularly relevant when a program treats a subset of the population with home characteristics that might differ broadly from those of the population at large, as in the case of HES-IE. Billing analysis better reflects and accounts for the differences in each participating household, which include assumptions around number of stories per household, and finished versus unfinished basements and attics. Additionally, as the PSD assumed a single-family prototype, there will be inconsistencies in expected savings for any multifamily application, where a single-family prototype would calculate savings assuming a thermal gradient with ambient air rather than with neighboring conditioned spaces.

- **Local weather profiles**. In the R16 billing analysis, participant billing data were weather-normalized based on temperature data from the nearest weather station as determined by participant zip code. Weather stations across state lines were used where they best represented the participant in question. By considering each participant's local weather when assessing savings, the estimates calculated for weather-sensitive measures, including air sealing, more

accurately reflected savings through the programs given the geographic distribution of participants across the state. The PSD offers savings estimates that must be more broadly applicable across programs, and it does not differentiate savings based on participant location; accuracy in developing program-specific savings may therefore be constrained.

Attic Insulation and Wall Insulation

Although Cadmus has found that the PSD's approach for attic and wall insulation incorporates mostly conservative assumptions, algorithms can generally result in overestimation of savings that stems from their many uncertainties. For the R16 evaluation, Cadmus compared aggregated *ex ante* savings, *ex post* savings derived from calibrated building simulations and billing analysis, and billing pre-upgrade consumption data.

Approach Differences

Realization rate drivers and key differences for the wall insulation measure closely mirror those for attic insulation. This is because the *ex ante* and *ex post* methodologies between attic and wall insulation measures were the same for the HES gas program, although the methodology used for HES-IE gas diverged from this approach.

Realization rates for wall insulation measures implemented through the HES-IE program are effected by many of the same factors influencing air sealing realization rates, as the evaluation used billing analysis for its savings estimates in both cases. As with the air sealing measure, behavioral and on-site factors influenced the evaluated savings for HES-IE wall insulation. In its review of insulation measures implemented in Connecticut, the R151 study found that 85% of the insulation jobs assessed on-site were Grade I or Grade II based on RESNET standards, indicating that quality of installation was typically high for insulation measures. However, R151 voices concern over the common use of fiberglass batt insulation, which may reduce the persistence of savings. If materials—and thereby savings—degrade within the first year after installation, evaluated savings derived through billing analysis will similarly decline.

The differing *ex ante* and *ex post* results for the insulation measures most likely derive from differences between the simulation and billing analysis approaches used by the evaluation team and the algorithmic approach employed in the PSD. The evaluation calibrated the simulations used to derive measure-specific savings by adjusting several building characteristics to match the simulation software to the average participant baseline consumption profile. Because billing data shaped the profile, the evaluation's final inputs better reflect participant characteristics and behaviors than the PSD assumptions. A prime example is equipment efficiency.

The PSD assumes one heating and one cooling equipment efficiency for all homes in Connecticut, whereas the R16 evaluation methods (iterative simulation process and billing analysis) rely on actual efficiency data for each population of electric and gas customers in the HES and HES-IE programs. As discussed above, these approaches better represent equipment efficiencies in each program than the PSD approach, which applies the same equipment efficiency assumption to all programs. This is likely to

impact gas savings in particular since efficiencies of furnaces on the market typically exceed the PSD's estimate of furnace efficiency.

Another driver behind the variation in the realization rates is the estimation of heating and cooling loads. For the attic and wall insulation measures, the PSD based all load estimates on average statewide estimates of HDD and CDD between 1979 and 2008; statewide HDD were estimated to be 5,885 °F-day per year. The R16 evaluation, in contrast, applied location-specific weather data to estimate more accurate heating and cooling loads for the participant population. The evaluation found that R16 participants were generally divided between Hartford and Bridgeport weather stations. Upon reviewing TMY3 data for Hartford (Hartford Bradley International Airport) and Bridgeport (Bridgeport Sikorsky Memorial Airport), estimates of Hartford HDD generally aligned with the statewide average from the PSD: 5,964 °F-day, 101% of the statewide average.[67] However, Bridgeport participants made up 37% to 50% of the population depending on program and fuel type, and had 5,524 °F-day: 6% fewer annual HDDs than the statewide average. Furthermore, 2011 saw substantially fewer HDD than normal in both locations, with 4,984 °F-day in Bridgeport and 5,680 °F-day in Hartford.[68] A reduction in HDD, and consequently heating load, in turn reduces *ex post* savings but better represents the savings achieved through the program. In its modeling effort, the evaluation's choice to employ different weather data for homes in Hartford and Bridgeport led to a difference in calculated gross savings, but is a more accurate approach than assuming a single weather profile for load calculations. When billing analysis was used, consumption was weather-normalized based on each customer's location, also improving the accuracy of results.

Another likely reason for the low realization rates of both measures is that wall and attic insulation measures are highly interactive. When more than one measure is implemented, particularly when HVAC improvements are made, this effect can be pronounced. The Connecticut PSD does not account for interaction factors between measures, so if the *ex ante* calculations follow PSD algorithms, the savings will be overstated. The R16 evaluation approach (billing analysis and simulation modeling) incorporated the interactive nature of attic insulation, wall insulation, and other building characteristics, and therefore resulted in smaller gross savings.

The combined effect of attic and wall insulation upgrades results in reduced savings compared to the sum of each individual measure's impact. For homes in the program that implemented both measures, the PSD calculates the savings for each measure assuming that each measure had the same pre-upgrade heating and cooling loads. A more accurate algorithm adjusts the savings for each measure when other interactive measures are implemented, basing the savings for one measure on the reduced heating and cooling load for the home resulting from implementation of other measures. This holds true not only for

---

[67] 1991–2005 Update: Typical Meteorological Year 3, Updated TMY3 Files, Individual Station Files. National Solar Radiation Data Base. Available online: http://rredc.nrel.gov/solar/old_data/nsrdb/1991-2005/tmy3/by_state_and_city.html.

[68] 1997–current Degree Day Statistics. National Oceanic and Atmospheric Administration. Available online: ftp://ftp.cpc.ncep.noaa.gov/htdocs/products/analysis_monitoring/cdus/degree_days/archives/Heating%20degree%20Days/Monthly%20City/2011/.

insulation measures, but for duct- and air-sealing measures as well. Without using a methodology that accounts for these interactions, the PSD will overestimate envelope savings whenever multiple HVAC or envelope measures are combined.

Measure Assumptions

In general, for wall and attic insulation measures the PSD incorporates conservative assumptions for its savings algorithms.

Figure 17 illustrates the adjustment factors added to the nominal R-values of the attic and wall insulation measures. The PSD adds these adjustment factors to the existing and new R-values in the PSD algorithm in order to account for the additional thermal resistance provided by the attic and wall construction. For small nominal R-values (typically the pre-upgrade condition), the adjustment factors are more positive than the larger nominal R-values (typically the post-upgrade condition). The nature of these adjustment factors will systematically cause more conservative savings than if the algorithm used nominal R-values.

Figure 17. PSD R-Value Adjustments as a Function of Nominal Value
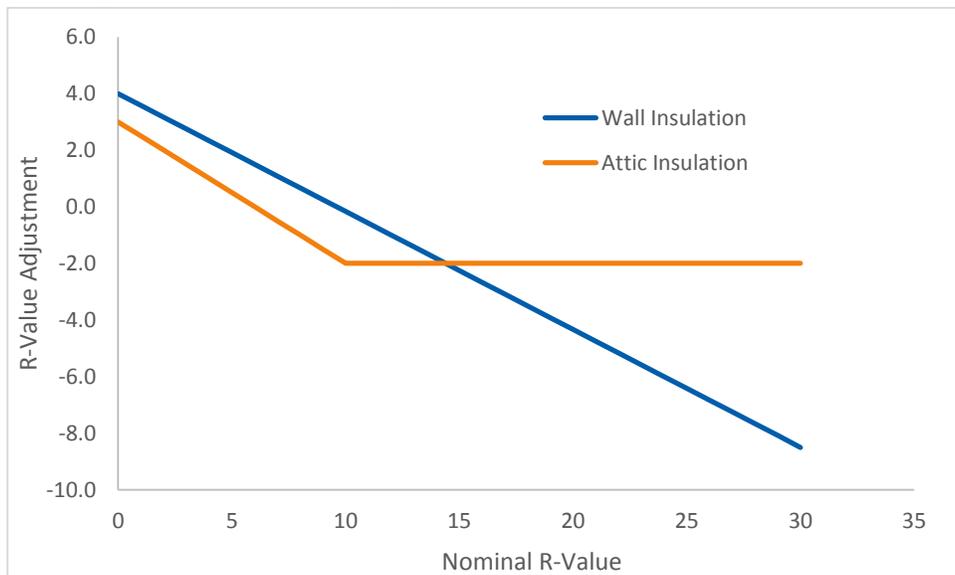
The 75% system efficiency assumption for fossil fuel heating reflects an engineering estimate, and requires further substantiation from a current source. The federal standard and market conditions in PY2011 exceed 80% AFUE for all except for mobile home furnaces (see Figure 16), and degradation over the age of a gas unit is minimal according to a 2015 NREL study.[69] According to RECS data for the New England region (excluding Massachusetts), the average home's heating equipment is approximately 13

---

[69] Brand, L., S. Yee, and J. Baker. Improving Gas Furnace Performance: A Field and Laboratory Study at End of Life. National Renewable Energy Laboratory. NREL Contract No. DE-AC36-08GO28308. February 2015. http://www.nrel.gov/docs/fy15osti/63702.pdf.

years old, with only 22% of homes having a system newer than five years old; the PSD also stipulates a 20-year estimated useful life for gas furnaces. While the federal standards for residential furnaces in place prior to 2015 first became effective in 1992, the market conditions under which these older units were installed may have offered lower efficiencies, on average, than were available in 2011. The extent to which distribution losses further reduce the system efficiency should be assessed using recent and, if possible, Connecticut-specific sources.

When a program population is assumed to have less efficient units, this assumption may cause an underestimate of savings. Furthermore, static efficiency assumptions are inherently imprecise. Efficiency of HVAC units vary during operation; variables such as outside air temperature, return air and water temperature, and heat exchanger soiling all affect efficiency during the year. The PSD algorithm does not account for these variations as a billing analysis or building simulation would. Billing analyses also incorporate unknown comfort-related behaviors (e.g., take back), as discussed above.

One assumption that requires further review is the HDD adjustment factor. For heating savings, the PSD uses HDDs to estimate the temperature difference across the building envelope in the winter. By definition, HDDs use a balance-point reference temperature to assume the outdoor air temperature at which the building does not require an HVAC load. For most residential applications and for this PSD algorithm, this balance point temperature is 65 °F (HDD base 65 °F). By using HDDs in these savings algorithms, the PSD assumes that the 65 °F balance point is the average indoor temperature in the winter. The PSD then multiplies HDD by a correction factor of 62%, in part to account for the fact that a heating system will not run at all times that the outside temperature falls below 65 °F. The adjustment factor may also account for other elements that could reduce the HDD, such as thermal gains. This factor is cited in the 1989 ASHRAE handbook, but is not referenced in newer versions, and this study could not verify how it was derived due to unavailability of the original source. Therefore, the study could not corroborate this factor or determine whether it is a reasonable correction to the degree-day estimate for Connecticut.

Upon further review, Cadmus found that the Ohio, Pennsylvania, and Mid-Atlantic TRMs use a similar factor when determining cooling savings, but no factor is applied to heating savings calculations, although the Mid-Atlantic assumes a reduced heating set point of 60 °F when calculating HDD. For cooling savings, these TRMs cite a 2008 study[70] by the Energy Center of Wisconsin that applies a 75% correction factor called the discretionary use adjustment (DUA) to account for the fact that air conditioners are unlikely to run for every hour the outside temperature is above the base temperature. However, these TRMs do not apply any degree-day adjustment to HDDs for heating savings, and the evaluation team cannot validate the assumed 62% correction factor; its vintage, however, suggests that a re-evaluation of this factor is needed.

---

[70]   Energy Center of Wisconsin. *Central Air Conditioning in Wisconsin: A Compilation of Recent Field Research*. May 2008, amended December 15, 2010. Available online: http://www.ecw.org/sites/default/files/241-1_0.pdf.

Future Recommendations

The R91 found during its review that differences in the results of the PSD and the R16 evaluation for the air sealing, duct sealing, attic insulation, and wall insulation measures stemmed from a variety of sources. In some cases, the review revealed differences simply in the approaches appropriate for developing *ex ante* and *ex post* savings; for example, while billing analysis can present a robust assessment of savings realized through program participation, this methodology is not suitable for developing savings to be claimed through a PSD. However, the team noted several areas where adjustments may improve future estimates of *ex ante* savings. To better align savings calculations with the best practices identified in Section 1 of this report, this report recommends reviewing the following parameters or approaches:

- **Update simulation models for air and duct sealing**. Revise models to use an hourly-iterative simulation software and draw upon participant home characteristics, differentiating between different building, customer, and HVAC types to award the most appropriate savings. Calibrate model prototypes to participant data to ensure that typical consumption patterns of Connecticut customers are reflected in savings computations. In future evaluations, ensure evaluators and PSD developers use an hourly-iterative software package that uses default assumptions and load shapes that are appropriate for residential applications (e.g., BEopt).

- **Differentiate savings values based on population segment.** Certain population segments may not be reflected accurately by the savings developed for an average participant home in the PSD, such as multifamily customers and the lower-income participants in the HES-IE program. Although the air infiltration measure does adapt savings for multifamily customers, the other measures reviewed do not contain a similar adjustment. By adjusting simulation or algorithm inputs and permitting appropriate savings to be awarded specific to these population segments, accuracy of the program-wide *ex ante* savings calculation may be improved.

- **Account for interactivity between HVAC and envelope measures**. Individual measure savings are lowered if installed concurrently; for example, performing duct sealing increases distribution efficiency so that if attic insulation is then installed, heating load drops by a much smaller amount than it would if ducts remained leaky. To account for this interactivity, make an adjustment to reduce savings when multiple shell- or duct-improvement measures are implemented through the program.

- **Consider whether additional weather and location assumptions can improve savings estimates**. The PSD currently uses only a single weather profile to estimate weather patterns that influence savings, which may not reflect the geographic distribution of participants across the state. Areas where a large number of participants are identified (e.g., Bridgeport) have notably lower HDDs than reflected by the statewide average or Hartford weather profiles.

- **Verify that heating HVAC efficiency assumptions remain valid.** Current HVAC system efficiency assumptions rely on estimates that should be validated, given the sensitivity of savings to efficiency values. If system efficiency assumption are found to be low for the participating population, savings may be overestimated. Lower furnace efficiencies require greater HVAC

energy consumption to meet winter setpoint temperatures; therefore, measures such as insulation, air sealing, and duct sealing, which reduce heating load, have an amplified effect. Furnace efficiency assumptions influence savings calculated both through building simulation and through the algorithmic approach applied for insulation measures.

- **Assess whether the HDD adjustment factor for insulation measures should be updated.** For attic and wall insulation savings, the current HDD correction factor, which draws from ASHRAE's 1989 handbook, could not be validated with a more current source. An updated value is not provided in more recent versions of this handbook. Provide transparency in what this value seeks to represent.

Conclusions

In reviewing the different approaches used to develop savings estimates in the PSD and in the R16 impact evaluation, the R91 study found meaningful differences in the methods that underlie the observed differences between *ex ante* and *ex post* savings estimates. Except in the case of the duct sealing measure, the R16 evaluation employed a different methodology than was used in the PSD; the divergent aims and applications for each methodology, outlined above in Methodology Strengths and Weaknesses, produced much of the difference between the evaluated and claimed savings values.

The evaluation based its savings estimates on actual changes in participant consumption arising from program participation in PY2011, either through billing analysis or model calibration. Evaluators can use both of these methodologies to examine a specific, retrospective year in order to obtain a relatively accurate estimate of savings for the programs in question. Conversely, the PSD employs more generic building simulations and algorithmic approaches in order to offer an adaptable and simple forward-looking savings calculation for any prescriptive Connecticut program. While the methodologies used both by the evaluation and the PSD are generally appropriate to their application and constraints, they will produce divergent *ex ante* and *ex post* savings values.

The R91 study identified several areas for improvement within the PSD that may lead to greater accuracy in PSD savings estimates and closer alignment with evaluated savings, as discussed in the Future Recommendations section above. Particularly where the methodology used in the evaluation and in the PSD is the same—building simulation, which was used to calculate savings for duct sealing measures—the differences in implementation of this approach are notable and point to potential methodological adjustments in line with the best practices discussed in Section 1 of this report (Section 1: Best Practices in Impact Evaluation). While PSD and evaluation estimates of measure savings may have different constraints, the PSD nevertheless aims to provide a tool with which to accurately estimate savings arising from measure installation; to do so, measure interactivity must be considered and periodic updates are critical to ensure that assumptions in both algorithms and building simulations correspond with the participant population.

References

Berry, L., and M. Schweitzer. *Metaevaluation of National Weatherization Assistance Program Based on State Studies, 1993–2002.* February, 2003. Available online: http://weatherization.ornl.gov/pdfs/ORNL_CON-488.pdf.

Brand, L., S. Yee, and J. Baker. Improving Gas Furnace Performance: A Field and Laboratory Study at End of Life. *National Renewable Energy Laboratory.* NREL Contract No. DE-AC36-08GO28308. February 2015. http://www.nrel.gov/docs/fy15osti/63702.pdf.

Brown et al. AMI System Security Requirements v1.01. December 2008. Prepared by the AMI-SEC Task Force. Available online: http://energy.gov/sites/prod/files/oeprod/DocumentsandMedia/14-AMI_System_Security_Requirements_updated.pdf.

California Public Utilities Commission (CPUC). *California Energy Efficiency Evaluation Protocols: Technical, Methodological, and Reporting Requirements for Evaluation Professionals.* Prepared by The TecMarket Works Team. April 2006. Available online: http://www.calmac.org/publications/EvaluatorsProtocols_Final_AdoptedviaRuling_06-19-2006.pdf.

California Public Utilities Commission, Rulemaking 07-01-041. *Decision Adopting Protocols for Estimating Demand Response Load Impacts*. *January 2007.* Available Online: http://www.calmac.org/events/FinalDecisionLoadImpact.pdf http://www.calmac.org/events/FinalDecision_AttachementA.pdf

Crossman, K., L. Tabor, M. Perussi, and D. Basak. *Dynamic Duo: How Combining Billing Analysis and Engineering Simulation Methods Improves Evaluation Quality and Understanding.* Proceedings of the 2013 International Energy Program Evaluation Conference, Chicago.

Department of Energy, Office of Energy Efficiency and Renewable Energy. 2011 Federal Register Technical Support Document: Energy Efficiency Program for Consumer Products: Residential Central Air Conditioners, Heat Pumps, and Furnaces. Chapter 3. Market and Technology Assessment. Prepared by Navigant Consulting, Inc., Lawrence Berkeley National Laboratory, and Pacific Northwest National Laboratory. June, 2011. Available online: http://www.regulations.gov/contentStreamer?documentId=EERE-2011-BT-STD-0011-0012&attachmentNumber=4&disposition=attachment&contentType=pdf.

Dimetrosky, S., K. Parkinson, and N. Lieb. 2014. "Chapter 6: Residential Lighting Evaluation Protocol." The Uniform Methods Project: Methods for Determining Energy Efficiency Savings for Specific Measures. Available Online: http://www.nrel.gov/extranet/ump/pdfs/20140514_ump_res_lighting_draft.pdf

Efficiency Maine Trust. *Residential Technical Reference Manual: Version 2016.1.* July 1, 2015. Available online: http://www.efficiencymaine.com/docs/EMT-TRM_Residential_v2016_1.pdf.

Efficiency Valuation Organization (EVO). *International Performance Measurement & Verification Protocol, Volume I: Concepts and Options for Determining Savings*. January 2012. Available online: http://www.evo-world.org/ipmvp.php

Efficiency Vermont. *Technical Reference User Manual (TRM): Measure Savings Algorithms and Cost Assumptions.* March 16, 2015. Available online: http://psb.vermont.gov/sites/psb/files/docketsandprojects/electric/majorpendingproceedings/TRM%20User%20Manual%20No.%202015-87C.pdf.

Energy Center of Wisconsin. *Central Air Conditioning in Wisconsin: A Compilation of Recent Field Research.* May 2008, emended December 15, 2010. Available online: http://www.ecw.org/sites/default/files/241-1_0.pdf.

Goldberg, M., and G. Agnew. February 2013. "Measurement and Verification for Demand Response." Prepared for the National Forum on the National Action Plan on Demand Response: Measurement and Verification Working Group. Available Online: https://emp.lbl.gov/sites/all/files/napdr-measurement-and-verification_0.pdf

Gupta, S. *EIA Energy Conference 2014 ElectriSense Presentation.* Available online: http://www.eia.gov/conference/2014/pdf/presentations/gupta.pdf.

IBM Whitepaper. *Managing big data for smart grids and smart meters.* Available online: http://www-935.ibm.com/services/multimedia/Managing_big_data_for_smart_grids_and_smart_meters.pdf.

International Building Performance Simulation Association. *BEST Directory 2015*. Available Online: http://www.buildingenergysoftwaretools.com.

Judkoff, R., and J. Nymark. 1998. "The BESTEST Method for Evaluating and Diagnosing Building Energy Software." *The American Council for an Energy-Efficient Economy (ACEEE).* Available Online: http://aceee.org/files/proceedings/1998/data/papers/0515.PDF.

Kushler, M., S. Nowak, and P. Witte. (February 2012). "A National Survey of State Policies and Practices for the Evaluation of Ratepayer- Funded Energy Efficiency Programs." Report Number U122. *American Council for an Energy- Efficient Economy (ACEEE).* www.aceee.org/research-report/u122.

Laughman et al. "Power Signature Analysis." *ICEE Power and Energy*. April 2003. Available online: http://www.cse.psu.edu/~sem284/cse598e-f11/papers/laughman.pdf.

Massachusetts Electric and Gas Energy Efficiency Program Administrators. *Massachusetts Technical Reference Manual for Estimating Savings from Energy Efficiency Measures.* June 2015. Available online: http://web1.env.state.ma.us/DPU/FileRoomAPI/api/Attachments/Get/?path=15-49%2FCMA_Appendix_3_Tech_Ref_Manual.pdf.

Murrill, B., E. Liu, and R. Thompson II. Smart Meter Data: Privacy and Cybersecurity. Congressional Research Service Report for Congress. Prepared for Members and Committees of Congress. February 2012. Available online: https://www.fas.org/sgp/crs/misc/R42338.pdf.

National Renewable Energy Laboratory, U.S. Department of Energy (NREL). *The Uniform Methods Project: Methods for Determining Energy Efficiency Savings for Specific Measures*. April 2013. Available online: http://www.nrel.gov/extranet/ump.

New York State Department of Public Service. *New York Standard Approach for Estimating Energy Savings from Energy Efficiency Programs – Residential, Multi-family, and Commercial/Industrial Measures: Version 2.* December 10, 2014. Available online: http://www3.dps.ny.gov/W/PSCWeb.nsf/ca7cd46b41e6d01f0525685800545955/06f2fee55575bd8a852576e4006f9af7/$FILE/TRM%20Version%202%20December%2010,%202014.pdf.

NMR Group, Inc. Connecticut HES Air Sealing, Duct Sealing, and Insulation Practices Report (R151) Draft. December 2015. Available online: http://www.energizect.com/sites/default/files/CT%20HES%20Air%20Sealing%2C%20Duct%20Sealing%2C%20and%20Insulation%20Practices%20%28R151%29_review%20draft_12.23.15.docx.

Northwest Regional Technical Forum (RTF). *Roadmap for the Assessment of Energy Efficiency Measure*. June 2014. Available online: http://www.nwcouncil.org/energy/rtf/Default.htm.

Oak Ridge National Laboratory. *National Evaluation of the Weatherization Assistance Program: Preliminary Evaluation Plan for Program Year 2006.* February, 2006. Available online: http://weatherization.ornl.gov/Retrospectivepdfs/ORNL_CON-498.pdf.

Oak Ridge National Laboratory. *National Weatherization Assistance Program Impact Evaluation: Energy Impacts for Single Family Homes.* September 2014. Available online: http://weatherization.ornl.gov/Retrospectivepdfs/ORNL_TM-2015_13.pdf.

Opinion Dynamics Corporation. *Delaware Department of Natural Resources and Environmental Control SEU and State Energy Efficiency Programs EM&V Report*. July 2012. Available online: http://www.dnrec.delaware.gov/energy/information/otherinfo/Documents/EM-and-V-guidance-documents/DNREC%20EMV%20REPT%20_FinalRev_2013.pdf.

State and Local Energy Efficiency Action Network (SEE Action). *Energy Efficiency Program Impact Evaluation Guide*. Prepared by Steven R. Schiller, Schiller Consulting, Inc. December 2012. Available online: http://www.seeaction.energy.gov.

State and Local Energy Efficiency Action Network (SEE Action). *Evaluation, Measurement, and Verification (EM&V) of Residential Behavior-Based Energy Efficiency Programs: Issues and Recommendations*. Prepared by Annika Todd, Elizabeth Stuart, and Charles Goldman of Lawrence Berkeley National Laboratory, and Steven R. Schiller, Schiller Consulting, Inc. May 2012. Available online: https://www4.eere.energy.gov/seeaction/system/files/documents/emv_behaviorbased_eeprograms.pdf

Stewart, J., and A. Todd. 2015. "Chapter 17: Residential Behavior Protocol." The Uniform Methods Project: Methods for Determining Energy Efficiency Savings for Specific Measures. Available Online: http://energy.gov/sites/prod/files/2015/02/f19/UMPChapter17-residential-behavior.pdf

Tabor, L., et al. *High Efficiency Heating Equipment Impact Evaluation.* March, 2015. Available online: http://www.neep.org/sites/default/files/resources/High-Efficiency-Heating-Equipment-Impact-Evaluation-Final-Report.pdf.

The Cadmus Group, Inc., and NMR Group, Inc. *Impact Evaluation: Home Energy Services–Income-Eligible and Home Energy Services Programs (R16)*. December 2014. Available online: http://www.energizect.com/your-town/hes-and-hes-ie-impact-evaluation-r16-final-report-12-31-14.

The Cadmus Group, Inc., and NMR Group, Inc. *Impact Evaluation: New Hampshire Home Performance with ENERGY STAR® Program.* June 13, 2011. Available online: http://www.puc.state.nh.us/electric/Monitoring%20and%20Evaluation%20Reports/124%20NH%20HPwES%20Impact%20Evaluation%20Report%20June%2013%202011.pdf.

The Cadmus Group, Inc. *Home Energy Services Impact Evaluation.* August 2012. Available online: http://ma-eeac.org/wordpress/wp-content/uploads/Home-Energy-Services-Impact-Evaluation-Report_Part-of-the-Massachusetts-2011-Residential-Retrofit-and-Low-Income-Program-Area-Evaluation.pdf.

The Cadmus Group, Inc. *Impact and Process Evaluation of Efficiency Maine Trust's Retro-Commissioning Pilot Program*. December 28, 2012. Available online: http://www.efficiencymaine.com/docs/Efficiency-Maine-RCx-Evaluation-Report_PDFFinal.pdf.

The Cadmus Group, Inc*. Low Income Single Family Program Impact Evaluation.* June 2012. Available online: http://ma-eeac.org/wordpress/wp-content/uploads/Low-Income-Single-Family-Program-Impact-Evaluation_Part-of-the-Massachusetts-Residential-Retrofit-Low-Income-Program-Area-Evaluation.pdf.

The United Illuminating Company and Connecticut Light & Power Company. *Connecticut Program Savings Document*. November 5, 2014. Available online:

http://www.dpuc.state.ct.us/DEEPEnergy.nsf/c6c6d525f7cdd1168525797d0047c5bf/8525797c0
0471adb85257db600666127/$FILE/2015%20Program%20Savings%20Documentation%20.pdf

U.S. Department of Energy, Energy Efficiency and Renewable Energy. "Residential Furnaces." Available
online:
https://www1.eere.energy.gov/buildings/appliance_standards/product.aspx/productid/72.

United States Census Bureau. *American Community Survey (ACS).* Available online: American Community
Survey (ACS).

United States Energy Information Administration. *Residential Energy Consumption Survey (RECS).*
http://www.eia.gov/consumption/residential/.

Appendix A: IPMVP M&V Option Selection Flowchart

Figure 18. IPMVP M&V Option Selection Flowchart
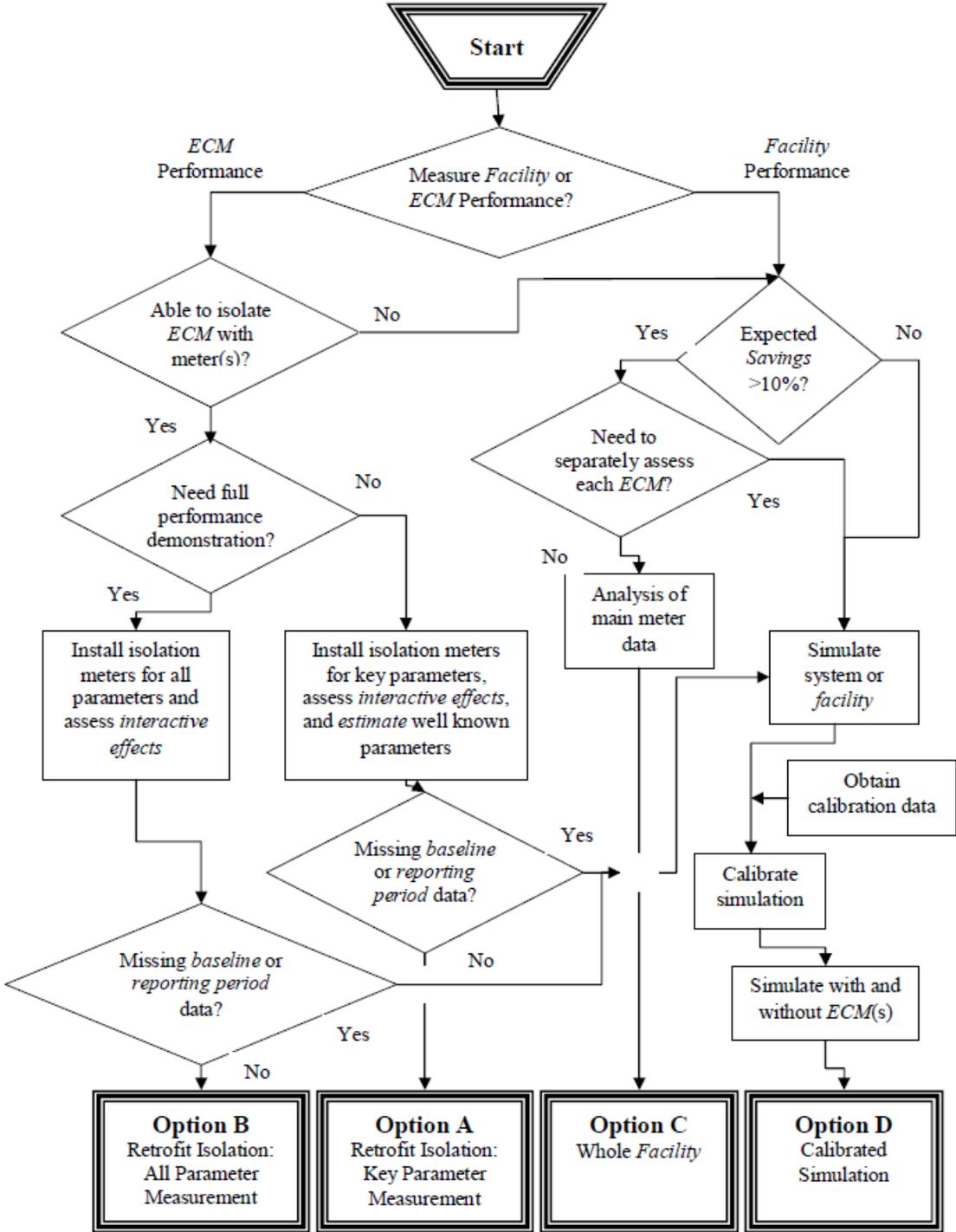
Source: Efficiency Valuation Organization (EVO). International Performance Measurement &Verification Protocol, Volume I: Concepts and Options for Determining Savings. January 2012. p. 33.

Appendix B: Suggested IPMVP M&V Options for Different Projects

Figure 19. Suggested IPMVP M&V Options for Different Projects

ECM Project Characteristic	Suggested Option			
	A	**B**	**C**	**D**
Need to assess *ECM*s individually	X	X		X
Need to assess only total facility performance			X	X
Expected *savings* less than 10% of utility meter	X	X		X
Significance of some *energy* driving variables is unclear		X	X	X
Interactive effects of ECM are signficant or unmeasurable			X	X
Many future changes expected within *measurement boundary*	X			X
Long term performance assessment needed	X		X	
Baseline data not available				X
Non-technical persons must understand reports	X	X	X	
Metering skill available	X	X		
Computer simulation skill available				X
Experience reading utility bills and performing regression analysis available			X	

Source: Efficiency Valuation Organization (EVO). International Performance Measurement & Verification Protocol, Volume I: Concepts and Options for Determining Savings. January 2012. p. 34.

Appendix C: SEE Action: Applications for Each IPMVP M&V Option

Figure 20. SEE Action: Applications for Each IPMVP M&V Option

OPTION A Retrofit Isolation – Key Parameter Measurement is best applied where:	OPTION B Retrofit Isolation – All Parameters Measurement is best applied where:	OPTION C Whole Facility is best applied where:	OPTION D Calibrated Simulation is best applied where:
• The magnitude of savings is low for the entire project or for the portion of the project to which Option A is applied • The project is simple with limited independent variables and unknowns • The risk of not achieving savings is low • Interactive effects are to be ignored or are stipulated using estimating methods	• The project involves simple equipment replacements • Energy savings values per individual measure are desired • Interactive effects are to be ignored or are stipulated using estimating methods • Independent variables are not complex	• The project is complex • Predicted savings are large (typically greater than 10%) compared to the recorded energy use • Energy savings values per individual measure are not needed • Interactive effects are to be included • Independent variables that affect energy use are not complex or excessively difficult to monitor	• New construction projects are involved • Energy savings values per measure are desired • Option C tools cannot cost-effectively evaluate particular measures • Complex baseline adjustments are anticipated • Baseline measurement data do not exist or are prohibitively expensive to collect

Source: Efficiency Valuation Organization (EVO). (2010). International Performance Measurement and Verification Protocol. IPMVP, EVO 10000—1:2010. www.evo-world.org.

## Appendix D: UMP Recommended Consumption Data Analysis Form

Figure 21: UMP Recommended Consumption Data Analysis Form

Program Condition	Consumption Data Analysis Form	Comparison Group	Gross or Net Savings	Unknown Biases
1. Randomized controlled trial experimental design	Two-stage or pooled	Randomly selected control group	Net	Spillover, if it exists
2. Stable program and target population over multiple years	Two-stage	Prior and future participants	Gross	Minimal
3. Participation staggered over at least one full year	Pooled	None: pooled specification with participants only	Gross	Minimal
4. Not randomized, not stable over multiple years, participants similar to general eligible population, nonparticipant spillover minimal	Two-stage or pooled	Matched comparison group	Likely between gross and net	Self-selection[6] and spillover
5. Not randomized, not stable over multiple years, participants unlike general eligible population, nonparticipant spillover minimal	Two-stage or pooled	General eligible nonparticipants	Likely between gross and net	Self-selection and spillover

Source: National Renewable Energy Laboratory, U.S. Department of Energy (NREL). The Uniform Methods Project: Methods for Determining Energy Efficiency Savings for Specific Measures. April 2013. Available online: *http://www.nrel.gov/extranet/ump*. p. 8–9.

Appendix E: RTF Roadmap for Assessment of Energy Efficiency Measures: Selecting a Method for Savings Estimation

Figure 22. RTF Roadmap: Selecting a Method for Savings Estimation
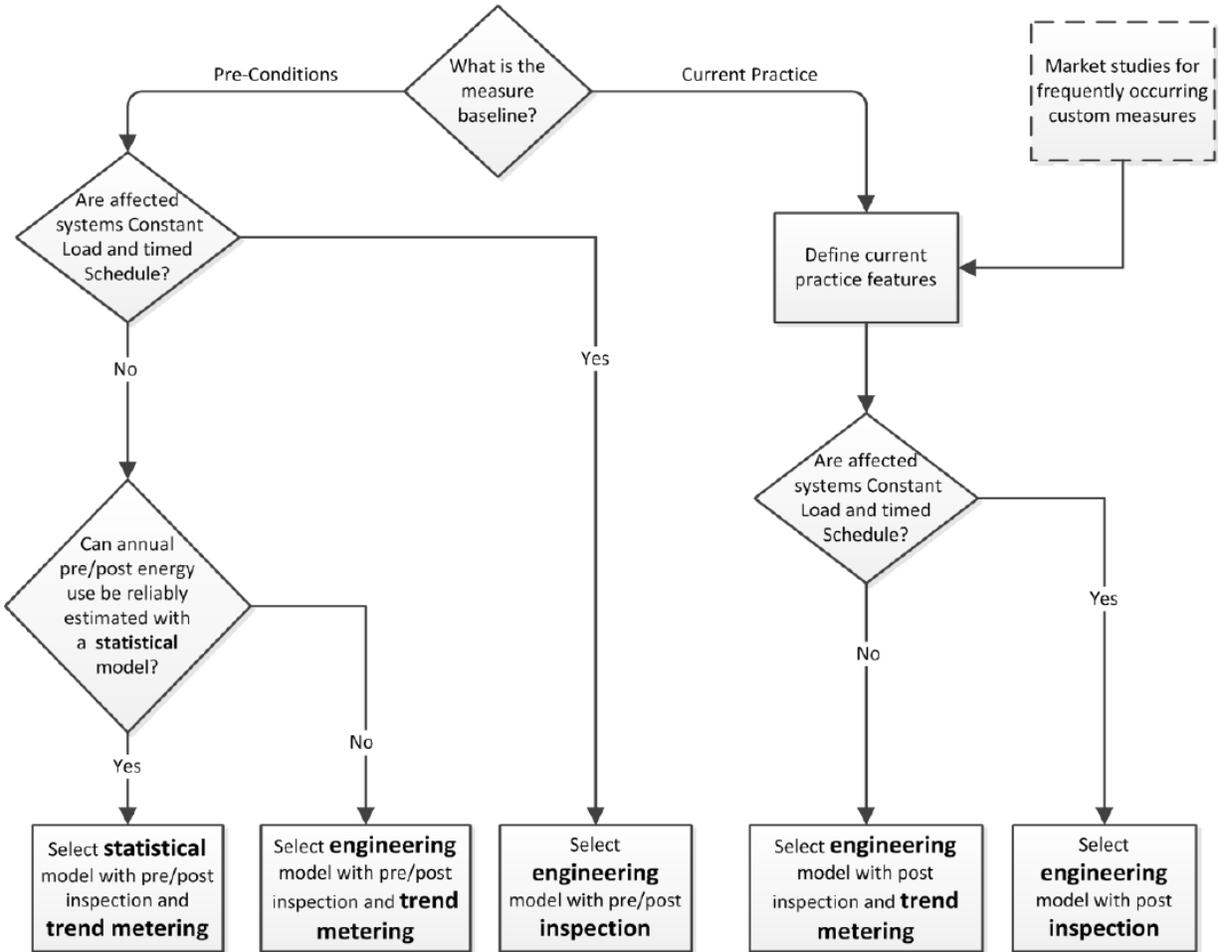
Source: Northwest Regional Technical Forum (RTF). Roadmap for the Assessment of Energy Efficiency Measures. June 2014. Available online: http://www.nwcouncil.org/energy/rtf/Default.htm. p. 35.