



Report Title (Arial 22pt Bold)

Date — Date (Period of Performance, Arial 18pt Bold)

A. Author, B. Author2, and C. Author3 (Arial 16pt)

Company Name

City, State – not abbreviated (Arial 14pt Italic)

A. Author, B. Author2, and C. Author3 (Arial 16pt)

Second Company Name

City, State – not abbreviated (Arial 14pt Italic)

NREL Technical Monitor:

**NREL is a national laboratory of the U.S. Department of Energy
Office of Energy Efficiency & Renewable Energy
Operated by the Alliance for Sustainable Energy, LLC**

This report is available at no cost from the National Renewable Energy Laboratory (NREL) at www.nrel.gov/publications.

Subcontract Report (Arial 11 pt Bold)

NREL/SR-xxxx-xxxxx

Month Year (Arial 11 pt)

Contract No. DE-AC36-08GO28308

Report Title (Arial 22pt Bold)

Date — Date (Period of Performance, Arial 18pt Bold)

A. Author, B. Author2, and C. Author3 (Arial 16pt)

Company Name

City, State – not abbreviated (Arial 14pt Italic)

A. Author, B. Author2, and C. Author3 (Arial 16pt)

Second Company Name

City, State – not abbreviated (Arial 14pt Italic)

NREL Technical Monitor: Name

Prepared under Subcontract No(s). ZAT-6-15179-02, TAM-7-16454-01

DRAFT: NOT FOR CITATION

**NREL is a national laboratory of the U.S. Department of Energy
Office of Energy Efficiency & Renewable Energy
Operated by the Alliance for Sustainable Energy, LLC**

This report is available at no cost from the National Renewable Energy Laboratory (NREL) at www.nrel.gov/publications.

Subcontract Report (Arial 11 pt Bold)

NREL/SR-xxxx-xxxxx

Month Year (Arial 11 pt)

NOTICE

This report was prepared as an account of work sponsored by an agency of the United States government. Neither the United States government nor any agency thereof, nor any of their employees, makes any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States government or any agency thereof. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States government or any agency thereof.

This report is available at no cost from the National Renewable Energy Laboratory (NREL) at www.nrel.gov/publications.

Available electronically at <http://www.osti.gov/scitech>

Available for a processing fee to U.S. Department of Energy and its contractors, in paper, from:

U.S. Department of Energy
Office of Scientific and Technical Information
P.O. Box 62
Oak Ridge, TN 37831-0062
phone: 865.576.8401
fax: 865.576.5728
email: <mailto:reports@adonis.osti.gov>

Available for sale to the public, in paper, from:

U.S. Department of Commerce
National Technical Information Service
5285 Port Royal Road
Springfield, VA 22161
phone: 800.553.6847
fax: 703.605.6900
email: orders@ntis.fedworld.gov
online ordering: <http://www.ntis.gov/help/ordermethods.aspx>

Cover Photos: (left to right) photo by Pat Corkery, NREL 16416, photo from SunEdison, NREL 17423, photo by Pat Corkery, NREL 16560, photo by Dennis Schroeder, NREL 17613, photo by Dean Armstrong, NREL 17436, photo by Pat Corkery, NREL 17721.



Printed on paper containing at least 50% wastepaper, including 10% post consumer waste.

Acknowledgments

~~FOURTH DRAFT (12-26-2013) – NOT TO BE CITED~~

~~Estimating Net Energy Saving: Methods and Practices~~

~~*Daniel M. Violette, Ph.D., Navigant Consulting, Inc.*~~

~~*Pam Rathbun, Tetra Tech*~~

~~DRAFT: NOT FOR CITATION~~

~~FOURTH DRAFT – NOT TO BE CITED~~

Table of Contents

Acknowledgements	ii
Estimating Net Energy Savings	3
1—Universality of the Net Impacts Challenge.....	3
2—Defining Gross and Net Savings for Practical Evaluation.....	4
2.1 Definition of Gross and Net Savings.....	4
2.2 Definitions of Factors Used in Net Savings Calculations.....	4
2.3 Uses of Net Savings Estimates in the EE Industry.....	8
2.4 The Net Savings Estimation Challenge—Establishing the Baseline.....	9
3—Methods for Net Savings Estimation.....	10
3.1 Randomized Controlled Trials and Quasi Experimental Designs.....	11
3.1.1—Randomized Control Trials.....	12
3.1.2—Quasi-Experimental Designs.....	16
3.2.1—Program Participant Surveys.....	23
3.2.2—Surveys of Program Nonparticipants.....	28
3.2.3—Market Actor Surveys.....	28
3.2.4—Case Studies for Estimating Net Savings Using Survey Approaches.....	29
3.3 Common Practice Baseline Approaches.....	35
3.4 Market Sales Data Analyses (Cross-Sectional Studies).....	40
3.5 Top-Down Evaluations (Macroeconomic Models).....	42
3.6 Structured Expert Judgment Approaches.....	48
3.7 Deemed or Stipulated NTG Ratios.....	50
3.8 Historical Tracing (or Case Study) Method.....	52
4—Conclusions and Recommendations.....	53
4.1 A Layered Evaluation Approach.....	53
4.2 Selecting the Primary Estimation Method.....	54
4.3 Methods Applicable for Different Conditions.....	57
4.4 Planning Net Savings Evaluations—Issues to be Considered.....	60
4.5 Trends and Recommendations in Estimating Net Savings.....	61
Appendix A: Price Elasticity Studies as a Component of Upstream Lighting Net Savings Studies	63
References	66

Acknowledgements

The chapter authors wish to thank and acknowledge the Uniform Methods Project Steering Committee and Net-to-Gross Technical Advisory Group members for their contributions to this chapter. The following people offered valuable input to the development of this chapter by providing subject-related materials, in-depth discussion, and careful review of draft versions:

- Michael Li, DOE
- Chuck Kurnik, NREL
- Michael Rufo, Itron
- Hossein Haeri, M. Sami Khawaja, Josh Keeling, Alexandra Rekkas, and Tina Jayaweera, Cadmus
- Tom Eckman, Northwest Power Planning Council

DRAFT #4 | Not for Citation

- Elizabeth Titus, Northeast Energy Efficiency Partnerships
- Steve Schiller, Schiller Consulting
- Rick Ridge, Ridge & Associates
- Ralph Prah, Ralph Prah & Associates
- Jane Peters [and Marjorie McRae](#), Research Into Action, Inc.
- Ken Seiden and Jeff Erickson, Navigant
- Lynn Hoefgen, NMR Group, Inc.
- Nick Hall, TecMarket Works

- [Miriam Goldberg, DNV GL](#)
- [Peter Miller, Natural Resources Defense Council](#)

Teri Lutz of Tetra Tech ~~worked~~ [made substantive contributions](#) across the entire chapter ~~to provide technical review, additions, and edits.~~

Definitions

<u>C&I</u>	<u>Commercial and industrial</u>
<u>CFL</u>	<u>Compact fluorescent lamp</u>
<u>DiD</u>	<u>Difference-in-differences</u>
<u>EE</u>	<u>Energy efficiency</u>
<u>FR</u>	<u>Free ridership</u>
<u>HER</u>	<u>Home Energy Report</u>
<u>IOU</u>	<u>Investor-owned utility</u>
<u>kWh</u>	<u>Kilowatt-hours</u>
<u>LFER</u>	<u>Linear fixed-effects regression</u>
<u>MCM</u>	<u>Macroconsumption metric</u>
<u>ME</u>	<u>Market Effects</u>
<u>NEEA</u>	<u>Northwest Energy Efficiency Alliance</u>
<u>NTG</u>	<u>Net-to-gross (ratio)</u>
<u>NW Council</u>	<u>Northwest Power and Conservation Council</u>
<u>RCT</u>	<u>Randomized control trial</u>
<u>RDD</u>	<u>Regression discontinuity design</u>
<u>RED</u>	<u>Random encouragement design</u>
<u>RTF</u>	<u>Regional Technical Forum</u>
<u>SMUD</u>	<u>Sacramento Municipal Utilities District</u>
<u>SO</u>	<u>Spillover</u>

DRAFT: NOT FOR CITATION

Contents

DRAFT: NOT FOR CITATION

Acknowledgments	
Estimating Net Energy Savings	1
1 Universality of the Net Impacts Challenge	2
2 Defining Gross and Net Savings for Practical Evaluation	3
2.1 Definition of Gross and Net Savings	3
2.2 Definitions of Factors Used in Net Savings Calculations	3
2.2.1 Free Ridership	3
2.2.2 Spillover	4
2.2.3 Market Effects	4
2.2.4 Net Savings Equations	5
2.3 Uses of Net Savings Estimates in the Energy Efficiency Industry	8
2.4 The Net Savings Estimation Challenge—Establishing the Baseline	9
3 Methods for Net Savings Estimation	12
3.1 Randomized Controlled Trials and Quasi-Experimental Designs	13
3.1.1 Randomized Control Trials	13
3.1.2 Quasi-Experimental Designs	19
3.2 Survey-Based Approaches	26
3.2.1 Program Participant Surveys	26
3.2.2 Surveys of Program Nonparticipants	31
3.2.3 Market Actor Surveys	32
3.2.4 Case Studies for Estimating Net Savings Using Survey Approaches	33
3.3 Common Practice Baseline Approaches	39
3.4 Market Sales Data Analyses (Cross-Sectional Studies)	46
3.5 Top-Down Evaluations (Macroconsumption Models)	48
3.5.1 Developing Top-Down Models	53
3.6 Structured Expert Judgment Approaches	55
3.7 Deemed or Stipulated Net-to-Gross Ratios	57
3.8 Historical Tracing (or Case Study) Method	59
4 Conclusions and Recommendations	62
4.1 A Layered Evaluation Approach	62
4.2 Selecting the Primary Estimation Method	63
4.3 Methods Applicable for Different Conditions	66
4.4 Planning Net Savings Evaluations—Issues to be Considered	70
4.5 Trends and Recommendations in Estimating Net Savings	71
References	74
Appendix: Price Elasticity Studies as a Component of Upstream Lighting Net Savings Studies ...	83

List of Tables

<u>Table 1. Applicability of Approaches for Estimating Net Savings Factors</u>	<u>12</u>
<u>Table 2: Randomized Control Trials (RCTs) —Summary View of Pros and Cons</u>	<u>18</u>
<u>Table 3. Quasi-Experimental Designs—Summary View of Pros and Cons</u>	<u>25</u>
<u>Table 4: Information Sources for the Three Levels of NTG Ratio Analysis</u>	<u>36</u>
<u>Table 5. Assignment of Free Ridership Score Based on Participant Responses</u>	<u>36</u>
<u>Table 6. Survey-Based Approaches—Summary View of Pros and Cons</u>	<u>39</u>
<u>Table 7. Common Practice Baseline Approach—Summary View of Pros and Cons</u>	<u>46</u>
<u>Table 8. Market Sales Data Analyses—Summary View of Pros and Cons.....</u>	<u>48</u>
<u>Table 9. Top-Down Evaluations (Macroeconomic Models)—Summary View of Pros and Cons</u>	<u>55</u>
<u>Table 10. Structured Expert Judgment Approaches—Summary View of Pros and Cons.....</u>	<u>57</u>
<u>Table 11. Deemed or Stipulated Approaches—Summary View of Pros and Cons</u>	<u>59</u>
<u>Table 12. Historical Tracing (or Case Study) Method—Summary View of Pros and Cons</u>	<u>61</u>

DRAFT: NOT FOR CITATION

Estimating Net Energy Savings

This chapter focuses on ~~the rationale for net savings estimation~~ methods used to estimate net savings in evaluation, measurement, and verification (EM&V) studies for energy efficiency (EE) ~~program evaluations, and where appropriate, how a conceptual view of net savings can influence the choice of methods. In this context, the purpose of programs. The most direct contribution from net savings~~ evaluation ~~studies~~ is to provide decision-makers¹ ~~with~~ the information needed to make good EE investment decisions ~~in EE~~. The specific audience for the evaluation effort can influence the methods used, the aspects of the evaluation that are emphasized, ~~and~~ the depth of ~~analysis, and the presentation of manner in which~~ the ~~work results~~ are presented.

Estimating net savings is central to many EE evaluation efforts and is broad in scope ~~since it focuses on defining. It requires the determination of~~ baselines (i.e., the counterfactual scenario) and savings levels ~~across many types of programs~~. The intent of this chapter is not to prescribe specific methods for estimating net savings, but ~~rather~~ to describe commonly used methods ~~and the~~, as well as other methods that are receiving attention in the evaluation community. ~~Information is provided on tradeoffs of each to enable each jurisdiction in the application of methods that will help jurisdictions and program administrators~~ make good decisions about ~~what the use of~~ net savings methods ~~to use~~.

The References section at the end of this chapter includes cited articles that ~~cover~~ address the ~~specific~~ ~~presented~~ methods in greater depth ~~than the scope of this chapter allows~~.

¹ Decision-makers that influence EE investments include regulators, utilities, program administrators, legislators, and implementation contractors ~~who conduct much program delivery field work~~.

1 Universality of the Net Impacts Challenge

Investment decisions result in allocating resources to achieve particular objectives. Regardless of the type of investment, once made, it is difficult to assess what would have happened absent that decision. This is the essence of evaluation: “What are the impacts of that investment decision?” These are termed *net impacts*, or *attributable impacts*. To address net impacts, a baseline is needed that represents what would have happened in the absence of this investment. ~~What would have happened in the absence of the investment.~~ This baseline is termed also called the “counterfactual scenario.”²

~~Journal publications and books that examine~~ The broader literature on evaluation practices ~~reveal~~ reveals a parallel between issues arising from estimating the net impacts of EE investments and ~~other~~ estimating the effects of other types of investments made in either the private or the public sector. Examples include:

- Healthcare: What would the health effects have been without an investment in water fluoridation?
- Tax subsidies for economic development: Would the project—or a variant of the project—have proceeded without a subsidy?
- Education subsidies: What would happen if school lunch programs were not subsidized or if low-interest loans for higher education were not offered?
- Military expenditures: What would have happened without an investment in a specific military program or technology?

Across industries and applications, program evaluators grapple with how to appropriately approximate the counterfactual scenario. ~~For EE programs, the counterfactual scenario often includes an assumption and determine impacts that some program participants would have installed some of the program promoted EE measures, even if the program had not existed.~~ are attributable to the investment being analyzed (Cook et al. 2010).²

² Some evaluators also view net savings estimation as an assessment of causality. This chapter uses the term attribution rather than causality, as it is more descriptive of the problem discussed, whereas causality has a wider range of interpretations that extends to metaphysics.

2 Defining Gross and Net Savings for Practical Evaluation

This section ~~discusses estimating net savings as an assessment of attribution.~~³ It defines key terms related to estimating net savings and summarizes ~~the different~~ various uses of net savings measurement in the industry. It also describes many ~~of the~~ issues evaluators face when estimating net savings, ~~which is tied to~~ in the context of developing an appropriate baseline against which program accomplishments are compared to estimate net impacts.

2.1 Definition of Gross and Net Savings

The Uniform Methods Project (Haeri, 2013) provides the following definitions of gross and net savings:

- **Gross Savings**~~savings~~: Changes in energy consumption that result directly from program-related actions taken by participants of an EE program, regardless of why they participated.
- **Net Savings**~~savings~~: Changes in energy use that are attributable to a particular EE program. These changes may implicitly or explicitly include the effects of ~~freeridership~~ free ridership, spillover, and induced market effects.

~~The term net to gross (NTG) ratio is almost synonymous with estimating net savings. The NTG ratio is commonly defined as the ratio of net to gross savings, and is multiplied by the gross savings to estimate net savings.~~

2.2 Definitions of Factors Used in Net Savings Calculations

The factors most often used to calculate net savings are ~~freeridership~~ free ridership, spillover (both participant and nonparticipant), and market effects. The definitions of these factors shown in Section 2.2.1 and Section 2.2.2 are consistent with those contained in the Energy Efficiency Program Impact Evaluation Guide (SEE Action, 2012b).

Freeridership

2.2.1 Freeridership~~Free Ridership~~

Free ridership is the ~~percentage of~~ program savings attributable to ~~freeriders~~. Freeriders are free riders (program participants who would have implemented a program measure or practice in the absence of the program). There are three types of ~~freeridership for program evaluators to address~~: free riders:

³ ~~Some evaluators also view net savings estimation as an assessment of causality. This chapter uses the term attribution rather than causality, as it is more descriptive of the problem discussed whereas causality has a wider range of literature interpretations that even extends to metaphysics. See discussion in (Cook, T. et al., 2010) on Causality in Contemporary Evaluation.~~

- **Total ~~Free~~ridersfree riders:** Participants who would have completely replicated the program measure(s) or practice(s) on their own and at the same time in the absence of the program.
- **Partial ~~Free~~ridersfree riders:** Participants who would have partially replicated the program measure(s) or practice(s) by implementing a ~~lessor~~lesser quantity or lower efficiency level.
- **Deferred ~~Free~~ridersfree riders:** Participants who would have completely or partially replicated the program measure(s) or practice(s) at a ~~future~~-time beyond~~after~~ the program timeframe.

~~2.2.12.2.2~~ Spillover

Spillover refers to additional reductions in energy consumption ~~and~~/or demand that are due to program influences beyond those directly associated with program participation. ~~Spillover accounts for~~As a result, these savings may not be recorded in the actions participants take without program financial or technical assistance.~~tracking system and credited to the program.~~ There are generally two types of spillover:

- **Participant ~~Spillover~~spillover:** This represents the additional energy savings that ~~occur~~are achieved when a program participant—as a result of the program’s influence—installs EE measures or practices *outside* ~~of~~ the efficiency program after having participated.

Evaluators have further defined the broad category of participant spillover into the following subcategories:

- *Inside ~~Spillover~~spillover:* Occurs when participants take additional program-induced actions at the project site.
- *Outside ~~Spillover~~spillover:* Occurs when program participants initiate actions that reduce energy use at sites that are not participating in the program.
- *Like ~~Spillover~~spillover:* Refers to program-induced actions participants make outside the program that are of the same type as those made through the program (at the project site or other sites).
- *Unlike ~~Spillover~~spillover:* Refers to EE actions participants make outside the program that are unlike program actions (at the project site or other sites) but that are influenced in some way by the program.
- **Nonparticipant ~~Spillover~~spillover:** This represents the additional energy savings that ~~occur~~are achieved when a nonparticipant implements EE measures or practices as a result of the program’s influence (for example, through exposure to the program) but is not accounted for in program savings.

~~2.2.22.2.3~~ Market Effects

Market effects refer to “a change in the structure of a market or the behavior of participants in a market that is reflective of an increase in the adoption of energy efficiency products, services, or practices and is causally related to market intervention(s)” (Eto et al., 1996). For example, programs can influence design professionals, vendors, and the market (through product

availability, practices, and prices), as well as ~~influencing~~ influence product or practice acceptance and customer expectations. All of these influences may induce consumers to adopt EE measures or actions. ~~As a result, an evaluator might conclude that some participants are current freeriders when, in fact, their actions represent market effects from prior year. Participants may not have previously adopted an EE measure, practice, or service because it did not exist in the marketplace or was not available at the same price without the utility programs. (Sebold et al. 2001).~~⁴ These freeriders can represent savings that resulted from programs in prior years due to market effects. It is important to recognize that evaluators may not have previously accounted for these ongoing effects. Program administrators and evaluators should consider nonparticipant spillover when developing the policy context for evaluating current year program results.

~~There is debate regarding the difference between spillover and market effects. Some experts suggest that market effects can be “best viewed as spillover savings that reflect significant program-induced changes in the structure or functioning of energy efficiency markets.”~~⁵ ~~While Prah et al. (2013) also suggest that market transformation is a subset of market effects (as the substantive and long-lasting effects). This view implies that market effects are a subset of spillover. Although~~ spillover and market effects are related, the methods used to quantify these two factors generally differ. ~~For that reason~~ Therefore, this chapter addresses them separately.⁶

2.2.4 Net Savings Equations

Evaluators use different factors to estimate net savings for various programs and jurisdictions, depending on how a jurisdiction views equity and responsibility (NMR ~~et al.~~, Group, Inc. and Research Into Action 2010). For example, some jurisdictions ~~only~~ include ~~freeridership~~ only free ridership in the calculation of net savings ~~while~~; others include both ~~freeridership~~ free ridership and spillover. Some jurisdictions estimate net savings without measuring ~~freeridership~~ free

⁴ ~~Sebold et al. (2001) sets out an expansive framework for assessing EE programs. When assessing EE policies in a broad context, it should be acknowledged that some participants identified as free riders in a current program might not have had the opportunity to adopt the EE measure or service were it not for the effects on the market from previous EE program efforts. These efforts may have contributed to that measure or service being available to customers in the current year. The importance of this issue to evaluation depends on the parameters of the evaluation. Most evaluations focus on set time periods spanning 1–3 years. Factors that are included are based on the incremental actions taken as a result of the EE program year being evaluated and the current state of the EE market. Actions taken that resulted from EE efforts in preceding years represent sunk costs and are not incremental to the current program being evaluated. However, this may be an important consideration in a broader policy assessment examining the overall trend in the adoption of EE measures and services across a longer time period. Market effects of previous years’ programs may not have been fully accounted for, and this can be a consideration in the broader policy context. However, for assessing the impacts of a given EE program for a given year, these effects from past programs are not generally considered. This is discussed in more detail in Section 3.3.~~

⁵ ~~Prah et al. (2013) also suggest that market transformation is a subset of market effects (as the substantive and long-lasting effects). This view implies that market effects are a subset of spillover.~~

⁶ ~~Some stakeholders view spillover as a subset of market effects, with market effects including long-lasting participant and nonparticipant spillover in addition to those that are program-induced. While the terminology varies across some applications, the estimation issues are the same.~~

ridership or spillover (market-level estimates of net savings).⁷ Messenger et al. (2010) also discuss differences across jurisdictions in the reporting of gross and net savings.

A practitioner who is trying to develop methods ~~for estimating~~ to estimate values for these factors will find the definitions provided in this section useful. However, the evaluator must work with the information available, which starts with the tracking system.⁸ Evaluators typically view the data in the tracking system as the initial estimate of gross savings. ~~Since freeridership~~ Because free ridership, spillover, and market effects are untracked values, evaluators ~~must~~ should estimate or account for them outside ~~of the~~ the program tracking system.⁹ A practical way to ~~account for~~ understand these values is to consider spillover and market effects as savings that are attributable to the program, but that are not included in the program tracking system. ~~Freeridership~~ Free ridership represents savings included in the program tracking system that are not attributable to the program.

To estimate net savings, the evaluator first estimates ~~these values~~ free ridership, spillover, and market effects, then makes appropriate adjustments to the values in the tracking database (or validated tracking database).¹⁰ as illustrated in equation 1.¹¹

⁷ Differences across jurisdictions in the reporting of gross and net savings is also discussed in Messenger et al. (2010, p. 19-21).

⁸ The definitions for ~~freeridership~~ free ridership, spillover, and market effects ~~must~~ should be integrated with ~~both~~ (1) how the utility tracks actual program participation data; and (2) how the utility records information about expected program impacts in the program tracking system. In general, the initial gross savings estimate (in terms of expected energy savings by participant or measure) comes from the tracking system. ~~Some of these~~ These data may include “deemed values” negotiated by the stakeholders. ~~In some cases, these~~ These deemed values may include factors that lower the savings of a measure, based on assessments of current practice, codes and standards, and ~~or~~ other factors that may directly or indirectly influence how the estimated gross savings are adjusted to estimate net savings. ~~As a result, it~~ It is important to understand how the gross savings are estimated by project and by participant. In fact, the first recommendation of NMR ~~Group Inc. and~~ Research into Action (2010) is that the Northeast Region needs a process leading to the development of a consistent definition of *adjusted gross savings*.

⁹ ~~There are direct~~ Direct estimation methods that can be used are available to address ~~freeridership~~ free ridership, spillover, and market effects without estimating each separately. This chapter addresses randomized control trials, quasi-experimental designs, and common practice baselines, each of which essentially is used to adjust the savings estimates in the program tracking system.

¹⁰ ~~A validated tracking database is simply a reviewed program tracking database. Programs that are equipment based use either a rebate or custom design and have a program tracking database that estimates the savings expected to be achieved by installing that particular equipment. A review of this tracking database can determine any obvious errors, whether adjustments can make the claimed (ex ante) savings entries more accurate, and whether any deemed savings values already include adjustments that account for net savings factors (for example, an adjusted baseline that captures market trends). The validated tracking system then contains the most accurate information on claimed savings for each participating site or project. The benefits of improved information in the tracking system are discussed in Violette et al. (1993).~~

¹¹ ~~Keating (2009) discusses issues concerning the math underlying net savings and NTG calculations, giving examples of an inappropriate multiplication algorithm using freeridership and spillover expressed as ratios.~~

A validated tracking database is simply a reviewed program tracking database. A review of the tracking database can determine obvious errors, whether adjustments can make the claimed (ex ante) savings entries more accurate, and whether any deemed savings values include adjustments that account for net savings factors (for example, an adjusted baseline that captures market trends). The validated tracking system then contains the most

Equation 1. Net Savings Including ~~Freeridership~~Free Ridership, Spillover, and Market Effects

$$\text{Net Savings} = \text{Gross Savings} - \text{FR} + \text{SO} + \text{ME not already captured by SO}$$

Where:

FR = ~~freeridership~~free ridership savings

SO = spillover savings

ME = market effects savings not already captured by SO

In much of the literature, the program evaluation approach involves a net-to-gross (NTG) ratio for which ~~freeridership~~free ridership, spillover, and market effects are expressed as a ratio to gross savings: (equation 2). These widely used NTG-ratios work well for some types of evaluation efforts (for example, survey-based estimations). The term is almost synonymous with estimating net savings and is commonly defined as the ratio of NTG savings for the sample. The population gross savings is then multiplied by the NTG ratio to estimate population net savings.

Equation 2. Net-to-Gross Ratio

$$\text{NTG Ratio} = 1 - \text{FR ratio} + \text{SO ratio} + \text{ME ratio (where the denominator in each ratio is the gross savings)}$$

When using the NTG ratio defined by specific ~~freeridership~~free ridership, spillover, and market effect factors (or ratios), evaluators use ~~the follow~~ equation 3 to calculate net savings:

Equation 3. Net Savings Calculation Using the Net-to-Gross Ratio

$$\text{Net Savings} = \text{NTG Ratio} * \text{Gross Savings}$$

~~While the above~~These definitions are essentially standard in the evaluation literature;¹² however, a given jurisdiction may decide not to include ~~freeridership~~free ridership, spillover, ~~and/or~~ or market effects ~~in the estimation of to estimate~~ net savings. For example, ~~while~~ evaluators almost always include ~~freeridership~~free ridership, but, because of policy choices made in a jurisdiction, most do not always fully consider spillover and market effects (see NMR ~~et al.,~~Group, Inc. and Research Into Action 2010 and; NEEP, 2012). ~~This is due to the policy choices made by that jurisdiction. 2012)~~. Most evaluators agree that spillover and market effects exist and have positive values, but ~~it can be difficult to determinedetermining~~ the magnitudes of these factors can be difficult. Increasingly, the trend is to include estimates of spillover in net savings evaluations. The inclusion of market effects is also increasing, but ~~not to the same~~ a lesser degree

accurate information on claimed savings for each participating site or project. The benefits of improved information in the tracking system are discussed by Violette et al. (1993).

¹² Other factors (sometimes called *net-impact factors*) are generally considered as adjustments to gross impact estimates. These include rebound, snapback, and persistence of savings. UMP Chapter 13—Persistence and Other Evaluation IssuesViolette (2013) addresses these factors ~~(Violette, 2013)~~. As with other NTG factors, evaluations do not treat net-impact factors consistently in gross impact calculations, and do not consistently adjust program gross impacts to calculate to a final net impacts number.

~~asthan~~ spillover. Methods are available to address ~~both~~ spillover and market effects and, ~~sincebecause~~ there is really no debate about whether they exist, these factors should be addressed when estimating net savings. ~~The spillover and market effects estimates may have some uncertainty, but no more than that in evaluation literature from other fields.~~ It is important to know the potential sizes of spillover and market effects for a given program or portfolio so ~~that~~ appropriate policy decisions can be made ~~regardingabout~~ EE investments.

2.3 Uses of Net Savings Estimates in the ~~EE~~ Energy Efficiency Industry

~~There is much discussion within many~~ Many regulatory jurisdictions ~~regarding~~ discuss the appropriate use of net savings estimates. This is due in part to: (1) the cost of the studies to produce these estimates;¹³ and (2) a perceived lack of confidence in the resulting estimates.¹⁴ However, evaluators and regulators recognize the advantages of consistently measuring net savings over time as a key metric for program performance (Fagan et al., 2009).

Evaluators generally agree ~~upon the following five uses for that~~ net savings ~~research can be useful for~~ (SEE Action 2012a, 2012b):¹⁵

- ~~Program planning~~ Gaining a better understanding of how the market responds to the program and using that information to modify the program design (for example, to set consumer including eligibility and target marketing and incentive levels).
 - ~~Assessing the degree to which programs cause a reduction in energy use and demand (net savings is one of numerous program success measures that should be assessed).~~
- ~~Obtaining~~ Gleaning insight into market transformation over time by tracking net savings across program years and determining the extent to which ~~freeridership~~ free ridership and spillover rates have changed over time. This insight ~~can potentially might~~ be used to define and implement a program exit strategy.
 - ~~Gaining a better understanding of how the market responds to the program and how to use that information to modify the program design (including how to define eligibility and target marketing).~~

¹³ GDS Associates (2012) provides additional information about the costs and benefits of evaluation, measurement, and verification approaches for small utilities (see <https://www.nreca.coop/wp-content/uploads/2013/12/EMVReportAugust2012.pdf>).

¹⁴ For additional information on the costs and benefits of different EM&V approaches for small utilities, see: <https://www.nreca.coop/wp-content/uploads/2013/12/EMVReportAugust2012.pdf>. Several experienced evaluators indicated in comments on earlier drafts of this chapter that in their experience, the required level of confidence and precision for estimates of net impacts within the EE field is generally greater than that used in other fields faced with similar types of questions and tradeoffs. The authors generally agree with this observation, but no meta-study comparing target levels of confidence and precision for EE program evaluation with similar evaluations in other fields has been conducted.

¹⁵ Other methods that can and should be used to inform program design and understand market response include process evaluations and market assessments.

- Informing resource supply and procurement plans, which requires an understanding of the relationship between efficiency levels embedded in base-case load forecasts and the additional net reductions from programs.
- Assessing the degree to which programs effect a reduction in energy use and demand (net savings is one program success measure that should be assessed).

With respect to the last bullet, Schiller (SEE Action, 2012b, pp. 2–5) also discusses the importance of consistently measuring savings across evaluation efforts and having consistent evaluation objectives. For example, evaluators in different jurisdictions assess the achievement of goals and targets as measures of overall EE program performance using different measures of savings: gross savings, net savings, or a combination of the two. There are also differences across jurisdictions in which the measure of EE program success is used for calculating financial incentives. There are arguments for basing financial incentives on net savings, as well as arguments for basing incentives on gross savings or a combination of the two.¹⁶

2.4 The Net Savings Estimation Challenge—Establishing the Baseline

This chapter discusses estimation methods that rely on the development of a baseline (the assumed counterfactual scenario). ~~The~~This baseline is used to measure the net impacts of a program. ~~To understand and defend the selection of a particular method for estimating net savings, evaluators must consider the implicit and explicit assumptions used for the baseline comparison group.~~ If evaluators could identify a “perfect baseline”¹⁶; i.e., a counterfactual scenario that exactly represents what would have happened if the EE program had not been offered, most of the issues associated with estimating net impacts would not ~~occur~~arise.

The evaluator is faced with the challenge of identifying a method that produces a baseline that best ~~representing~~represents the counterfactual scenario—in other words, what the participant group (and the market) would have done in the absence of the program.¹⁷ ~~The evaluator must~~To understand and defend the selection of a particular method for estimating net savings, the evaluator should consider the implicit and explicit assumptions used for the baseline comparison

¹⁶ —As more jurisdictions begin to consider the delivery of EE programs as a business process that requires an investment of resources, they are considering the return on investment (~~ROI~~), ~~ROI~~ (more commonly termed *incentives*), which is typically coupled with performance targets. Jurisdictions can base targets on reaching a certain level of gross savings or on achieving a certain level of net savings—each has pros and cons. A gross savings target ~~provides~~may provide a ~~more clear~~clearer incentive structure for the program administrator, and there is generally less controversy over whether the target is achieved. The fact that incentives are usually based on a calculation of shared benefits, where the predominant share of benefits goes to ratepayers, creates an equitable incentive structure: the program administrator receives fewer benefits and even if attributed (net) savings are ~~less~~lower than expected, the ratepayers still receive ~~the majority~~most of the benefits. For example, under an 80–20% split of the benefits (80% of benefits are realized by ratepayers and 20% ~~are realized~~ by the administrator), having attributed savings reduced by 50% still implies that 70% of the benefits go to ratepayers. See Rufo (2009) for other views on aligning incentives with the outputs of program evaluation.

¹⁷ —Chapter 8 of the DOE Uniform Methods Project (Agnew et al., 2013) provides DAN CAN YOU CLARIFY—Should be Agnew and Goldberg (2013) in all instances. Agnew and Goldberg (2013) provide a number of choices for selecting control groups for use in billing analyses (for example, comparing changes in energy use for ~~both~~ participants and a control group). It also discusses using regression analysis as a tool for making appropriate comparisons and arriving at alternative net savings values.

~~group. For example, when considering the use of nonparticipants as a candidate baseline, the evaluator needs to account for issues that pertain to the similarity, or matching, of the participant and program participants with customers that may comprise the nonparticipant/ comparison groups.~~ The evaluator ~~must~~ should also account for any effects the program might have had on the comparison group (that is, any interactions between the participant group and the comparison group that may impact the program net savings). ~~In addition to the baseline estimation methodology issues described in more detail in the next section, self-selection bias, freeridership, and spillover can cause concern when estimating a baseline.~~¹⁸ influence the program net savings).

Self-selection ~~bias~~ can be viewed as a baseline issue arises when a program is voluntary and participants select themselves into the program, suggesting the potential for systematic differences between program participants and nonparticipants. This issue is not unique to EE evaluations and ~~is present~~ arises in any policy or program assessment involving self-selection. ~~Freeridership is one specific variant of self-selection bias. This is~~ In this context, free ridership is viewed as a baseline issue when the actions of the nonparticipant comparison/control group do not accurately reflect the actions participants would have taken in the absence of the program. Specifically, the assumption in this case is that the self-selected participants are those who would have taken more conservation actions than the general nonparticipant comparison group.¹⁹

~~While freeridership~~ Free ridership reduces net program savings, ~~there are in this example case, but~~ other variants of self-selection ~~bias that~~ might increase net savings. ~~when a participant group is compared to a nonparticipant baseline.~~ For example, if the customers who self-select ~~themselves~~ into the program need the financial incentives to justify the EE investment, an adjustment for self-selection might increase overall net savings. ~~The fact that participants are self-selected does not indicate whether net savings are over or under estimated.~~

¹⁸ ~~Self selection, freeridership, and spillover issues are common in other applications as well. Consider a business decision to downsize in order to produce net benefits. Self selection would be addressed when designing the business initiative. Freeriders would be considered, such as whether employees who are most confident and have the best chance to get new jobs would take a potential early retirement package. Spillover impacts are considered, such as whether productivity is impacted for employees that remain on the job after their coworkers are downsized. While self selection, freeridership, and spillover pose challenges for EE evaluation, they are part of assessing many investment and business decisions.~~

¹⁹ ~~In this context, freeriders are a subset of the self-selection bias case, the nonparticipant baseline does not fully correct for free riders, resulting in estimated net savings that are biased upward. Other self-selection bias factors could result in cause the participant and nonparticipant groups behaving to behave differently. For example, if participants really need the rebate financial assistance to make the investment and nonparticipants do not need the rebate to take EE actions, then the baseline comparison group would might take more EE actions than the participant group. The result is a low estimate of savings, rather than an estimate that is too high such as occurs under in the commonly assumed freerider self-selection hypothesis. Developing a better comparison group in absence of the program. In this case, a nonparticipant baseline would correct produce estimated net savings that are biased downward and appropriately correcting for the this self-selection bias and effect would increase the estimated program savings-net savings. The authors have observed that often there is an assumption that addressing self-selection will always lower estimated net savings by reducing bias caused by free riders, but this is not always the case.~~

Spillover ~~is another~~ can also be viewed as a baseline issue. For example, nonparticipant spillover can occur when the energy consumption of the comparison group of nonparticipants is not indicative of what the energy consumption for this group would have been in the absence of the program. In this case, the comparison group is *contaminated*: ~~the existence of~~ the program affected the behavior of those in the comparison group.

This section ~~discusses~~ discussed issues related to establishing an appropriate baseline as an approximation of the counterfactual scenario. Understanding that ~~free ridership~~ free ridership, spillover, and market effects ~~are~~ can be viewed as baseline issues can help the evaluator focus on ~~those~~ the factors that are most important to selecting an appropriate method.²⁰ In many applications, selecting the baseline is a core issue in choosing an appropriate estimation method. When ~~presentating~~ presenting the net savings results of a program, the evaluator should include a description of the baseline and the assumptions implicit in the estimation method.

²⁰ Self-selection, free ridership, and spillover issues are not unique to EE evaluation—they are common in other settings as well. Consider a business decision made to produce net benefits, such as downsizing. Might self-selection be important to address in assessing this business initiative? Employees who have the best experience and are the most confident in their ability to land new jobs might (if able) self-select into the downsizing option. Might there be some free riders if the downsizing effort includes personnel who were planning to leave anyway? Also, there might be spillover impacts from the downsizing program where having workers leave reduces the productivity of employees who remain. Although self-selection, free ridership, and spillover pose challenges for EE evaluation, these same issues often have to be addressed in evaluating investment decisions in other fields and contexts.

3 Methods for Net Savings Estimation

This section discusses ~~different~~ methods for estimating net savings, as well as some of the ~~key~~ advantages and challenges associated with each ~~method~~. Evaluators use a variety of methods, some of which address ~~freeridership free ridership and/or spillover~~ (for example, self-report surveys) ~~while other methods are focused~~; ~~others focus~~ on market effects (for example, structured judgment approaches or historical tracing). The methods addressed in this section are:²¹

- Randomized control trials (RCTs) and quasi-experimental designs
- Survey-based approaches
- Common practice baseline approaches
- Market sales data analyses
- Top-down evaluations (or macroeconomic models)
- Structured expert judgment approaches
- Deemed or stipulated NTG ratios
- Historical tracing (or case study) method.

Table 1 lists ~~which~~ methods ~~that~~ are applicable for estimating ~~freeridership free ridership~~, spillover, and market effects. This table ~~only~~ indicates the general applicability of the ~~approaches methods~~. The following sections review the specific applications, caveats, limitations, and other key information in greater detail ~~needed to understand explain~~ how to assess the methods for each net savings component.

Table 1: Applicability of ~~approaches~~ ~~Approaches for estimating net savings factors~~ Estimating Net Savings Factors

Method	Freeridership <u>Free Ridership</u>	Spillover	Market Effects
Randomized controlled trials RCTs and quasi-experimental designs	Controls for freeriders free riders ²²	Controls for participant spillover ²³	Not generally used
Survey-based approaches	Is applicable	Is applicable	In conjunction with structured expert judgment
Common practice baseline	Is applicable	Is Not applicable ²⁴	Not applicable

²¹ ~~Price elasticity studies examine how consumers respond to reductions in price for an EE product. To date, these studies have examined programs that lower the costs of lighting products (for example, CFLs), but have not expanded to other EE products. Please see Appendix A for a discussion of this methodology.~~

²² ~~Does not provide a direct estimate freeridership free ridership, but rather controls for freeriders free riders through experimental design.~~

²³ ~~Does not estimate spillover, but rather controls for participant spillover through experimental design. A separate study of control group members is required to address nonparticipant spillover if it is expected to be significant and affect the net impacts.~~

Method	Free ridership Free Ridership	Spillover	Market Effects
methods			
Market sales data analysis	Is applicable	Is applicable	Is applicable
Top-down evaluations	Assesses the overall change in energy use, and therefore there is no adjustment needed for free ridership, spillover, and market effects		
Structured expert judgment ²⁵	Is applicable	Is applicable	Is applicable
Deemed or stipulated NTG ratios	Is applicable	Is applicable	Not generally used
Historical tracing	Is applicable	Is applicable	Is applicable

3.1 Randomized Controlled Trials and Quasi-Experimental Designs

This section discusses two methods for selecting a baseline against which to compare program impacts: RCTs and quasi-experimental designs. RCTs represent the ideal approach, but may not always be possible. When an RCT is not possible, a quasi-experimental design is an alternative. These approaches are increasingly being used to evaluate behavioral programs, information programs, and pricing programs designed to increase efficiency.²⁶ Generally, these programs have a large number of participants that are typically in the residential sector.

3.1.1 Randomized Control Trials

An RCT design is ideal for assessing the net impacts of a program—particularly the free ridership and short-term spillover components. If the RCT is short term (that is, one year or less), then it may not capture longer term spillover and market effects.

For the RCT, the study population is defined first, then consumers from the study population are randomly assigned to either a treatment group (participants in the EE program) or to a control group that does not receive the treatment (nonparticipants). Random assignment is a key feature of this method. By using random probability to assign consumers to either the treatment or the control group, the influence of observable differences between the two groups is eliminated (for example, location of home, age of home, and appliance stock). Unobservable differences are also eliminated (for example, attitudes toward energy use, expectations about future energy prices, and expertise of household members in areas that might induce participation).²⁷ As a result, this

²⁴ Spillover could arguably be addressed through surveys of participants and nonparticipants, but this is not generally viewed as being part of the common practice baseline method, and the use of surveys would make this more similar to survey-based estimation methods discussed in Section 3.2.

²⁵ — This approach is only applicable only if the experts are knowledgeable about the specific market being studied.

²⁶ The SEE Action (2012a) report, focused on information and behavioral programs, was authored/written for the Customer Information and Behavior Working Group and the Evaluation, Measurement, and Verification Working Group. More information is available at www.seeaction.energy.gov.

²⁷ References addressing the RCT and quasi-experimental designs include: NMR Group, Inc. and Research Into Action (2010) and two reports by SEE Action (2012a, 2012b). The

SEE Action
reports can be downloaded at:

<http://www1.eere.energy.gov/seeaction/index.html>.

[\(NMR Group, Inc. and Research Into Action \[2010\]; SEE Action \[2012a, 2012b\]\)](#). This method, when implemented properly, can provide a near-perfect baseline that results in reliable net savings estimates.

The net savings calculations are relatively straightforward when [an](#) RCT is designed properly. The literature generally covers three methods for calculating net savings:

1. **Use a simple post-period comparison to determine the differences in energy use between the control and treatment groups after participation in the program.** For example, if participating households are using 15,000 kilowatt hours (kWh) on average and the control households are using 17,000 kWh, ~~then~~ the net savings estimate is 2,000 kWh.
2. **Use a difference-in-differences (DiD) approach, ~~comparing to compare~~ the change in energy use for the two groups between the pre- and post-participation periods.** For example, assume participants used 17,500 kWh prior to program participation and 15,000 after participation, for a difference of 2,500 kWh between the pre- and post-periods. Assume also that the well-matched control group has similar pre-period energy use (approximately 17,500 kWh), but the group's post-period energy use is 17,000 kWh (that is, slightly ~~less~~lower, possibly ~~due to~~because of weather), for a difference of 500 kWh. Applying the DiD method results in an estimated savings of 2,000 kWh (the 2,500 kWh change for participants minus the 500 kWh change for nonparticipants).
3. **Use ~~of~~ a linear fixed-effects regression (LFER) approach,** where the regression model identifies the ~~effect~~effects of the program by comparing pre- and post-program billing data for the treatment group to the billing data for the control group. A key feature of the LFER approach is the addition of a customer-specific intercept term that captures customer-specific effects on electricity ~~usage~~use that do not change over time, including those that are unobservable. Examples ~~of these fixed effects~~ include the square footage of a residence, the number of occupants, and thermostat settings.²⁸ [\(see Provencher and Glinsmann \[2013\] for an example and additional discussion of the LFER method\).](#)²⁹

Even if randomizing the ~~participant~~treatment and control groups, an evaluator may use a method other than the simple post-period comparison ~~in an effort~~ to be as thorough as possible and use all the available data to develop the estimate. The DiD method tracks trends over time, and the fixed-effects component of the LFER adds an extra control for the differences between

²⁸ ~~See Provencher and Glinsmann (2013) for an example and additional discussion of the LFER method.~~

²⁹ ~~A number of the methods discussed in this chapter use regression approaches; some of these. Some are~~ [more](#)fairly simplistic ~~while~~; others are quite sophisticated, requiring expertise in econometrics. Each section provides citations to applied studies, many of which describe the econometric techniques employed. For example, Stuart (2010) lists econometric software and routines that can be useful in matching. Also, [Chapter 8 of the DOE UMP](#) (Agnew et al., (2013) ~~discusses~~discuss regression models in more detail, but provides a limited set of literature references. SEE Action (2012a) recommends [Econometric Analysis](#) by Greene (2011) as a useful reference on regression techniques. Wooldridge (20022010) focuses on cross-section and panel data models that are often used in evaluation. [The Guide to Econometrics](#) by P. Kennedy (2008) and [Mostly Harmless Econometrics: An Empiricist's Companion](#) by Angrist and Pischke (2008) are useful supplements to any econometrics ~~text book~~textbook.

consumers that are constant during the period being examined. ~~In theory, all~~ All three methods should produce the same result generate unbiased estimates, as all three are based on the assumption that randomization fully accounts for the ensures no systematic differences between the treatment and control groups that influence in the estimated drivers of program net savings energy use, so the three methods would be expected to generate similar, but not necessarily identical, results.

The RCT approach is simple in concept, but may be more difficult to implement given available data, timing, and/or program design issues. ~~For example, suppose an evaluator selects the study population and performs the random assignment for a Home Energy Reports (HERs) program in which program administrators send energy use reports by mail. This program is designed to generate energy savings by providing residential consumers with information about their energy use and energy conservation. The reports give consumers various types of information, including: (1) how their recent energy use compares to their energy use in the past; (2) tips on how to reduce energy consumption, some of which are tailored to each consumer's circumstances; and (3) information on how their energy use compares to that of neighbors with similar homes. Even though this is a random design, the evaluator still needs to check for any inadvertent systematic differences between the participant and control groups. The evaluator may discover that several thousand program households have addresses where mail is not delivered. As a result, these households were inadvertently dropped from the participant group. If those consumers are dropped from the participant group, should the consumers with undeliverable mail addresses also be dropped from the control group? This situation occurred in a randomized trial of a behavioral program (see Provencher and Glinsmann, 2013), and the evaluators adjusted the samples and tested for conformance with RCT assumptions.~~

It is becoming standard practice for evaluators to ~~test the likelihood that the program groups and control groups are appropriately randomized. They can apply~~ use statistical methods to test whether the ~~data in the two groups are consistent with random assignment of consumers to the allocation of customers between the treatment group and the control group. This is consistent with what would be expected from a random assignment of consumers to the treatment and control groups. For billing data, this~~ type of analysis often involves comparing the means of the two groups with respect to demographic variables (if available) and monthly energy use in the pre-program year. ~~To test the validity of assumptions used in an RCT, evaluators can check that the difference in mean consumption between the treatment and the control groups does not fall~~ For example, if the differences in means for the two groups falls outside a 90% confidence bound for more than ~~two~~ 2 months of the pre-program year. ~~If mean consumption does fall outside a 90% confidence bound for more than two months, there is cause for concern that assignment to the two groups is not random. (See an example of an application of this test for consistency with RCT expectations in Provencher and Glinsmann [2013] and other tests in Stuart [2010].) If this is the case, it does not prove that is worth examining how the random assignment was~~ not conducted, but it does provide a signal that to ensure no inadvertent elements of the

process ~~used to perform the random~~ assignment ~~needs to be reviewed.~~³⁰ ~~to the treatment and control groups.~~

~~The RCT approach to estimating program impacts reflects the “intent to treat” effect. Generally, it is not appropriate to drop customers after the random assignment, though the consequences of doing so vary. For example, questions may arise about what to do with consumers who opt out. Consider, for instance, a program involving Home Energy Reports (HERs), in which program administrators send energy use reports by mail. This program was designed to generate energy savings by providing residential consumers information about their energy use and energy conservation. Some percentage of consumers will opt out of the program. They should remain in the analysis because the similar set of control consumers who would have opted out of the program could not be identified if they were to receive the report. Also, on average, these consumers might have different energy use than the other control consumers, causing the reported impact to be biased if the treatment group is adjusted to remove the opt-out consumers. At the other extreme, HERs might not be deliverable because of observable address characteristics. If this same address characteristic can also be identified for control consumers, the estimate of program impacts after eliminating treatment and control consumers with this characteristic is, strictly speaking, an unbiased estimate of the effect of intent-to-treat conditional on the address characteristic. These examples are meant to show that careful analysis is needed in the application of all methods, including RCTs. In addition, Duflo et al. (2007) caution that excessive investigations of subgroups not specified *ex ante* constitute a form of data mining that should be avoided. The case discussed above where address characteristics are available for the treatment and control groups does not fall in this category, but this caution deserves emphasis.~~

To maintain ~~an~~ RCT over a period of time, evaluators must take care when working with the data across the treatment and control groups. For example, a behavioral program (such as HERs) may be rolled out to 20,000 high-use residential consumers in program year 1. In program year 2, an additional 20,000 consumers of all energy use classifications may enroll, and another 30,000 consumers may enroll in program year 3. Additionally, some consumers in program year 1 may have dropped out (requested to not receive the ~~home energy reports~~).~~HERs~~).³¹

~~Inevitably, there are also issues with~~Issues inevitably arise about the consumer energy use data. Researchers have used the following criteria, ~~among others~~, as indicators of problems with consumer billing data:

- Having ~~less~~fewer than 11 or more than 13 bills during a program year
- Having ~~less~~fewer than 11 or more than 13 bills during the pre-program year
- Energy consumption outside a reasonable range (that is, an outlier observation with average daily consumption that is ~~less~~lower than the 1st percentile or ~~greater~~higher than the 99th percentile)

³⁰ ~~See an example of an application of this test for consistency with RCT expectations in Provencher and Glimsmann (2013) and other tests in Stuart (2010).~~

³¹ ~~This is not an unusual problem in the utility industry. Utilities have for many years addressed similar issues in maintaining random customer samples for load.~~

- Observations with ~~less~~fewer than 20 or more than 40 days in the billing cycle.

Agnew and Goldberg (2013) also discuss issues with consumer energy use data in residential settings. Even programs that have operated for several years are likely to have issues. Using the HERs example, this could include consumer records that are missing the date when the first report was sent ~~for~~ or entries in ~~customer~~consumer records that indicate issues with that observation.

After addressing data issues, ~~does~~ the evaluator probably still ~~have~~has a good RCT? ~~The answer is probably,~~ unless ~~there are~~ a large number of consumers ~~who~~ are affected by these data issues or ~~there are~~ consumers are disproportionately affected across the participant and control groups. ³²Mort (2013) presents additional criteria that can cause sites to be excluded and suggestions about what to do if the number of removed sites exceeds 5%.

The ability to disseminate information to large groups of consumers has led to an increase in RCTs in EE evaluation.³³ In general, these RCT-based evaluations have focused on residential behavior-based ~~efficiency~~EE programs such as HERs programs. These programs lend themselves to random trials in that they: (1) provide information only; (2) can be implemented for a large number of consumers at the same time; and (3) allow for an RCT design. These characteristics, however, are not generally present for many large-scale EE programs that tend to ~~comprise much~~account for many of the EE portfolio savings.

In summary, the RCT approach is generally viewed as the most accurate method for estimating net impacts. The RCT controls for ~~freeriders~~free riders and near-term participant spillover, which are two important factors. To the extent that the program affects the control group, nonparticipant spillover is not addressed. This effect is likely to be small over the short run in many behavioral programs.

If nonparticipant spillover is large, net impacts will be underestimated because ~~they include freeriders that~~ there are actually nonparticipants ~~that~~who were affected by the program; ~~that is, freeridership, and the baseline~~ will be treated incorrectly in the net savings calculation inaccurate. To appropriately address this issue, the evaluator would need to conduct a separate study of control group members to address nonparticipant spillover. ~~Since~~Because market effects are longer-term spillover effects, ~~it is unlikely that they would likely not be included in~~ any RCT net savings approach that spans just a few years ~~would include market effects~~.

³² ~~Chapter 9 of the DOE UMP (Mort, 2013) presents additional criteria that can result in the exclusion of sites (see p. 26-27) and suggestions on what to do if the number of removed sites becomes large (that is, greater than 5%).~~

³³ ~~Some evaluations~~Evaluations of HERs programs that used ~~RCT~~RCTs include: Sacramento Municipal Utility District (2011), Puget Sound Energy (2012), AEP (2012), SMUD (2011), PG&E (2013), Commonwealth Edison (2012), and ComED (2012). Some of these studies actually compare the Pacific Gas & Electric (2013). Some ongoing evaluations use RCT methods for HERs programs, and will produce additional practical information on RCT design to procedures which match participants to nonparticipants applications. Another useful study, but one focused on evaluating pricing programs, which used an RCT design is SMUD the Sacramento Municipal Utility District (2013). This study assesses different pricing structures in the residential sector; however, the methods used are good examples of what can also be applied in EE evaluations in an RCT context.

Although the RCT method can produce an accurate baseline when constructed correctly, it may not be possible to apply an RCT to evaluations of EE programs for a variety of reasons. RCT generally requires planning in advance of program implementation. As pointed out in Chapter 8 (Agnew and Goldberg, 2013) of these protocols, "...evaluation concerns have been less likely to drive program planning." Also, an RCT approach may involve denying or delaying participation for a subset of the eligible and willing population. In some cases, the random assignment may result in providing services to consumers who either do not want them or may not use them- (see [Table 2 for pros and cons of RCTs](#)).

Other characteristics of programs that can make an RCT difficult to implement include:

- Programs that require significant investments, such as a commercial and industrial (C&I) major retrofit program in which the expenditures are in the tens of thousands of dollars. Typically, these programs are opt-in, and random ~~assignment~~ assignments within an eligible study population might include consumers who either do not need the equipment or services or do not want to make that investment. Programs that involve relatively large investments in measures and services across ~~both~~ the residential and C&I sectors are clearly not amenable to ~~a randomized trial~~ an RCT design.
- Participants in some C&I programs can be relatively unique, with few similar consumers ~~that~~ who might be candidates for a control group.
- To achieve savings targets, many programs must be rolled out over an entire year, with consumers opting in every month. As a result, consumers self-select into the participant group, which is unknown until after ~~one~~ 1 year of the program implementation. Evaluators can more easily apply RCT to programs with a common start date for a large number of participants (for example, HERs programs).

Table 2: Randomized Control Trials— (RCTs)—Summary View of Pros and Cons

Pros	<ul style="list-style-type: none"> • Random assignment reduces and limits bias in estimates • Increases reliability and validity • Controls for free riders free riders and participant spillover • Widely accepted in natural and social sciences as the gold standard of research designs
Cons	<ul style="list-style-type: none"> • Bias can result if random assignment occurs among volunteers or if the program drop-out rate differs by key characteristics • Does not address nonparticipant spillover • Equity/ethical concerns about assigning some ratepayers to a control group and not allowing them to participate in the program for a period of time • It is generally Generally not applicable to programs that involve large investments in measures and services • Participants in some C&I programs may be relatively unique and with few control group candidates • Needs to be planned as part of program implementation to allow for appropriate randomization of program participants and a control group

**This summary of pros and cons is not meant to replace the more detailed discussion in the text for guidance in application.*

3.1.2 Quasi-Experimental Designs

For most EE programs, either practical concerns or design factors will limit the use of RCT methods. In these situations, quasi-experimental designs are often a good option. ~~The use of quasi-~~Quasi-experimental designs ~~is~~are not unique to EE evaluations and ~~is~~are often used in evaluations of private and public investments. Stuart (2010) reviews the evolving research on matching and propensity scoring methods in quasi-experimental designs and states that such methods "... are gaining popularity in fields such as economics, epidemiology, medicine, and political science."^{34,35}

Quasi-experimental designs have similarities to ~~RCT~~RCTs, except that random assignment is not possible. In a quasi-experimental design, consumers typically select themselves into the participant group, and the evaluation researcher must then develop the comparison group. To avoid confusion, quasi-experimental designs use the term "~~comparison group,~~"³⁵ and RCT designs use the term "~~control group,~~"³⁶

The evaluator's goal is to select a comparison group that matches the participant group in terms of the actions that influence energy use. If done well, the only significant difference between the two groups will be participation in the program. Still, how well the comparison group actually matches the participant group will always be subject to some uncertainty, as there may be *unobservable* variables that affect energy use, the attribute of interest. Stuart (2010) defines the problem this way:

One of the key benefits of randomized experiments for estimating causal effects is that the treated and control groups are guaranteed to be only randomly different from one another on all background covariates, both observed and unobserved. Work on matching methods has examined how to replicate this as much as possible for observed covariates with observational (nonrandomized) data...

³⁴ ——— Stuart (2010) also provides a guide to software for matching ~~since,~~ because software limitations have made it difficult to implement many of the more advanced matching methods. However, recent advances have made these methods more and more accessible. This section lists some of the major matching procedures available. A continuously updated version is also available at <http://www.biostat.jhsph.edu/~estuart/propensityscoresoftware.html>. Common statistical software packages such as STATA, SAS, and R address most of the current matching approaches.

³⁵ ——— The majority of Most attribution analyses assessing business decisions and public or private investments use quasi-experimental designs, as many practical factors result in the use of this method. As an extreme example, consider a study that is designed to assess the health effects of smoking. Would it be appropriate to select a study population of 9,000 18-year-olds and assign one-third to a group that does not smoke, one-third to a group that smokes a pack of cigarettes a day, and one-third to a group that smokes a pack a day, but with some mitigating medications? Clearly, this type of RCT would pose ethical issues. As a result, natural quasi-experiments are used where ~~existing~~ smokers are matched with a comparison group of ~~non-smokers~~ nonsmokers that is as representative as possible. The methods of matching on observable characteristics have become quite advanced in the past decade.

³⁶ Technically, quasi-experimental designs do not always include a nonparticipant comparison group. For example, the interrupted time-series design (Shadish et al. 2002) relies only on aggregate participant data over time and shows this method can help control for threats to internal validity: i.e., that the results of the study are appropriately estimated for the participating customers. External validity involves generalizing; i.e., the ability of the study results to be extrapolated to other groups of customers.

While extensive time and effort [are] put into the careful design of randomized experiments, relatively little effort is put into the corresponding “design” of nonexperimental [quasi-experimental] studies. In fact, precisely because nonexperimental studies do not have the benefit of randomization, they require even more careful design.

“Matching” is broadly defined in the literature to be any method that aims to equate (or balance) the distribution of covariates in the treatment group and the comparison group. This may involve methods such as 1:1 matching (in which each participant is matched to another ~~customer that~~ consumer who did not participate), weighting, or subclassification. ~~(see Stuart 2010).~~

3.1.2.1 Matching Methods

Chapter 8 of the Uniform Method Project discusses consumption data analyses, including alternatives for constructing comparison groups. Also, the two SEE Action guides (2012a and 2012b) address matching. Matching methods include:

- **Participants as the ~~Comparison Group~~ comparison group:** SEE Action (2012b, pp. 3–6) states that among quasi-experimental approaches, “perhaps the most common [is] the ‘pre-post’ approach. With this approach, sites in the treatment group after they were enrolled in the program are compared with the same sites’ historical energy use prior to program enrollment. In effect, this means that each site in the treatment group is its own nonrandom control group.”

By using the participant group as its own comparison group, the energy use of the participants during a period before they participated in the program is used as the comparison or baseline. A statistical consumption analysis is used that also includes factors that are expected to influence energy use and may vary across the pre-post time periods. Weather is the most obvious additional variable that should be controlled, but there may be other variables as well, such as economic factors if the periods cover a ~~two~~ two-year period or ~~greater~~ longer. Agnew ~~et al.~~ and Goldberg (2013) provide a useful set of algorithms for making weather adjustments.³⁷

- **Nonparticipants ~~is~~ as the Comparison Group:** The trend in the literature is to move away from the simple approach of using participants as their own control group in a time-series analysis and, instead, to develop cross-sectional time-series data that ~~includes~~ include data on participants and matched nonparticipants.³⁸ Stuart (2010), Ho et

³⁷ ~~There are other~~ Other approaches ~~that~~ can be used for weather normalization, particularly if the evaluator is interested in changes in monthly peak demand in addition to average monthly energy use. Additional weather normalization approaches are discussed ~~in~~ by Eto (1988) and ~~in~~ McMenamin (2008).

³⁸ ~~Practical references for matching methods include:~~ (1) Stuart, E.A. (2010); (2) Ho, D. et al. (2007) and (3) Abadie and Ibens (2011).

[al. \(2007\)](#), and [Abadie and Imbens \(2011\)](#) present [practical matching methods](#). These ~~data sets~~[datasets](#) allow for the use of panel models³⁹ and DiD methods.

The simplest form of matching uses data that are already available. In the early days of ~~evaluation of~~[evaluating](#) residential programs, evaluators matched by ZIP codes, based on the assumption that consumers within the same ZIP code would have similar characteristics. However, this method is not very refined.

More recent approaches have focused on matching by energy-use and energy-use distributions across months and seasons. These matching methods can be simple or sophisticated, even when matching is confined to [available](#) energy-use data ~~already available~~ (that is, no additional surveys of nonparticipants are conducted). Matching on energy use can be as simple as stratifying participants and nonparticipants by their energy consumption (season, year, or month) and then drawing nonparticipants to match the participants' distribution of energy use.

As discussed ~~in~~[by](#) [Stuart \(2010\)](#), the literature on matching based on energy use is expanding. [Provencher et al. and Glinsmann \(2013\)](#) ~~focuses~~[focus](#) on a comparison of the distribution of energy across ~~both~~ months and seasons. The analysis follows the approach advocated by [Ho et al. \(2007\)](#) and [Stuart \(2010\)](#). The procedure used by [Provencher and Glinsmann](#) ~~involves~~[matching](#) ~~(2013)~~ ~~matches~~ each participant household to a comparison household based on a minimum distance criterion—in this case, the minimum sum of squared deviations in monthly energy consumption for the ~~three~~[3](#) months of the specified season in the pre-program year.⁴⁰

In the second step, a panel ~~data set~~[dataset](#) consisting of the monthly energy use by program households and their matched comparisons are constructed for the same season in the program year and used in a regression model predicting monthly energy use for the season. This matching is viewed by many as preferable to that involving the distribution of households across ZIP codes or demographic variables. This is because the estimate of program energy savings is based on the assumption that the comparison households are “just like” treatment households in their energy use, except for the effect of the program. Energy use is then the variable of greatest concern for the ~~non-random~~[nonrandom](#) assignment of households into the treatment and ~~the~~ control groups. To the extent that additional variables (such as heat type) are available at the ~~customer~~[consumer](#) level, the evaluator's validation of the two-stage RCT can be extended to these. However, [Provencher and Glinsmann](#) state that this is not necessary:

³⁹ ——— Panel (data) analysis is a statistical method widely used in [social science](#), ~~epidemiology~~, [social science](#), [epidemiology](#), and ~~econometrics~~[econometrics](#), which deals with two-dimensional (cross-sectional/times series) ~~panel data~~[panel data](#). The data are usually collected over time and for the same individuals.

⁴⁰ ——— In the program evaluation literature, matching often involves matching on variables with different metrics; for example, energy use and square footage of the household. These variables are normalized in the application of the distance criterion, usually using the full covariance matrix for the variables, or the inverse of the standard error for each variable (the Mahalanobis metric). When you only consider past energy use, such as monthly energy use, this sort of normalization isn't necessary because all measures are in the same units. The Mahalanobis metric is used frequently in most propensity scoring applications. The original reference is Mahalanobis (1936) and the use of the metric is covered ~~in~~[by](#) [Stuart \(2010\)](#). One application, among many examples, is [Feng \(2006\)](#), which also includes the SAS[®] code for this method.

Strong evidence that groups of households have the same distribution of energy use in the pre-program period is sufficient to establish that estimates of program savings will be unbiased. Differences that matter, such as heat type, would be revealed in the comparison of monthly energy use in the pre-program period.

These matching methods tend to follow the literature reviewed ~~in~~by Stuart (2010). Stuart indicates that matching methods have four key steps, with the first three representing the “design” and the fourth the “analysis.” These steps are:

1. ~~Defining~~“Define closeness”~~;~~: the distance measure used to determine whether an individual is a good match for another~~;~~
2. ~~Implementing~~Implement a matching method appropriate to the measure of closeness~~;~~
3. ~~Assessing~~Assess the quality of the resulting matched samples (and perhaps ~~iterating~~iterate Step 1 and Step 2 until well-matched samples result~~); and~~ ~~and~~ and
4. ~~Analyzing~~Analyze the outcome and ~~estimating~~estimate the treatment effect, given the matching done in Step 3.

In Step 1, “closeness” is often defined as a minimum distance value as used in Provencher and Glinsmann.

Another approach for identifying nonparticipants is “propensity scoring.” The most common method used in propensity score estimation involves the estimation of a logistic regression. This model uses information ~~on both~~about participants and nonparticipants to estimate a dependent variable assigned the value of 1 if that ~~customer~~consumer is a participant or 0 if the ~~customer~~consumer is a nonparticipant. This process allows for identification of nonparticipants who ~~are~~have similar ~~to participants in terms of a~~ propensity ~~score~~scores to nonparticipants (that is, similar attributes between participants and nonparticipants). This approach has a long history in in the EE evaluation literature.^{41, 42}

⁴¹ ~~The use of discrete choice methods to address self-selection bias in evaluations of EE programs has been presented in early evaluation handbooks. See: Violette et al. (1991) and Oak Ridge National Laboratories (1991). The use of discrete choice methods to address self-selection in evaluations of EE programs has been presented in early evaluation handbooks. See Violette et al. (1991) and Oak Ridge National Laboratory (1991). More recently, Bodmann (2013) used a discrete choice model to develop an instrumental variable to address omitted variable bias. However, most of these applications occurred in the 1990s, probably because the development of a discrete choice model that has adequate predictive power requires large sample sizes, which make the surveys expensive to conduct. The discrete choice model needs to be able to predict customers who choose to participate and customers who choose not to participate with appropriate reliability. This approach thus requires both participant and nonparticipant surveys. This more advanced econometric topic is not dealt with in detail in this chapter; however, several reviewers believed it was important to provide references to these methods. Heckman (1979) originally developed the two-stage model for treating self-selection. These techniques are addressed both under instrumental variables and self-selection by Kennedy (2008), who states: “Selection is not well understood by practitioners. It rests on the fundamentally on the role of an unmeasured variable and so is similar to bias created by the omission of a relevant explanatory variable” (p. 286). An updated discussion of the Heckman models for self-selection, along with appropriate caveats, can be found in Guo and Fraser (2010). Note: a link to this chapter is provided in the References section. Guo and Fraser also show how the Heckman models relate to propensity scoring.~~

3.1.2.2 Regression Discontinuity Design

SEE Action evaluation guides (2012a, 2012b) discuss the regression discontinuity design (RDD) ~~for matching~~. This method is becoming more widely used, but applies to programs where ~~there is~~ a cutoff point or other discontinuity ~~that~~ separates otherwise likely program participants into two groups. This approach ~~to matching~~ examines the impacts of a program by using a cutoff value that puts consumers ~~in~~ into or out of the program through a design that does not involve their selecting themselves into the program or choosing not to participate. As a result, this approach addresses the self-selection issue.⁴³ By comparing observations lying closely on either side of a cutoff or threshold, ~~it is possible to estimate~~ the average treatment effect in environments where randomization is not possible ~~can be estimated~~.⁴⁴ The underlying assumption in RDD is that assignment to participant and nonparticipant groups is effectively random at the threshold for treatment. If this holds, ~~then~~ those who just met the threshold for participating are comparable to those who just missed the cutoff and did not participate in the program.

The SEE Action reports indicate that RDD is a good candidate for yielding unbiased estimates of energy savings. The example used by SEE Action is based on an eligibility requirement for households to participate in a program. This requirement might be that a ~~customer~~ consumer whose energy consumption exceeds 900 kWh ~~per~~ /month would be eligible to participate in a behavior-based efficiency program, while consumers who ~~consume~~ use less than 900 kWh ~~per~~ /month would be ineligible. Thus, the group of households immediately below the usage cutoff level might be used as the comparison group.

For participating and nonparticipating households near the cutoff point of 900 kWh in monthly consumption, RDD is likely to be ~~an extremely~~ a good design. In the larger context, this RDD assumes that the program impact is constant across all ranges of the eligibility requirement variable (that is, the impact is the same for households at all levels of energy ~~usage~~ use). Evaluators ~~must~~ should consider this assumption carefully for participating households that might consume much more than 900 kWh ~~per~~ /month (for example, 2,000 kWh or more for some participants). Households with greater consumption may have greater opportunities for energy use reductions (although, the change might be constant as a percentage). In this example, potential concerns about the consistency of program impacts across different levels of household energy use makes Stuart's third step important: assessing the quality of the resulting matched samples.

Applications in the EE arena include Dubin and McFadden (1984), Goldberg and Kademan (1995), and Bodmann (2013), who used a discrete choice model to develop an instrumental variable to address omitted variable bias.

⁴² ~~—~~ Southern California Edison (2012) ~~provides~~ provides a recent behavioral impact application using propensity scoring.

⁴³ ~~—~~ In the recent years, there has been a strong movement ~~towards~~ toward focusing on the “identification” issue evaluation, that is, the issue that in the absence of a ~~randomized controlled trial~~ an RCT you do not really know if the error term in a regression is correlated with the explanatory variable of interest, so your estimate of the coefficient on that explanatory variable should be assumed to be biased in the absence of “sound” corrective action. A regression discontinuity design addresses this issue.

⁴⁴ ~~—~~ The ~~regression discontinuity design~~ (RDD) has a history in evaluation dating back to the 1960s. This approach has been used to assess a wide variety of attribution analyses in the fields of education, health, and policy. Recently, this approach has been used more often. For a review of RDD see: Imbens, ~~G.~~ and Lemieux, ~~T.~~ Regression Discontinuity Designs: A Guide to Practice, (2010, *Journal of Economic Literature* 48, 281–355).

~~The previous example is only one instance of Another~~ discontinuity. ~~Another~~ example is a time-based cutoff point. Because utilities often have annual budgets for certain programs, it is not uncommon for a program to exhaust its budget before the year is finished, sometimes within ~~six~~6 months. In this case, a date-based cutoff is useful. Consumers who apply for the program after the enrollment cutoff date imposed by budget restrictions may be similar to the program participants accepted into the program during the first ~~six~~6 months of the year. Also, both groups of consumers may have a more similar distribution of energy use per month (the focus of an impact assessment).

3.1.2.3 Random Encouragement Design

Random encouragement ~~designs~~design (RED) ~~are~~is also applicable to the types of data available for EE program evaluation. Like RDD, it is another way to incorporate randomization into the evaluation design. RED involves taking a randomly selected group of participants to receive extra encouragement, which typically takes the form of additional information or incentives. A successful encouragement design allows ~~estimation of the effect~~effects of the intervention ~~as well as the effect of the and~~ encouragement ~~itself to be estimated~~ (Diamond and Haninmueller, 2007; ~~and McKinzie, 2009~~⁴⁵). In this case, there may be an EE program for which all consumers can decide to opt in. This could be a residential audit program or a commercial audit or controls programs. A group of randomly selected consumers ~~are~~is then provided extra encouragement in terms of information and/or financial incentives. This randomization can ameliorate the effects of self-selection.⁴⁶

Fowlie and Wolfram (2009) outline an application of RED to a residential weatherization program and address the design of the study. They point out that:

REDs are particularly useful when:

- *Randomization of access or mandatory participation is not practical or desirable.*
- *There is no need to ration available services (that is, demand does not exceed supply).*
- *The effects of both participation and outreach are of interest to policy makers.*

Rather than randomize over the intervention ~~itself~~, we randomly manipulate encouragement to participate.

⁴⁵ In a position statement closely related to what EE program evaluators face, McKenzie states that “Rigorous impact evaluations, which compare the outcomes of a program or policy against an explicit counterfactual of what would have happened without the program or policy, are one of the most important tools that can be used along with appropriate economic theory for understanding “what works”. Despite this, until recently impact evaluations have been rare, especially outside the areas of health and education.”

⁴⁶ ——— The underlying estimation concept in RED is explained ~~in~~by the U.S. Department of Energy (2010): “In RED, researchers indirectly manipulate program participation using an encouragement ~~“instrument”~~“instrument” so as to generate the exogenous variation in program participation that is so essential for causal inference. This exogenous variation can then be used to identify the effect of the program on those households whose participation was contingent upon the encouragement.” Other useful references to RED are Bradlow (1998) and West (2008) as well as two documents by Fowlie and Wolfram (undated, 2010-2011); links are included in the BibliographyReferences section.

This allows the effect of the encouragement to produce exogenous variation in program participation, which can help identify the effect of the program on participants (U.S. Department of Energy, 2010).

~~There are~~ Evaluators should take certain practical issues ~~evaluators must take~~ into account in any research design, and RED is no exception. The sample sizes needed for ~~an~~ RED study are typically larger than for a pure RCT, and groups receiving the encouragement need to show different participation rates.⁴⁷ Evaluators should consider this research design when estimating net savings, as it aligns well with many standard EE program implementation plans. The random variation is designed not by excluding participants, but simply ~~through the provision of~~ ~~by~~ providing enhanced information and/or incentives offered to ~~the selected~~ consumers. ~~Work that is ongoing~~ Ongoing research work using RED should provide useful information for practitioners, but ~~few examples exist in~~ the EE evaluation literature to date ~~has few examples~~.

3.1.2.4 Summary of Quasi-Experimental Designs—Matching and Randomized Designs

~~While~~ Although it is impossible to determine definitively whether the matching, RDD, or RED designs discussed above provide an appropriate comparison group, there are tests that can provide evidence that either supports or discounts the validity of the RDD design and other quasi-experimental designs. Additionally, Fowlie ~~et al.~~ and Wolfram (2009) point out that there have been studies comparing these designs to the ideal RCT and with comparison studies that do not address systematic bias between the participant and control groups. (see Table 3). The finding is that randomized designs (either RDD or RED) ~~are an improvement over~~ improve on simple comparison approaches. RDD depends on the program having a cutoff point for participation that ~~which~~ allows for random selection. RED may be a good fit with many EE programs that have a large number of participants, but appropriate design in the types of information and incentives ~~are~~ is required. These methods should be viewed as options whenever a program contains a large number of participants, preferably 500 or more.

Importantly, these methods ~~must~~ should be considered in advance of program implementation to allow for the appropriate data, or the design of the information or incentives that will be offered potential participants, ~~to effectively implement these evaluation methods~~. It has always been important to consider evaluation when designing or revising EE programs, but the consideration of these randomized overlays to assist in evaluation makes this even more critical.

Table 3: Quasi-Experimental Designs—Summary View of Pros and Cons

Pros	<ul style="list-style-type: none"> • Limits bias if a matched comparison group can be identified regarding the actions that influence energy use • Unlike RCT, can be applied after program implementation. • Increases reliability and validity • Controls for free riders free riders and participant spillover
------	--

⁴⁷ ——— This can be one of the challenges in the design of ~~an~~ RED approach. The design of the encouragement given to a random sample of participants must be effective—; that is, produce higher acceptance rates than for the balance of the participant group.

	<ul style="list-style-type: none"> • Widely accepted in natural and social sciences when random assignment cannot be used
Cons	<ul style="list-style-type: none"> • May be difficult to identify a matched comparison group if there are unobservable variables that affect energy use • Does not address nonparticipant spillover • <u>Participants in some</u>Some C&I programs may <u>be relatively</u>have unique, <u>with participants and</u> few control group candidates

3.2 Survey-Based Approaches

This section describes the survey-based approach and the analytic use of the data obtained. Commonly conducted, surveys collect NTG-related data. ~~Despite the many drawbacks discussed within this section, this~~ This approach is typically the most can be a cost-effective, transparent, and flexible method for estimating NTG, and it has become one of the most often-used methods in EE net savings estimation. Consequently, it is the most frequently employed NTG methodology important to understand good survey design, and the strengths and weakness of these methods.

Surveys may target up to three types of respondents: (1) program participants, (2) program nonparticipants, and (3) market actors.⁴⁸ ~~While this~~ This section individually describes surveys with these three types of respondents; best practices recommend triangulating and using multiple survey approaches (for example, enhanced self-report) or multiple net savings estimation approaches.

The methods discussed in the preceding section provide estimates of net savings directly. That is, those approaches ~~either~~ compare a participant group to either a random control group (as part of an RCT) or to a comparison group from a well-designed, quasi-experimental application, and ~~these~~ these approaches do not require a separate effort to estimate ~~free ridership~~ free ridership, spillover, or market effects.⁴⁹

Survey-based approaches are used in evaluations that start with gross estimates, and then adjust for NTG factors. ~~As mentioned, surveys~~ Surveys can be a cost-efficient means to estimate NTG factors, but they are not without issues, as discussed in the following subsections. Chapter 12, Survey Design and Implementation, of the Uniform Methods Project (Baumgartner, Baumgartner (2013) also discusses many of the issues involved in using surveys to estimate NTG.

3.2.1 Program Participant Surveys

Survey-based methods for estimating net savings from program participants who are aware of the program incentives/services ~~program~~ use questions about the program's influence on the

⁴⁸ _____ Note that a Delphi panel, which also uses surveys of a panel of experts, is discussed underin Section 4.6 of this chapter.

⁴⁹ _____ Market effects can be viewed as longer-term spillover effects; therefore, it is unlikely that any market effects are included in aan RCT net savings approach spanning just a few years.

participants' actions and decision-making. Participants answer a series of closed-ended and open-ended questions on these topics:

- Why they installed the program-eligible equipment.
- What they would have done in the absence of the program incentive and ~~other~~ services.
- What further actions they took on their own because of their experiences with the program.

As noted ~~in the Chapter 12, Survey Design and Implementation, of the Uniform Methods Project (Baumgartner, by Baumgartner (2013),~~ best practice survey design for attitudes and behavior measurement use multiple-item scales to better represent the construct. ~~Since~~Because participant decision-making is complex, the survey ~~must~~should ask a carefully designed series of questions rather than a single question, as that could result in misleading findings.⁵⁰ Refer to SEE Action (2012b), Megdal et al. (2009), Haeri and Khawaja (2012), and New York Department of Public Service (2013b) for discussions about the sequencing of a series of questions.

The primary benefits of a survey-based approach are ~~as follows:~~

- ~~Implementing a~~A survey ~~typically costs~~approach can be less expensive than ~~many~~ other approaches, particularly if the effort is combined with data collection activities that are already planned for process and impact evaluations.
- The evaluator has the flexibility to tailor questions based on variations in program design or implementation methods.
- It can yield estimates of ~~freeridership~~free ridership and spillover without the need for a nonparticipant control group (NMR ~~et al.,~~Group, Inc. and Research Into Action 2010). However, participant surveys ~~only~~capture only a subset of market effects,⁵¹ a key piece of NTG.

Despite these benefits and the wide use of a survey-based self-report approach, significant concerns have been raised (Ridge et al., 2009 ~~and;~~ Peters ~~et al.,~~and McRae 2008). The main concerns are:

- ~~There is a~~A potential bias related to ~~respondents'~~respondents giving socially desirable answers.⁵²

⁵⁰ ~~Discussions of the sequencing of a series of questions can be found in SEE Action (2012b), Megdal et al. (2009), Haeri and Khawaja (2012), as well as the recent evaluation standards adopted in New York (New York Department of Public Service, July, 2013)~~

⁵¹ ~~Participant surveys can, in theory, capture end user market effects;~~ for example, changes in end-user awareness, knowledge, and efficiency-related procurement practices, ~~etc.~~

⁵² ~~Participants may also have a bias toward overstating program impacts~~ due to the desire because they want to retain incentives, although this has not been widely documented.

DRAFT #4 | Not for Citation

- The inability of consumers to know what they would have done in a hypothetical alternative situation, especially in current program designs that use multiple methods to influence behavior.
- The tendency of respondents to rationalize past decisions ~~on choices~~.
- ~~There is a~~ potential for arbitrariness in the scoring methods that translate responses into ~~free rider~~ free rider estimates.
- Consumers may fail to recognize the influence of the program on other parties who influenced their decisions. For example, a program having market effects may have influenced contractor practices, which, in turn, may have indirectly impacted the participants' (and nonparticipants') decisions.

Ridge et al. (2009) point out that, although these concerns are valid, ~~they are widely acknowledged by social scientists who have worked on a variety of methods over the years to address them.~~ It is also important to ~~not recognize~~ that all ~~methodologies~~ methods have ~~inherent~~ potential biases.⁵³ For example, market sales analysis,⁵⁴ which is based on objective sales data, can be biased if the market actors who provide data for the analysis operate differently from those not participating in the study or if the comparison area is systematically ~~non-comparable~~ noncomparable.

Ridge et al. (2009) point out that it does not make sense to compare all self-report approaches equally, as some conform to best practice, ~~while and~~ others do not. Keating (2009) adds that many of the criticisms of the self-report approach can be alleviated through careful research design, sampling, survey timing, and wording of questions.

~~In Chapter 12 of the Uniform Methods Project,~~ Baumgartner (2013) presents guidelines for selecting appropriate survey designs and recommends procedures for administering best practice surveys. The literature also contains a number of best practice elements for survey design, data collection, and analytic methods specific to estimating net savings (NYS DPS, New York State Department of Public Service 2013; Tetra Tech et al., 2011; ~~Ridge et al. 2009~~). This literature notes the importance of making the entire process transparent so ~~that~~ stakeholders can understand how each question and its ~~responses impact~~ response impacts the final estimate. Thus, the report should contain details of critical elements such as the question sequence, scoring algorithms, and the handling of inconsistent and/or missing data.

Tetra Tech et al. (2011) present some of the best practices regarding survey design, data collection, and analytic elements related to net savings estimation ~~are presented here (Tetra Tech et al. 2011).~~

⁵³ This is, of course, the primary motivation for triangulation.

⁵⁴ ———Market sales analysis captures the total net effect of a program. Ideally, this method involves obtaining comprehensive pre- and post-market sales data in both the area of interest and in an appropriate comparison area and examining the change in the program area compared with the change in the ~~non-program~~ nonprogram area (Tetra Tech et al., 2011).

3.2.1.1 Survey Design Elements⁵⁵

A number of design elements need to be considered when developing surveys. Best practices for choosing design elements include:

- ~~Identifying~~Identify the key decision-maker(s) for the specific ~~energy-efficient~~EE project. For downstream programs, a key decision-maker in the household or business is likely to be responsible for making the final decision, although they may assert that their vendor was the most influential in their decision. Although consumers ultimately decide ~~on~~ what they will purchase, ~~consumers~~they may not be aware of the influence of the interventions for upstream programs where trade ally decisions are driving change (for example, original equipment manufacturers determine equipment ~~energy efficiency~~EE levels and retailers determine what equipment to stock and market, or advertise as a result of upstream program incentives).
- ~~Using set-up~~Use setup or ~~warm-up~~warmup questions to help the decision-maker(s) recall the sequence of past events and how these events affected their decision to adopt the measure.
- ~~Using~~Use multiple questions to limit the potential for misunderstanding or the influence of individual anomalous responses.
- ~~Using~~Use questions that rule out rival hypotheses for installing the efficient equipment.
- ~~Testing~~Test the questions for validity and reliability.
- ~~Using~~Use consistency checks when conducting the survey to immediately clarify inconsistent responses.
- ~~Using~~Use measure-specific questions to improve the respondent’s ability to provide concrete answers, and ~~recognizing~~recognize that respondents may have different motivations for installing different measures.
- ~~Using~~Use questions that capture partial efficiency ~~improvement~~improvements (accounting for savings above baseline but less than program eligible), quantity purchased, and timing of the purchase (where applicable for a measure) to estimate partial ~~freeridership~~free ridership.
- ~~Using~~Use neutral language that does not lead the respondent to an “expected” answer.
- ~~Using~~Use combinations of open- and close-ended questions to balance hearing from the end users in their own words and ~~creating~~create an efficient, structured, and internally consistent ~~data set~~dataset.

3.2.1.2 Data Collection Elements

Even when the survey design is effective, data collection ~~must~~should also follow best practices for collecting reliable information and calculating valid estimates. These ~~data collection~~ practices include:

⁵⁵ ———Comments received from chapter reviewers and, in particular ~~Mr.~~Michael Rufo, Itron Inc., provided additional contribution to this section.

- ~~Pre-testing~~Pretest the survey instrument to ensure that questions are understandable, skip patterns are correct, and the interview flows smoothly. The pretesting should use, when possible, cognitive interviewing techniques (Miller 2011).⁵⁶
- ~~Using~~Use techniques to minimize nonresponse bias, such as advance letters on utility or program administrator letterhead (the organization for which the participant will most likely associate the program) and multiple follow-ups over a number of weeks.
- ~~Following~~Follow professional standards for conducting surveys, which include training and monitoring interviewers.⁵⁷
- ~~Determining~~Determine the necessary expertise of the interviewer based on the complexity and value of the interview (for example, it is better for trained evaluation professionals rather than general telephone surveyors to address the largest, most complex projects in custom programs ~~rather than general telephone surveyors~~).
- ~~Timing~~Time the data collection so it occurs as soon as possible after ~~installation of a measure~~ is installed, as this minimizes recall bias and provides timely feedback on program design. Recognize, however, that timely data collection for estimating ~~freeridership~~free ridership will underestimate participant spillover, as little time may have passed since program participation. ~~Although, conducting~~Conducting a separate spillover survey at a later date with these same participants can alleviate this. Having a separate survey will increase data collection costs, but may be warranted if spillover effects are likely to have occurred.
- ~~Sampling~~Sample (or oversample) a census of ~~(or oversampling)~~ the largest savers and, depending ~~upon~~on program participation, ~~samplingsample~~ end- uses with few installations to ensure the measures are sufficiently represented in the survey sample.

3.2.1.3 Analytic Elements

In addition to discussing survey design and data collection elements, much of the literature discusses best practices for analysis such as:

- ~~Treating~~Treat acceleration of the installation of the EE measures appropriately to produce lifetime net savings rather than first-year net savings (this requires understanding the program's influence on the timing of the project).⁵⁸

⁵⁶ In cognitive interviews, respondents are asked to describe how and why they answered the question as they did. Miller (2011) notes that “through the interviewing process, various types of question response problems that would not normally be identified in a traditional survey interview, such as interpretive errors and recall accuracy, are uncovered” (p. 54).

⁵⁷ _____ Data collections surveys can be conducted via telephone, ~~web~~the Web (including ~~smart~~ phones~~smartphones~~), postal mail, and in-person. For large complex C&I projects, an energy engineer who is knowledgeable ~~with~~about the type of project and technology should conduct the interviews.

⁵⁸ A chapter review comment provided by Mr. Michael Rufo, Itron, notes that “A focus on program induced early replacement versus the effect on efficiency level is gaining attention in the evaluation field. In cases where there is early replacement, two net savings components may be needed to appropriately characterize overall net savings: (1) the early replacement period that uses an in situ baseline; and, (2) the efficiency increment above

- ~~Incorporating~~Incorporate the influence of previous participation in the program.
- ~~Establishing~~Establish *a priori* rules for treatment of missing/~~don't~~don't knows in the scoring algorithm.
- ~~Weighting~~Weight the estimates by annual savings to account for the size of the savings impacts for each ~~customer~~consumer.
- ~~Sampling, calculating~~Sample, calculate, and ~~reporting~~report the precision⁵⁹ of the estimate for the design element of interest (measure, project type, or end-use).
- ~~Conducting~~Conduct sensitivity testing of the scoring algorithm.
- ~~Defining~~Define what the spillover measurement is—and is not; attempting to estimate and justify the use of an approach.
- ~~Employing~~Employ, where feasible, a preponderance of evidence (or triangulation of results) approach that uses data from multiple sources,⁶⁰ ~~(see Itron, Inc. 2010)~~, especially for large savers and complex decision-making cases. Potential data sources could include project file reviews, program staff and account manager interviews, vendor interviews, and observations from site visits.

~~The~~ New York Department of Public Service (2012) developed additional guidelines specific to the estimation of spillover savings to address recurring methodological limitations that the New York Department of Public Service staff and its contractor team observed in the estimation of spillover in New York and the industry as a whole. Prahl et al. (2013) ~~summarizes~~summarize this work and the critical decisions that evaluators must make before deciding whether and how to estimate spillover. That paper also discusses how the estimation of per-unit gross savings, estimation of program influence, and documentation of causal mechanisms varies for different levels of rigor.

3.2.2 Surveys of Program Nonparticipants

Self-report surveys with nonparticipants are commonly used to triangulate participant self-report responses and collect data for calculating nonparticipant spillover or market effects. These surveys help evaluators understand what EE actions nonparticipants have taken and whether they took those actions because of program influences (nonparticipant spillover). Conducting surveys with nonparticipants poses its own unique challenges. ~~First, there:~~

minimum or standard practice at the end of the early adoption period (that is, one for the RUL (remaining useful life) period and one for the remainder of the EUL.²² [effective useful life].”

⁵⁹ ~~Appendix H of the Evaluation Plan Guidance for EEPS Program Administrators in New York (New York DPS, 2013)~~ The New York Department of Public Service (2013a) presents guidelines for calculating the relative precision of program net savings estimates for different types of estimates, including the NTG ratio based on the self-report method and for spillover savings. Additional discussion of sampling for evaluation can be found in Khawaja et al. (2013).

⁶⁰ ~~See, for example, Itron, February, 2010. 2006–2008 Evaluation Report for PG&E Fabrication, Process and Manufacturing Contract Group, for the California Public Utilities Commission, Energy Division.~~

- There is no record of the equipment purchase, and identifying a group of nonparticipants who have installed energy-efficient equipment on their own can be time consuming and costly.⁶¹ ~~Second, establishing~~
- Establishing causality entails estimating gross unit savings (often with limited evidence other than the consumer self-report) and establishing how the program may have influenced the consumer's decision. The consumer may not have been aware, for example, of the influence the program had on the equipment's availability or the market actor's stocking practices.

3.2.3 Market Actor Surveys

When estimating net savings, it is important to consider all the ~~various~~ points of program influence. In addition to targeting consumers, upstream and midstream programs often target program services and/or funding to market actors (such as contractors, auditors, and design specialists) with the goal of influencing their design, specification, recommendation, and installation practices. ~~Thus, in~~ upstream and midstream programs, consumers may not be aware of program influences on sales, stocking practices, or prices (discussed in the Appendix A).⁶² ~~As a result, it is not appropriate to use~~.⁶³ Thus, using only participant self-reports when estimating net savings: is inappropriate. In these cases, evaluators use market actor self-report surveys to examine the effecteffects of these upstream influences.

These market actor self-report surveys can be designed as qualitative in-depth interviews or as structured surveys with a statistically designed sample of contractors. The use and application of the data determine the format ~~of the survey~~. For example, evaluators may use:

⁶¹ _____
⁶¹ _____ One approach to mitigating the efficiency and cost of this is to use one nonparticipant survey that asks about a variety of program eligible measures and use the results across multiple programs.

⁶² ~~There are studies that focus on examining how a change in the price of an energy-efficient product influences consumer purchases. Two approaches were used: (1) stated preference experiments that systematically ask potential consumers what they would choose from a set of options with different features and prices and (2) revealed preference studies observe the actual choices consumers make from true choices available to them when making purchases. To obtain accurate revealed preference information, it is usually necessary to observe the items purchased. Consumers cannot reliably report the efficiency levels of recently purchased equipment. Direct observation can be accomplished via store intercepts for small items such as light bulbs, or via onsite visits for large items such as refrigerators. The remaining challenge for this method is the potential nonresponse bias; that is, potential differences between consumers willing to have their purchases observed and those who decline. An example of a study that focuses on how changes in price influence consumer purchases of energy-efficient products is Cadmus Group (2012). See Appendix A for additional information.~~

⁶³ There are studies that focus on examining how a change in the price of an energy-efficient product influences consumer purchases. Two approaches were used: (1) stated preference experiments that systematically ask potential consumers what they would choose from a set of options with different features and prices; and (2) revealed preference studies observe the actual choices consumers make from true choices available to them when making purchases. To obtain accurate revealed preference information, it is usually necessary to observe the items purchased. Consumers cannot reliably report the efficiency levels of recently purchased equipment. Direct observation can be accomplished via store intercepts for small items such as light bulbs, or via onsite visits for large items such as refrigerators. The remaining challenge for this method is the potential nonresponse bias; that is, potential differences between consumers who are willing to have their purchases observed and those who decline. An example of a study that focuses on how changes in price influence consumer purchases of energy efficient products is Cadmus (2012b) See the Appendix for additional information.

DRAFT #4 | Not for Citation

- Qualitative, open-ended data based on a small sample of market actors to contextualize market actors' practices (best used for triangulation purposes).
- Quantitative market actor data to calculate [freeridershipfree ridership](#) and spillover rates specifically related to the practices of those market actors. The calculated rates can then be directly integrated with participant self-report results, triangulated with participant self-report results, and/or used as the sole source for [freeridershipfree ridership](#) and spillover rates. ([See, for example, KEMA, Inc. \[2010\].](#))

Evaluations can also include market actor survey data to estimate nonparticipant spillover and market effects. An important issue related to the quantification of nonparticipant spillover savings using only surveys of consumers is valuing the savings of measures installed outside the program. As previously noted, during telephone interviews, consumers often cannot provide adequate equipment-specific data on new equipment installed either through or outside a program. Although they [arecan](#) usually [able to](#) report what type of equipment was installed, consumers typically cannot provide sufficient information about the quantity, size, efficiency, and/or operation of that equipment to [makeenable](#) a determination about its program eligibility.

One approach to estimating nonparticipant spillover and market effects via market actors is to ask market actors questions such as:

- What percentage of their sales meets or exceeds the program standards for each program measure category installed through the program(s)?
- What percentage of these sales did not receive an incentive?

The market actors should then be asked several questions about the program's impact on their [decisiondecisions](#) to recommend and/or install this efficient equipment outside the program.

3.2.4 Case Studies for Estimating Net Savings Using Survey Approaches

This section presents three examples of estimating net savings with self-report surveys. The first example demonstrates how the participant self-reports method was used to calculate [freeridershipfree ridership](#) of nonresidential programs in California. The second demonstrates how a sample set of survey questions were used in conjunction with a matrix to estimate [freeridership-free ridership](#). The final example summarizes an approach used by the Energy Trust of Oregon (Castor, 2012) that calculates low, mid, and high scenario NTG ratios to account for "Don't Know" responses to certain questions. This example addresses the best practice of conducting sensitivity analysis on the algorithm used to estimate NTG.

Example 1. Nonresidential Programs [FreeridershipFree Ridership](#) Assessment

The Large Nonresidential Freeridership Approach, developed by the [nonresidentialNonresidential](#) Net-to-Gross Ratio Working Group for the Energy Division of the California Public Utilities Commission (2012), was developed to address the unique needs of large nonresidential customer projects developed through [energy efficiencyEE](#) programs offered by the four California investor-owned utilities and other third-parties. [The Large Nonresidential Freeridership Approach is based on an approach that has been evolving for more than 15 years.](#) As described in the framework, the method relies exclusively on the self-report approach to estimate project- and program-level NTG ratios, [asbecause](#) the working group notes that other

available methods and research designs are generally not feasible for large nonresidential customer programs. This methodology provides a standard framework, including decision rules, for integrating findings from ~~both~~ quantitative and qualitative information in the ~~calculation of the net to gross ratio in a~~ systematic and consistent ~~manner~~ calculation of the NTG ratio.

The approach describes three levels of ~~freeridership~~ free ridership analysis. The most detailed level of analysis, the Standard – Very Large Project NTG ratio, is applied to the largest and most complex projects (representing ~~10 to~~ 20% of the total projects) with the greatest expected levels of gross savings. The Standard NTG ratio, involving a somewhat less detailed level of analysis, is applied to projects with moderately high levels of gross savings. The Basic NTG ratio is applied to all remaining projects.

~~There are five~~ Five potential sources of ~~freeridership~~ free ridership information are discussed in this study. Each level of analysis relies on information from one or more of these sources:

- **Program files**, which can include various pieces of information relevant to the analysis of ~~freeridership~~ free ridership. Program files may include ~~as~~ letters written by the utility's customer representatives that document what the ~~eustomer~~ consumer had planned to do in the absence of the rebate and explain the ~~eustomer's~~ consumer's motivation for implementing the ~~efficiency~~ EE measure. It can also include information on the measure payback with and without the rebate.
- **Decision-maker surveys**, conducted with the person involved in the decision-making process that led to the implementation of measures under the program. This survey obtains highly structured responses concerning the probability that the ~~eustomer~~ consumer would have implemented the same measure in the absence of the program. ~~First,~~ participants
 - Participants are asked about the timing of their program awareness relative to their decision to purchase or implement the ~~energy efficiency~~ EE measure. ~~Next,~~ they
 - They are asked to rate the importance of the program versus ~~non-~~ program influences in their decision-making. ~~Third,~~ they
 - They are asked to rate the significance of various factors and events that may have led to their decision to implement the ~~energy efficiency~~ EE measure at the time that they did (for example, age or condition of the equipment, information from a facility audit, standard business practices, and ~~prior~~ experience with the program or measure).

~~In addition, the~~ The survey ~~obtains a description of what~~ also asks participants to describe what they would have done in the absence of the program, beginning with whether the implementation was an early replacement action. ~~If not, the~~ The decision-makers are asked to ~~provide a description of~~ describe the equipment they would have installed in the absence of the program, including the efficiency ~~level~~ levels and quantities. This information is used to adjust the gross engineering savings estimate for partial ~~freeridership~~ free ridership.

DRAFT #4 | Not for Citation

This survey contains a core set of questions for Basic NTG ratio sites, and several supplemental questions for both Standard and Standard – Very Large NTG ratio sites. For example, if Standard or Standard – Very Large respondents indicate that a financial calculation entered highly into their decision, they are asked additional questions about their *financial criteria* for investments and their rationale for the current project. These questions are intended to provide a deeper understanding of the decision-making process and the likely level of program influence versus these internal policies and procedures. Responses to these questions also serve as a basis for consistency checks to investigate conflicting answers ~~regarding about~~ the relative importance of the program and other elements in influencing the decision. ~~In addition,~~ Standard – Very Large respondents may ~~also~~ receive additional detailed probing on various aspects of their installation decision based on industry- or technology-specific issues, as determined by review of other information sources. For Standard – Very Large sites, the respondent data are used to construct an internally consistent “story” that supports the NTG ratio calculated, based on the overall feedback.

- **Vendor ~~Surveyssurveys~~** are completed for all Standard and Standard – Very Large participants ~~thatwho~~ used vendors, as well as for Basic participants ~~thatwho~~ indicate a high level of vendor influence in the decision to implement the EE measure. For participants ~~thatwho~~ indicate the vendor was very influential in decision-making, the vendor survey results are incorporated directly into the NTG ratio scoring.
- **Utility and ~~Program Staff Interviews~~ program staff interviews** for the Standard and Standard – Very Large NTG ratio analyses. Interviews with utility staff and program staff are also conducted to gather information on the historical background of the ~~customer's~~ consumer's decision to install the efficient equipment, the role of the utility and program staff in this decision, and the ~~name~~ names and contact information of vendors involved in the specification and installation of the equipment.
- **Other information** for Standard – Very Large Project NTG ratio sites includes secondary research of other pertinent data sources. For example, this could include a review of standard and best practices through industry associations, industry experts, and information from secondary sources (such as the U.S. Department of ~~Energy's~~ Energy's Industrial Technologies Program's Best Practices website).⁶⁴ In addition, the Standard – Very Large NTG ratio analysis calls for interviews with other employees at the participant's firm, sometimes in other states, and equipment vendor experts from other states where the rebated equipment is installed (some without rebates) to provide further input on standard practice within each company.

Table 4 shows the data sources used in each of the three levels of ~~freeridership~~ free ridership analysis. Although more than one level of analysis may share the same source, the amount of information ~~utilized~~ used in the analysis may vary. For example, all three levels of analysis obtain core question data from the decision-maker survey.

⁶⁴ This website can be found at: <http://www1.eere.energy.gov/industry/bestpractices/>.

Table 4: Information Sources for the Three Levels of NTG Ratio Analysis

	Program File	Decision-Maker Survey Core Question	Vendor Surveys	Decision-Maker Survey Supplemental Questions	Utility & Program Staff Interviews	Other Research Findings
Basic NTG ratio	√	√	√ ¹		√ ²	
Standard NTG ratio	√	√	√ ¹	√	√	
Standard NTG ratio—Very Large Projects	√	√	√ ³	√	√	√

¹ Only performed for sites that indicate a vendor influence score greater than maximum of the other program element scores.

² Only performed for sites that have a utility account representative.

³ Only performed if significant vendor influence is reported or if secondary research indicates the installed measure may be becoming standard practice.

Example 2. Free Ridership Assessment for an Equipment Rebate Program

This example shows how to calculate an NTG ratio and how to use a sample set of survey questions in conjunction with a matrix to estimate free ridership (see Table 5). The example is from Chapter 5 of the Energy Efficiency Program Impact Evaluation Guide (SEE Action, 2012b). In this case, the evaluators assign a free ridership score based on a participant’s response to six questions.

Table 5. Assignment of Free Ridership Score Based on Participant Responses

Free Ridership Score	Already Ordered or Installed	Would Have Installed without Program	Same Efficiency	Would Have Installed All the Measures	Planning to Install Soon	Already in Budget
100%	Yes	Yes	—	—	—	—
0%	No	No	—	—	—	—
0%	No	Yes	No	—	—	—
50%	No	Yes	Yes	Yes	Yes	Yes
25%	No	Yes	Yes	Yes	No	Yes
25%	No	Yes	Yes	Yes	Yes	No
0%	No	Yes	Yes	Yes	No	No
25%	No	Yes	Yes	No	Yes	Yes
12.5%	No	Yes	Yes	No	No	Yes
12.5%	No	Yes	Yes	No	Yes	No
0%	No	Yes	Yes	No	No	No

Source: SEE Action (2012b) based on example provided by Cadmus.

One issue with this method is the somewhat arbitrary nature of assigning free ridership scores based on sets of question responses, as they are dependent on the judgment of the particular evaluator. Different researchers may assign different free ridership scores to different sets of respondent answers. To address this, the literature

recommends using sensitivity analyses around the [freeridershipfree ridership](#) scores, based on the judgments of people familiar with the program.⁶⁵ An example of increasing the robustness of this method is found in an assessment of residential heating and cooling equipment for the Electric and Gas Program Administrators of Massachusetts.⁶⁶

~~Other approaches use upper and lower bounds on freeridership~~ Another useful exercise is to assess the reliability of the assignment of free ridership scores by the evaluators. Inter-rater reliability scores⁶⁷ can be calculated to assess the reliability of these assignments. To the extent that evaluators assign the same free ridership scores to the same set of response patterns, then reliability will be increased. Other approaches use upper and lower bounds on free ridership developed directly from survey respondents.⁶⁸

Example 3. Commercial, Industrial, and Residential Scenario Analysis

The Energy Trust of Oregon uses an approach (Castor, 2012) to calculate low, mid, and high scenario NTG ratios to account for the “Don’t Know” responses to certain questions. The report appendix describes this approach. The project’s [freeridershipfree ridership](#) score is composed of two elements: (1) a project change score and (2) an influence score.

⁶⁵ ——— Issues may arise if these [freeridershipfree ridership](#) scores are viewed as categories rather than as continuous variables. A 50% score may imply a higher level of [freeridershipfree ridership](#) than does a 25% score, but it may not denote that the 50% score implies that [freeridershipfree ridership](#) is, in fact, twice as high when compared to respondents placed in 25% [freeridershipfree ridership](#) score category. It is possible to perform arithmetic on these numbers and use the values to generate a mean value and even a variance, but this may not be appropriate. The lack of an accurate “distance” factor in these numbers makes the calculated variance hard to interpret. For variables that are meant to represent categories rather than continuous numeric values, frequencies are the more often used descriptive statistic.

⁶⁶ ——— This work was conducted by a consortium of consultants under a prime contract led by ~~The Cadmus Group~~, supported by Navigant, and Opinion Dynamics Corporation. ~~(Cited (cited as Cadmus-, Navigant Consulting; Opinion Dynamics Corporation (2012).~~

⁶⁷ [Inter-rater reliability, inter-observer reliability, and inter-judge agreement are some terms that have been used in the literature to designate a wide variety of concepts. All these terms, however, refer to the extent of agreement among raters, judges, and observers \(Gwet 2010, 2012\).](#)

⁶⁸ ——— Violette et al. (2005) discuss approaches used in the net savings and attribution assessment for a large-scale C&I retrofit program. [freeridershipFreeridership](#) was assessed using a series of survey questions asked of various actors, including participating end-use consumers and vendors/contractors/consultants. [freeridershipFreeridership](#) was asked in both direct freeridership questions and in supporting, or influencing, questions. Participating owners and [ESCOenergy service companies](#)/contractors in a large-scale C&I retrofit program were each asked for direct estimates of: (1) ~~The~~the “proportion” of the savings or measures that would have been installed without the program; and (2) ~~The~~the “likelihood” that the measures would have been installed without the program. A three-step approach was used. Step 1 focused on whether the respondent believed that freeridership existed at all; if the respondent believed it existed in this project, Step 2 established bounds on the freeridership effect, that is, what was the smallest value that seemed reasonable and what might have been the highest reasonable freeridership value. Step 3 used questions to obtain where within this range the freeridership value was likely to fall. Appendices to Violette et al. (2005) discuss alternative approaches. This program had some unique characteristics that made this approach more tractable. It involved large-scale C&I projects and the survey respondents were provided with summaries of the technologies and measures installed. [Other efforts that used similar approaches include Violette, Ozog and Cooney \(2003\) for addressing net savings from regional and market transformation programs in the Pacific Northwest, and Navigant \(2013b\) which assesses the net impacts of U.S. DOE’s Wind Powering America Initiative.](#)

DRAFT #4 | Not for Citation

The project change score is based on the respondent's answer to the question, "Which of the following statements describe the actions you would have taken if Energy Trust incentives and information were not available"?

Possible answer choices are assigned a number between 0 and 0.5, with 0 indicating no ~~freeridership~~free ridership and 0.5 indicating that the participant was a full ~~freerider~~. ~~Since respondents~~free rider. ~~Because a respondent~~ can select multiple responses to the question, ~~their~~the answer choice with the lowest score is selected. If the respondent selects "Don't Know," two scores are created to account for the range of possible answers (0 and 0.5).

For commercial projects, respondents are asked this follow-up question when they report they would not have done anything differently in the absence of the program: "If your firm had not received the incentive, would it have made available the funds needed to cover the entire cost of the project"? If the respondents select "Yes," their project change score is 0.5. If the respondents select "No," their project change score is 0. However, if the respondents select "Don't Know," they are given two scores for project change, as previously described.

The influence score is based on respondents' answers to questions about the influence of Energy Trust incentives, program representatives, contractor/salesperson, studies, and other program elements. The answer choices are given a value between 0 (element's influence was a 5, extremely influential) and 0.5 (element's influence was a 1, not at all influential). The score for the most influential element is taken as the influence score. If respondents answer "Don't Know" for all elements, they are given two influence scores to account for the range of possible answers (0 and 0.5).

To generate the ~~freeridership~~free ridership score for each project, the project change and influence scores are added. For respondents who do not provide "Don't Know" answers, this score will be a single number between 0 (no ~~freeridership~~free ridership) and 1 (full ~~freeridership~~free ridership). For those who gave a "Don't ~~know~~Know" answer to one of the questions, there are two ~~freeridership~~free ridership scores—one high and one low. For those who answered "Don't ~~know~~Know" to *both* the project change and influence questions, no score is calculated.

~~Freeridership~~Free ridership scores are averaged for all respondents in each program/measure group and the result is shown as a percentage rather than a decimal- (see Table 6 for pros and cons of survey-based approaches).

- "Low Scenario" is the average of the ~~freeridership~~free ridership scores where the low score is used for those who answered "Don't ~~know~~Know" to a question.
- "High Scenario" is the average where the high score is used for those who answered "Don't know" to a question.
- "Mid Scenario" is the average of the Low and High Scenarios. In the case of C&I projects, individual scores are weighted by their share in the ~~electrification~~electricity or gas savings of all respondents of their group before the scores are averaged for scenarios.

Table 6-2. Survey-Based Approaches—Summary View of Pros and Cons

Pros	<ul style="list-style-type: none"> • Can provide useful information to support process and impact evaluations (for example, source of awareness, satisfaction, and demographics) • Flexible approach that allows the evaluator to tailor questions to the program design or implementation methods • Can yield estimates of freeridership free ridership and spillover without the need for a nonparticipant control group
Cons	<ul style="list-style-type: none"> • Potential biases related to respondents' giving "socially desirable" answers • Consumers' inability to know what they would have done in a hypothetical alternative situation, especially in current program designs that use multiple methods to influence behavior • The tendency of respondents to rationalize past choices • Potential arbitrariness of scoring methods based on evaluator judgment that translate responses into freerider free rider estimates • Consumers may fail to recognize the influence the program may have had on other parties who influenced their decisions (for example, program may have influenced contractor practices, which in turn impacted the participant) • Participant surveys only capture only a subset of market effects

3.3 Common Practice Baseline Approaches

The common practice baseline approach⁶⁹ is ~~relatively new in the broader evaluation literature and its application has been limited. However, the Northwest Power and Conservation Council (NW Council) in the Pacific Northwest has applied it also is receiving attention as a method for a number of years.~~⁷⁰ [estimating net savings.](#) SEE [Action describes \(2012b\) has defined](#) the common practice baseline as [follows](#):

Common practice baselines are estimates of what a typical consumer would have done at the time of the project implementation. Essentially, what is “commonly done” becomes the basis for baseline energy consumption (SEE Action, 2012b, p. 7-2).^{71,72}

⁶⁹ —The Common Practice Baseline section gave rise to a number of comments. Some reviewers did not see this method as parallel to the other methods presented in this chapter, as it focuses on *ex ante* values of the mean of market behavior and does not look at *ex post* information on actions or program participants. In this context, this approach was viewed as more of an *ex ante* deemed net savings approach (see [section Section 3.7](#) on deemed NTG values [below](#)). After considering these comments, the Common Practice Baseline approach was viewed as warranting a separate section due, in part, to the recent attention given this approach to net savings.

⁷⁰ —Comments provided by Mr. Tom Eckman of the Northwest Power and Conservation Council (NW Council) indicated that this general approach has been applied in setting deemed savings since the 1980s, and it was designed to fit with the NW Council integrated planning process, that is, it is meant to provide an estimate of the increment of savings beyond what system planners assume for naturally (or currently) occurring efficiency in their demand models. Additional information on this can be found at the Regional Technical Forum website of the NW Council — <http://rtf.nwccouncil.org>.

⁷¹ —SEE Action (2012b) illustrates this “commonly done” baseline using an appliance example. “For example, if the program involves incenting consumers to buy high-efficiency refrigerators that use 20% less energy than the minimum requirements for ENERGY STAR[®] refrigerators, the common practice baseline would be refrigerators that consumers typically buy. This might be non-ENERGY STAR refrigerators, or ENERGY STAR refrigerators, or, on average, something in between.”

~~This approach is based on the development of a market-based common practice baseline (sometimes called standard industry practice). This baseline includes a “consideration of what typically would have been done in the absence of the efficiency action” (SEE Action 2012b). This approach is under development in several jurisdictions and will certainly evolve in its application. In general, it is based on using available information to develop an *ex ante* estimate of net savings, with limited adjustments based on *ex post* data and analysis. This approach has many appealing qualities, but the tradeoffs need to be clarified, both in terms of potential biases and the real costs associated with this approach.~~

~~The common practice baseline method is relatively new in the broader evaluation literature and its application has been somewhat limited; however, the Northwest Power and Conservation Council (NW Council) in the Pacific Northwest has applied a variant of this method for a number of years in estimating *ex ante* net savings.⁷³ The NW Council continues to evolve this approach with new protocols developed by the Regional Technical Forum (RTF 2012).⁷⁴ Ridge et al., 2012b)-(2013) indicate that, in addition to the NW Council, three other jurisdictions are working with variants of the common practice baseline approach: Northwest Energy Efficiency Alliance (NEEA), Indiana, and Delaware.~~

As with other net savings approaches, the common practice baseline approach is designed to assess the savings attributable to EE program activities. One advantage claimed for the common baseline approach is that it avoids double counting of ~~freeriders~~ free riders. The concern is that the two-step approach—where (1) gross savings is estimated ~~first~~ *ex post* using current practice as the baseline; and ~~then a second step estimates freeridership and spillover~~ (2) an NTG ratio is applied to the *ex post* gross savings—can double count at least some ~~freeriders~~ free riders (Ridge et al. 2013; Hall et al. 2013). The argument is that the estimated claimed (*ex ante*) gross savings

⁷² ~~The SEE Action report~~ (2012b) defines common practice baselines in its glossary as “The predominant technology(ies) implemented or practice(s) undertaken in a particular region or sector.” (p. A-4).

⁷³ Tom Eckman of NW Council indicated that this general approach has been applied in setting deemed savings since the 1980s, and it was designed to fit with the NW Council integrated planning process; that is, it is meant to provide an estimate of the increment of savings beyond what system planners assume for naturally (or currently) occurring efficiency in their demand models. Additional information can be found at the RTF website of the NW Council and in RTF (2012).

⁷⁴ Some reviewers indicated that this double counting problem may be the result of inconsistent program rules as set out by the program administrators and regulators, and was not an estimation issue. Further, a number of reviewers indicated that rather than over-estimating freeriders, this approach underestimates freeriders due to selection bias (discussed in the main body text below). The RTF guidelines (dated August 15, 2012) sets out the current practice baseline approach most directly in its definition of savings: “Savings is defined as the difference in energy use between the baseline (see section 2.2) and post (after measure delivery) periods, which is caused by the delivery of a measure. The terms “net” or “gross” are intentionally not used to modify the term “savings,” as they may conflict with the definition of “baseline,” provided in section 2.2. The current practice baseline defines directly the conditions that would prevail in the absence of the program (the counterfactual), as dictated by codes and standards or the current practices of the market. The most important conflict would arise if savings were estimated against a current practice baseline and then those savings were further adjusted by a net-to-gross ratio, where the net-to-gross ratio was the probability that the measure would have been delivered in the absence of program influence.” Note that the RTF uses the term *current baseline* rather than *common practice baseline* used elsewhere.

may be closer to net savings than the estimates of net savings calculated by adjusting the gross savings estimates by ~~freeridership~~ free ridership, spillover, and market effects. This is because some of these factors are already contained in the process used to produce the gross savings estimates. Hall et al. (2013) ~~state~~ point out that “if a baseline approach already has ~~freeriders~~ incorporated free riders in ~~its construction~~, there is often no need to readjust the savings calculation to account for ~~freeriders~~ free riders a second time.”⁷⁵ This emphasizes the need to: (1) understand the derivation of gross estimates as part of the EE evaluation process, and (2) to explicitly set out the assumed counterfactual scenario in the net savings method used. Taking these two steps avoids the double counting that results in higher-than-appropriate ~~freeridership~~ estimates-free ridership estimates.⁷⁵~~76~~ ~~Ridge et al. (2013) indicates that, in addition to the NW Council using it,~~⁷⁷ three other jurisdictions—the Northwest Energy Efficiency Alliance (NEEA), Indiana, and Delaware—have recently adopted the common practice baseline approach as one method for addressing net savings.

Examples from ~~the~~ guidelines on common practice baselines include:

- NW Council’s ~~Guidelines Savings Estimation Methods~~ guidelines savings estimation methods: The current practice baseline defines directly the conditions that would prevail in the absence of the program (the counterfactual scenario), as dictated by codes and standards or the current practices of the market. (RTF, 2012, p. 2). In the guidelines developed by the RTF, the impact estimation methods are grouped by the type of RTF

⁷⁵ Some reviewers indicated that this double counting problem may be the result of inconsistent program rules as set out by the program administrators and regulators, and is not an estimation issue. If this is the case, evaluators still must decide whether the *ex ante* savings are net, gross, or somewhere between, because the *ex post* estimates must be used in an internally consistent way to adjust the claimed *ex ante* savings. Further, a number of reviewers indicated that rather than overestimating free riders, this approach is likely to underestimate free riders because of selection bias (discussed below in this section).

⁷⁶ ~~Some reviewers indicated that this double counting problem may be the result of inconsistent program rules as set out by the program administrators and regulators, and was not an estimation issue. Further, a number of reviewers indicated that rather than over-estimating freeriders, this approach underestimates freeriders due to selection bias (discussed in the main body text below). The RTF guidelines (dated August 15, 2012) sets out the current practice baseline approach most directly in its definition of savings: “Savings is defined as the difference in energy use between the baseline (see section 2.2) and post (after measure delivery) periods, which is caused by the delivery of a measure. The terms “net” or “gross” are intentionally not used to modify the term “savings,” as they may conflict with the definition of “baseline,” provided in section 2.2. The current practice baseline defines directly the conditions that would prevail in the absence of the program (the counterfactual), as dictated by codes and standards or the current practices of the market. The most important conflict would arise if savings were estimated against a current practice baseline and then those savings were further adjusted by a net to gross ratio, where the net to gross ratio was the probability that the measure would have been delivered in the absence of program influence.” Note that the RTF uses the term *current baseline* rather than *common practice baseline* used elsewhere.~~

⁷⁷ ~~Comments provided by Mr. Tom Eckman of the NW Council’s Regional Technical Forum (RTF) has evolved over since its initial introduction in the 1980s and that it is part of an integrated planning process, including program design, setting deemed savings values for measures, and the production of demand forecasts for integrated planning; and, this process has evolved over a number of years.~~

measure: (1) unit energy savings measures, (2) standard protocol measures, and (3) custom protocol measures. Depending on the measure type, the research design could be relatively straightforward because the RTF might have already established the unit energy savings values. For other measure types that might have used a current practice baseline, the evaluator could determine the baseline based on a vendor's description of what it would normally do for this type of end user, information on recent shipments or sales of relevant equipment, or services gathered from manufacturers, trade associations, distributors, retailers or other studies and databases that establish current practice, or statistical approaches such as regression models involving participants and nonparticipants.

- **Evaluation protocols for NEEA commercial sector initiatives:** At any point in time, consumers are making decisions about equipment purchases, design features, or operational practices. The average efficiency that results from these decisions constitutes an estimate of what would have happened in the absence of NEEA's initiatives. This is the current-practice baseline in the *RTF Guidelines* and represents the counterfactual scenario. The difference between the efficient equipment that NEEA promotes through its initiatives and the counterfactual scenario (which varies by measure) constitutes the savings that NEEA has caused. Any additional adjustments, such as the application of an NTG ratio, are unnecessary (see Ridge et al. 2013). *Indiana's Evaluation Framework*
- **Indiana's evaluation framework:** This framework discusses the use of the standard market practice to estimate net savings: ~~The standard market practice (SMP)~~This approach is a way to set energy impact analysis baselines so that the baseline already incorporates the influence of ~~freeriders~~free riders. In this approach, a ~~freerider~~free rider assessment is not needed because the ~~use of market is already using~~ a standard market practice baseline ~~is already what the market is doing~~ without the program's direct influence. ~~The SMP~~This baseline is typically set at the mean of the level of ~~energy efficiency~~EE being installed across the market being targeted by the program- (TecMarket Works, ~~et al.~~ 2012, p. 55).

Similar excerpts from ~~the NEEA and~~ Delaware guidelines for net energy savings estimation can be found in ~~(Ridge et al., (2013).~~

Gross impact estimation ~~itself~~ is a value that requires a baseline. In other words, the gross savings from an ~~energy efficiency~~EE measure is the difference between the energy use of the installed high-efficiency equipment and an alternative equipment specification. The baseline for the gross impacts estimate may be any of the following: (1) the energy use of the equipment that was replaced during a retrofit; (2) the energy use of standard-efficiency technology that likely would have been installed by the ~~customer~~consumer; or (3) the energy use of the equipment required by codes and standards (assuming stringent enforcement of the codes and standards). In fact, Ridge et al. (2013) point out that the actual equipment baseline used to estimate gross impacts may not be clear cut and that "there are gradations in the way baselines are established in the energy-efficiency industry."

The case for the use of a common practice baseline appears to stem from two issues:

1. The definition of gross savings may ~~actually~~ include factors that are more appropriately viewed as components of net savings, and additional adjustments are not needed to these original estimates. This is essentially an *ex ante* estimate of net savings using current practice as the baseline ~~and with~~ net savings ~~is estimated as~~ the reduction in energy use resulting from the change to more efficient technologies.^{-78,79}
2. Program evaluations that report net savings may do so inconsistently. Unfortunately, the components of the net savings calculation differ between jurisdictions, and ~~those components~~ are often based on what ~~is viewed as appropriate and measureable~~ by the jurisdiction's stakeholders (~~See~~ ~~view as appropriate and measurable~~ (see NEEP, 2012). Although ~~there is wide recognition that~~ spillover ~~exists~~ ~~is widely recognized~~ and can be significant, ~~there is resistance in~~ a number of jurisdictions ~~to resist~~ estimating spillover values and including them in the net savings calculations. Market effects values have faced similar challenges.⁻⁸⁰

SEE Action (2012b, p. 7) indicates that appropriate common practice baselines can be estimated through surveys of participants and nonparticipants as well as analysis of market data. The process of developing a working definition of common practice baselines may pose some challenges. Currently, there is not ~~wide spread~~ ~~widespread~~ experience in developing common practice baselines allowing for a determination of best practices. The RTF of the NW Council has the most experience in developing these baselines, with its methods emphasizing the use of

⁷⁸ ~~Mr.~~ Tom Eckman of the NW Council expands on this point ~~in personal communication~~, stating that: “What is occurring prior to program launch is a better measure of what would have occurred absent the program (that is, the counterfactual scenario) than a determination made after the program has influenced the market.” Essentially, the NW Council ~~performs~~ ~~performed~~ an *ex ante* net analysis when they developed deemed savings estimates that are by design viewed as net savings. For the NW Council's purposes, this is viewed as being as accurate as performing complex studies after the program has been implemented. More information on the NW Council approach can be found ~~at in RTF (2012) and at the RTF website~~ <http://rtf.nwcouncil.org/>.

⁷⁹ ~~The~~ common practice approach as applied by the NW Council works best when the forecasts are made at the measure level. Covering all the measures ~~which that~~ combine to make a program can be time consuming and expensive to update. Also, this is short term in that over time ~~The, the~~ control group (that is, nonparticipants) would likely have evolved their actions from one year to the next as conditions change and ~~that~~ accounting for these effects is ~~also~~ important in determining net savings. As with all approaches discussed in this section, there are pros and cons and the selection of the approach to use has to recognize the context in which this choice is made. For example, ~~there are no~~ ~~Tom Eckman of the NW council indicated that this method may be less controversial in the Northwest because some entities do not have~~ financial incentives tied to ~~estimates of net savings value among members of the NW Council~~.

⁸⁰ ~~To~~ further illustrate, net savings as presented in the findings of EE evaluations ~~is are~~ always presented as “net” of something; however, it may be gross savings net freeridership, or it may be gross savings net freeridership and spillover, or, in some cases, market effects may be included in the defined net savings estimates. Navigant (2013) found that ~~the majority of most~~ jurisdictions defined net savings as “gross savings adjusted only for freeridership.” (The review of net savings methodologies in Navigant ~~(2013)~~ ~~[2013a]~~ focused only on C&I programs. ~~Out of thirty eight~~ ~~Of 38~~ C&I program evaluations reviewed, ~~twenty eight~~ ~~28~~ estimated net savings as gross savings adjusted for freeridership only. Three estimated net savings as gross adjusted for freeridership plus participant spillover, and seven studies adjusted for freeridership and both participant and nonparticipant spillover. None of the studies attempted to address market effects in addition to the spillover values.)

DRAFT #4 | Not for Citation

market data,⁸¹ and the RTF has produced guidelines for the development and maintenance of ~~savings~~ savings estimation methods based on the common practice baseline approach (RTF; 2012).

A significant concern is that self-~~selection~~ selection bias ~~may still be an~~ is viewed as a likely issue with common practice baselines. An EE program that allows consumers to select themselves into the program may attract ~~those~~ consumers among the common practice baseline who would have taken the high-efficiency actions anyway. If an EE program ~~only~~ attracted ~~those only~~ consumers who were predisposed to install the high-efficiency equipment promoted by the program, ~~then~~ net savings could be overestimated by not fully accounting for all ~~freeridership~~ free ridership. Additionally, to the extent that the program results in nonparticipant spillover, it is not clear how the common practice baseline approach would capture those savings.⁸²

Another point made ~~in~~ by Ridge et al. (2013) is that ~~prior~~ previous EE programs have affected the markets for EE equipment through spillover and market effects. This results in current ~~standard~~ common practice baselines that are more efficient than ~~what they~~ would have been ~~the~~ ease if these past EE programs were not offered. In this case, using market average can contain a fair number of past participants (for example, end users, installers, and distributors) and nonparticipants who ~~were already have been~~ influenced by the program. The effect of these past programs is to lower the annual energy use of the measures that constitute the current practice. This argument seems to be partly analytical and partly a policy consideration. Ideally, past evaluations of EE programs should have included all the impacts attributable to the programs, ~~but, since~~ because spillover and market effects were generally omitted from past evaluations, they have not been counted. The annual energy use that is represented by current practice is lower than it would have been if these past programs were not offered. From this perspective, the use of unadjusted current practice baselines as estimates of net savings seems to be an effort to make up for mistakes in past evaluations (that is, the omission of spillover and market effects that impact the overall market).

A jurisdiction may view ~~this savings that accrue today from programs in previous years along with the savings from current programs~~ as a reasonable estimate of EE program impacts over time, which the long term; and, that this best represents the overall return on investment in EE. Alternatively, it may take the position that each EE program should be evaluated as an incremental investment (that is, a program implemented in 2014 should be evaluated against what is attributable to that investment only—all impacts from prior years' programs are essentially sunk costs and should not be considered). This is an example of where policy and analytic views of net savings estimation are linked. ~~It is not possible to definitively recommend a net savings approach across program types and jurisdictions without considering the appropriateness of the decision to include impacts from prior programs on the current practice baseline.~~

⁸¹ _____The RTF of the NW Council believes that the emphasis on market research for developing common practice baselines will also help produce better program designs.

⁸² This will not be an issue in applications where market-wide sales data are available on standard and energy-efficient equipment, but these data are unavailable in most markets targeted by EE programs.

Another factor is that the common practice baseline is essentially a snapshot in time. The common practice baseline will change over time and periodic updates will be needed.⁸³ The complexity of the update will depend on the program type. For essentially a one-technology program (for example, refrigerator recycling), the update may be straightforward. Updating common practice baselines for a large C&I custom program where many technologies and end uses are impacted may be more difficult. In such cases, it might be more cost effective to focus exclusively on measures that account for the greatest savings.

The bottom line for assessing the common practice baseline approach is the same process as that is used in all other methods: (1) understand the construction of the baseline used in the evaluation; and, (2) analyze the implications of this baseline against what is an appropriate counterfactual scenario for that program. Based on this standard approach, decisions can be made regarding about the net savings estimation method that is most appropriate for the evaluation of an EE program.

When an evaluator encounters a jurisdiction that is using a “current practice baseline” method and refers to these savings as net savings, the evaluator should proceed in an internally consistent manner.⁸⁴ For example, it is important that the evaluator explain what the utility/agency/regional body is calling gross savings and what, if any, adjustments have been made in the establishment of the baseline to produce a net savings value.

The common practice baseline has not been advocated as applicable to all programs, even within a single jurisdiction. An evaluator can select from among the many other methods for estimating net savings, each with its own sources of error, and decide which is most likely to produce estimates that have the least error. Hall et al. (2013) state that they “are not suggesting that the direct net analysis approaches (i.e., common practice baselines) should be used in all evaluations or that they can be applied to all types of program configurations or target markets.” As a result, the common practice baseline approach should be considered as another method in the toolkit that evaluators can use to address net savings, based on an analysis of the market and the appropriate counterfactual scenario.

⁸³ This is no different than programs evaluated using more traditional methods. The fundamental question is, “What is the shelf life of any evaluation given that many things (e.g., program intervention strategies, technologies promoted, targeted customers, and local and regional economic conditions) can change that would affect the program’s ability to deliver net savings?” That is, all evaluations are essentially a snapshot in time.

⁸⁴ Reviewers of this section have commented that the evaluator might conduct multiple current baseline studies, calculate *ex post* net savings, and calculate a net realization rate to test the robustness of the approach; however, the cost of the analyses becomes a factor. Analyzing the market and different baselines has been presented as useful for understanding EE programs. This view may be most appropriate for jurisdictions that have EE measure and equipment specific data. These data may be limited to certain types of programs, and require a commitment to gathering data at the measure level. Also, before taking this approach, the evaluator might want to make sure that self-selection, nonparticipant spillover, and market effects are not serious sources of bias. If serious bias is suspected, the evaluator could select the baseline from the multiple baseline approaches above as the one that produces the most conservative results; however, there may be little analytic support for this selection. Another suggestion advanced in this newly developed literature is to augment the results using a survey based self-report NTG ratio, but this seems to defeat the purpose of using the common practice baseline method as an *ex ante* method of producing net savings. It increases costs and brings in the issues involved in using appropriate survey methods, and it may thereby reduce some of the advantages claimed for the common practice baseline approach.

In summary, several jurisdictions ~~have adopted~~ looking toward the use of common practice baselines in their EE evaluation guidelines. As with all methods, there are pros and cons: (see Table 7). A potential strength of the common practice baseline approach is its use in upstream and market transformation EE programs. It can be applied market-wide and, unlike randomized trials and quasi-experimental designs, it does not require participants to be identified. if appropriate sales data are available. However, ~~one of the challenges with~~ this method is controlling for more susceptible to self-selection (that is, the average consumer may not be the type of consumer who participates in the program). It is not clear how this can be addressed, other than by conducting surveys to determine specific characteristics of purchasers of efficient equipment relative to the common practice baseline. However, this survey effort would negate the unique aspects claimed for the common practice baseline approach; i.e., specific consumers who have and have not purchased the high efficiency equipment would need to be identified. This makes this approach more similar to the survey method approaches discussed in Section 3.2.

Another factor to consider is that ~~the common practice baseline is essentially a snapshot in time~~. ~~The common practice baseline will change over time and there will be a need for periodic updates~~. ~~The complexity of the update will depend on the program type~~. If it is essentially a one technology program (for example, refrigerator recycling), then the update may be straight forward. Updating common practice baselines for a large C&I custom program where many technologies and end-uses are impacted may be more difficult. ~~Hall et al. (2013) state that they “are not suggesting that direct net analysis approaches (that is, common practice baselines) should be used in all evaluations or that they can be applied to all types of program configurations or target markets.”~~ As a result, the common practice baseline approach is another technique in the toolkit that evaluators can use to address net savings, based on an analysis of the market and the appropriate counterfactual scenario.

Table 7. Common Practice Baseline Approach—Summary View of Pros and Cons

Pros	<ul style="list-style-type: none"> • Can help to avoid double counting of freeridership <u>free ridership</u> in circumstances where gross impacts incorporate some net savings factors • Can be used in upstream and market transformation programs • Can be applied market-wide
Cons	<ul style="list-style-type: none"> • Self-selection bias is not addressed <u>and methods for addressing self-selection are not readily apparent</u> • Does not capture nonparticipant spillover • Common-practice baselines for measures and technologies will change over time and require updating • <u>Determining average market practice has accuracy challenges</u> • Approach <u>has been</u> applied in the Pacific Northwest, along with other net savings estimation methods, but is relatively new and still evolving as a general net savings estimation method

3.4 Market Sales Data Analyses (Cross-Sectional Studies)

A market sales data method can capture the total net effect of the program, including both ~~freeridership~~ free ridership and participant and nonparticipant “like” spillover. As described in a

residential ~~freeridership~~ free ridership and spillover methodology study prepared for the Massachusetts Program Administrators (NMR ~~et al.~~ Group, Inc. and Tetra Tech 2011), the total net effects of a program can be estimated ~~through~~ via an analysis of market sales data.

The most common approach is a cross-sectional, comparison area method in which post-program data are compared with data from a ~~non-program~~ nonprogram comparison area (or multiple comparison areas) for the same point in time. Thus, evaluators can make a comparison between the change in the program area from the pre-program period to the post-program period *and* the change in the ~~non-program~~ nonprogram area over the same period.

The NMR ~~et al.~~ Group, Inc. and Tetra Tech (2011) study lists three important factors to consider when deciding if an approach is appropriate ~~to use~~ for a particular program:

- **Does an appropriate comparison area(s) exist?** Comparison area(s) must represent a credible baseline for the area of interest. This may entail using a set of systematic adjustments to control for differences in total size of, or demographics for, the areas. As EE programs become more prevalent, ~~it is becoming more difficult to find~~ finding comparison areas that do not have similar program activities ~~is becoming more difficult~~.
- **~~Is~~ Are the market data available and complete?** Market data analysis requires comprehensive market data for ~~both~~ the area of interest and an appropriate comparison area(s). or areas. The complication here is that comprehensive sales/shipment tracking systems have not been available for most markets. Absent comprehensive sales data, a general picture of market coverage can be obtained by conducting surveys or in-depth interviews. These are typically conducted with vendors and contractors about sales volumes and efficient equipment sales shares for conditions with and without the program, or for in-territory and comparison area sales. In some cases, the self-reported purchases of participating end-~~users~~ users can provide market data if the sample is sufficiently large and representative of the market. Also, it can be expensive to gather the market sales and shipment data, and even ~~with~~ a diligent data collection effort, ~~there~~ may be leave gaps in the data.
- **What are the features of the program?** Market data analysis is usually appropriate for programs that promote large numbers of homogenous measures and that have substantial influence upstream to the end-user.

As an example of this approach, Cadmus et al. (2012) tracked ENERGY STAR[®] appliances, lighting, and home electronics product sales in New York and then compared those sales to sales of the same products in Washington, D.C., Houston, Texas, and Ohio. All ~~of~~ these baseline areas were areas without significant utility efforts to promote ENERGY STAR products. The market data were used to estimate both the market share and the energy savings attributable to the New York Energy ~~\$mart~~ SM mart Products Initiative Program administered by the New York State Energy ~~and~~ Research and Development Authority.⁸⁵

⁸⁵ ~~Discussion with Mr.~~ Scott Dimetrosky indicated that this study developed savings from product sales and installations. These savings were derived by first estimating the market share for ENERGY STAR products

Another example of a market sales approach entails interviewing or surveying a panel of trade allies who are either program participants or nonparticipants. This could include ~~contactors~~ contractors, retailers, builders, and installers. These trade allies are offered monetary compensation for information ~~on~~ about projects or sales completed within a specified time period. (see Table 8 for pros and cons of this approach). The types of information requested can include manufacturer, efficiency levels, size, price, installation date, installation ZIP code, types of incentives received, and an assessment of the program’s impact on ~~incented and non-~~ incented/nonincented efficiency actions. With annual updates, this method could provide context for tracking longer term ongoing program impacts or market effects. This method could also work in tandem with other approaches for estimating net savings and provide a market context for estimates that may otherwise focus only on short-term impacts.

Table 8: Market Sales Data Analyses—Summary View of Pros and Cons

Pros	<ul style="list-style-type: none"> • Can estimate the total net effect of a program • Uses information on actual consumer behavior • Addresses trends in an entire market • Most appropriate for programs that promote a large numbers of homogeneous measures and have substantial influence upstream
Cons	<ul style="list-style-type: none"> • There may be a low availability and quality of sales and shipment data in the area of interest and in an appropriate comparison area(s) • Data may be expensive to acquire and/or may have gaps that can be misleading • May be difficult to determine the appropriateness of a comparison area

3.5 Top-Down Evaluations (Macroeconomic Macroconsumption Models)

Top-down evaluations use macrodata on energy consumption in a model that relates changes in energy consumption to a measure of EE effort (usually expressed as EE expenditures ~~on EE~~). Top-down evaluation produced ~~what has been termed as~~ “macroconsumption metrics” (MCMs) in two recent pilot applications in California (~~See~~ Cadmus, 2012 2012a; Demand Research, LLC; 2012). The broader literature refers to these ~~methods~~ as top-down methods, ~~but~~ and the MCM notation adopted in the recent California pilot studies refers to the same set of methods and cites top-down studies as background for its pilot work.

The top-down approach has much appeal since it directly addresses To date, this method’s application has been somewhat limited to national or large regional (i.e., multistate) applications. Applications to utility level programs have been limited to pilot studies and the general applicability of these methods has not been demonstrated. Still, the top-down approaches have

through estimates of total market size and sales of ENERGY STAR products. Next, portions of the market share were allocated to exogenous, non-NYESP ~~Program~~ effects, including the impact of the national U.S. Environmental Protection Agency/U.S. Department of Energy ENERGY STAR Program, naturally occurring adoption (including the impact of higher energy prices and interest generated by programs in neighboring states), and the impacts of other NYSERDA New York State Energy Research and Development Authority residential programs. The remaining market share, after netting out these other effects, was considered attributable to the NYESP New York Energy Smart Products Initiative Program.

appeal because they directly address overall net savings. The dependent variable is overall energy use (often expressed as energy use per capita) and this method simply examines the change in energy use ~~due to~~ resulting from EE efforts. ~~As a result~~ Thus, there is no need to adjust for ~~free ridership~~ free ridership and spillover, or even for market effects, in estimating overall net savings. In addition, the regression analyses provide confidence and precision levels around these estimates.

~~There~~ However, there are challenges in estimating the relationship between EE efforts and changes in overall energy consumption, such as the size of the impact isolated by the model.

Developing

The development of a model that can measure a ~~1 to~~ 1-2% change in total energy use annually and is attributable to EE programs requires a reasonably sophisticated structure. For example, the model must have an appropriate lag structure because the impacts from one year's expenditures will occur over a number of years.⁸⁶ In addition, the number of observations and quality of data needed to identify a small effect can be challenging. The data platform needed to support this top-down or MCM model approach requires the following:

- A measure of EE expenditures (or another metric of EE effort for different cross-sections, such as utilities or program administrators).
- The A large number of observations ~~needed~~ to identify the ~~effect~~ effects of EE over a number of years, taking into account the lag structure of EE impacts. As a result, most top-down studies include multiutility or multistate efforts that can provide a reasonably large number of cross-sectional areas for the analyses.
- Matching demographic and macroeconomic data to utility service areas, or subareas of utilities that are used as observations in the analyses.
 - ~~High-quality data regarding~~ about energy consumption for each cross-section analyzed.
 - ~~As a result, most top-down studies include multi-utility or multistate efforts that can provide a reasonably large number of cross-sectional areas for the analyses. The California pilot study used top-down methods to estimate overall EE impacts for the state.~~

Questions that evaluators should consider when deciding on the appropriateness or applicability of top-down models are:

⁸⁶ ~~BC Hydro (2012) demonstrates the importance of the relationship between current expenditures on EE and future savings. It also shows the importance of letting the data determine the most appropriate lag structure as opposed to implementing a fixed structure that acts as a constraint. How~~ The estimate of energy savings is influenced by the manner in which lagged effects are handled in the regression model ~~influences the estimated energy savings.~~

DRAFT #4 | Not for Citation

- What information will be produced by these top-down models if they are successfully estimated, recognizing that a large number of cross-sections with varying levels of EE investment are needed for estimation?
- How does this information compare to what is produced by other methods?

~~For example, top~~ Top-down models may be useful for:

- Estimating overall average change in energy ~~usage due to~~ use from the EE programs for a region. A top-down model that provides a good fit, meets reasonable assumptions, and has acceptable levels of statistical significance levels can provide information on the average change in overall energy use (or energy use per capita) from overall EE efforts.
- Estimating regional environmental impacts. Aggregate models can be useful in assessing state and regional environmental impacts such as the impact on carbon emissions.
- Providing evidence of estimated energy-savings at a regional level. The model can confirm—at an aggregate level—whether the expected energy savings are actually reflected in the ~~macro consumption~~ macroconsumption data.
- Estimating overall cost savings ~~due to~~ from EE programs. Top-down models can also be used to estimate an overall cost savings per ~~kWh~~ kilowatt-hour saved and confirm the efficacy of the overall EE effort.

Top-down models, however, ~~are not able to~~ cannot provide information ~~on~~ about:

- Savings produced by specific measures or programs.
- Where to make additional investments in EE at the program- or measure- level.
- How to improve existing programs.
- How to use estimates of ~~free ridership~~ free ridership and spillover to suggest program improvements.
- Quality assurance/quality control processes needed for regulatory oversight.

The relative importance jurisdictions and stakeholders place on program-level ~~information~~ versus aggregated information ~~at~~ will influence decisions to implement these ~~different~~ types of evaluation frameworks. Top-down approaches seem complementary to results produced by program-level evaluations; however, there may be concerns about using these ~~top-down~~ methods ~~as a replacement for~~ to replace program-level evaluations. Some view the program-level research as essential in that it helps ensure that the right set of programs comprise the EE portfolio and it is useful in addressing program- and portfolio-specific questions ~~regarding~~ about implementation. Top-down methods and program-level evaluation ~~both~~ provide useful, but different, perspectives on the accomplishments of EE efforts.

Cadmus (~~2012~~ 2012a) reviewed a number of ~~the leading~~ top-down studies that ~~all~~ expressed energy consumption as a function of a metric meant to measure EE effort including:

- ~~Parmark~~ Parfomak and Lave (1996) used a panel ~~data set~~ dataset of 39 utilities from 1970 to 1993. The claimed savings by utilities for their C&I programs was used as a proxy for

the level of EE effort. The regression analysis was similar to a realization rate regression analysis model, where the coefficient on the claimed utility savings indicated what fraction of those savings ~~were able to~~ could be found in the data. The ~~study~~ authors estimated the realization rate for the utility's claimed savings at 99%.

- Auffhammer et al. (2008)—working with data developed by Loughran ~~et al. and Kulick (2004)~~—used what has become the more traditional formulation. Here, EE effort was expressed in the econometric model as program expenditures reported to the U.S. Energy Information Administration ~~(EIA)~~. The ~~study~~ authors found that average utility reported savings (2% ~~to~~ %–3%) fell within the 95% confidence interval for estimated savings. The cost of saved energy was approximately \$0.06 ~~per~~/kWh.
- Arimura et al. (2011) also used the ~~EIA~~ Energy Information Administration data on program expenditures across 307 U.S. utilities to examine the ~~impact~~ impacts of EE investments ~~in EE~~ on overall energy consumption.⁸⁷ ~~Using~~ The authors used utility ~~EIA~~ Energy Information Administration data from 1989 to 2006, ~~the study authors found to determine~~ electricity savings of 1.8% annually and estimated the cost of saved energy at approximately \$0.05 ~~per~~/kWh.

The California Pilot Project on top-down methods involved two efforts, Cadmus, ~~(2012)~~ (2012a) and Demand Research, LLC (2012).

Example 1: ~~The Cadmus Group~~ California Top-Down Pilot Study

Cadmus used expenditures on EE programs as the level of EE effort in its models. The models were estimated at the utility level for residential and nonresidential energy savings. Cadmus worked with data at the utility level using information from the three investor-owned utilities (IOUs) and from large public utilities in California such as Los Angeles Department of Water and Power ~~(LADWP)~~ and the Sacramento Municipal Utility District ~~(SMUD)~~. Data were also collected from some small public utilities, but ~~this information was~~ were generally inconsistent.

A number of ~~different~~ models estimated the relationship between utility energy consumption for residential and nonresidential customer segments and EE expenditures ~~on EE~~.⁸⁸ Overall, it was difficult to obtain significant results across the models. The best model produced significant coefficients on the EE expenditures variable using only data from the three IOUs. To demonstrate the information that can be produced by top-down models, Cadmus developed estimates of savings from EE efforts over a ~~six~~6-year period and calculated the cost of energy saved. Savings from EE spending from 2005 to 2010 were estimated at 8%, and the cost per

⁸⁷ —Arimura et al. (2011) also advance the state of the practice by modeling energy prices and utility ~~energy-efficiency~~ EE program expenditures as endogenous and allowing consumption to depend on program expenditures in a flexible way. The literature on top-down models represents sophisticated applications of econometric methods. Problems of ~~endogeneity~~ endogeneity and autocorrelation with flexible lag structures have become common issues that are addressed by these models.

⁸⁸ —Cadmus ~~(2012)~~ (2012a) did not try to estimate separate models for commercial and industrial consumers ~~as~~ because the time-series was inconsistent. In some years, commercial sector consumption would increase and industrial consumption would decrease by approximately the same amount. This suggested that there was some switching in the definition on the commercial and industrial rate classes. As a result, the two classes were modeled together.

~~kWh~~ kilowatt-hour saved was estimated at \$0.05. The results of the Cadmus study indicated savings were within 10% of the net savings reported by California IOUs for the 2006 to 2008 program cycle. The estimates of ~~both~~ energy savings and cost per ~~kWh~~ kilowatt-hour saved had large confidence intervals: $\pm 66\%$ on the energy savings estimate and ~~over more than~~ $\pm 100\%$ on cost per ~~kWh~~ kilowatt-hour saved. The ~~number of 48~~ observations (~~48 total observations~~) in the top-down IOU model resulted in lower precision than studies with much larger sample sizes.

Cadmus did look into disaggregating the data beyond the IOU level to gain more cross-sections for the analysis; however, there was concern about the ability to allocate EE program expenditures to smaller geographic areas. One specific concern was the savings from compact fluorescent lamps (CFLs). ~~Over~~ More than 50% of the expected savings were from CFLs and these sales were tracked at point of sale instead of the location where they were used, making it difficult to align the energy consumption and the impact of EE expenditures for smaller geographic areas.

Example 2: Demand Research, LLC California Top-Down Pilot Study

Demand Research (2012) developed an MCM model working with California utilities and program contractors that disaggregated residential energy use and estimates of residential sector EE efforts into a database of cross-sectional observations at the census tract level. ~~Commercial and industrial~~ C&I sector energy use and metrics for EE efforts were disaggregated down to the county level. Instead of using energy expenditures, the Demand Research, LLC study used the utilities' *ex ante* estimates of energy saved by census tract as the metric of residential EE effort.⁸⁹ ~~Parfomak and Lave (1996) used a similar approach.~~ For the ~~commercial and industrial~~ C&I sectors, county-level data were developed. The independent variable for the EE level of effort in the commercial sector model was a metric related to incentives paid; however, *ex ante* energy savings was used as the metric for EE effort by county for the industrial sector.^{90, 91}

The findings from the Demand Research, LLC study were:

⁸⁹ ~~This approach is similar to that used by Parfomak and Lave (1996).~~

⁹⁰ ~~The reason why different~~ Different metrics for EE level of effort were used in the ~~commercial and industrial~~ C&I sector model ~~was due to~~ because the method selected to address ~~endogeneity~~ endogeneity in the commercial sector model, ~~that is, ensuring~~ ensured that the EE level of effort variables uncorrelated with the error term.

⁹¹ ~~Considerable work went into creating these sector~~ the census tract databases ~~for the residential model and the county level databases used in the commercial and industrial models.~~ The details can be found in the full study, but as an overview of the effort -- key energy consumption and program tracking data by fuel and segment were inspected prior to modeling for missing values, seemingly erroneous data or outliers, and high and low end values that might skew the sample statistics or suggest ~~multi-modal~~ multimodal distributions. Other adjustments to the ~~data sets~~ datasets were made, including the use of a "restricted" commercial sector ~~data set~~ dataset that included only counties with high *ex ante* energy savings values in this pilot test. Dropping sites from statistical analyses that likely provide no information because the expected savings from those sites are so small is not uncommon. The usual justification is that the total savings number is not likely to be influenced by their exclusion ~~since~~ because the expected savings were so small.

DRAFT #4 | Not for Citation

- The residential models estimated by Demand Research, LLC (2012) showed that higher levels of the EE effort variable resulted in reduced energy use with ~~estimates statistically significant~~ estimates at a 95% confidence interval.
- The commercial sector model produced the expected sign on the EE effort variable, but the results were not statistically significant.
- The industrial sector model did produce statistically significant results for the EE effort variable.
- The residential, ~~commercial~~, and industrial C&I sector models produced statewide savings estimates of 7.3% for the five 5-year period from 2006 to 2010.
- The relative precision for the aggregate savings estimate was $\pm 31\%$ (or a 90% confidence interval of ~~5.0% to~~ -9.5%).
- The estimated statewide savings of 7.3% exceeded the utility *ex ante* estimates of 4.8%.

The aggregate statewide estimate of energy savings across all three sectors was forecasted with reasonable confidence and precision. Looking at the results at one level of disaggregation lower (at the sector level results) shows a high degree of variability. For example:

- The estimated industrial energy savings (all three utilities combined) were ~~much about~~ 745% higher than the utilities' *ex ante* values, ~~about 745% higher~~ (Demand Research, LLC; 2012, p. 36).
- The commercial sector ~~kWh~~ kilowatt-hour savings estimates (all three IOUs combined) were ~~much about~~ 27% lower than the utilities' *ex ante* estimates ~~(about 27% of the ex ante savings)~~.
- The residential sector savings estimates from the ~~estimated~~ MCM model for PG&E Pacific Gas & Electric and SDG&E (SCESan Diego Gas & Electric (Southern California Edison was not estimated) were substantially higher than the utilities' *ex ante* values.

When these sector-level results are aggregated up to a statewide number, the wide discrepancies at the sector level tend to offset each other. It is important to recognize that this was a pilot effort and views will differ ~~on~~ about the overall robustness of findings at the sector and statewide levels.

3.5.1 Developing Top-Down Models

Cadmus (~~2012~~ 2012a) and Demand Research, LLC (2012) took different paths to developing a top-down MCM model for this California Pilot Study. Both study teams concluded that the work to date indicated this was a potentially useful research path for developing statewide estimates of energy savings attributable to EE policies. In its study report, Cadmus discussed the potential applications of these methods:

DRAFT #4 | Not for Citation

- Top-down ~~macro-consumption~~macroconsumption methods could yield inexpensive⁹² estimates of energy savings from utility ~~energy efficiency~~EE programs and building codes at an aggregate level.
- These methods are attractive because it is possible to produce confidence and precision levels for the net energy savings estimates, ~~something that~~which is not as easily accomplished in bottom-up evaluation studies.⁹³
- Top-down studies can be used to verify statewide ~~energy efficiency~~EE program savings estimates based on bottom-up evaluation by looking at aggregate energy consumption data.
- These methods can be useful in tracking a state’s progress in reducing greenhouse gas emissions and developing forecasts of energy savings from future program spending at an aggregate level.

Next steps that might provide additional insights into this top-down application –are to: (1) replicate the results of Cadmus and Demand Research, LLC using the datasets already developed;⁹⁴ and (2) continue improving the data platform⁹⁴ used for these analyses—both studies contained recommendations for improving the data. Violette et al. (2012) discuss the importance of the data platform on which these top-down models are estimated. Other considerations pertain to the sensitivity of the results to model specification (that is, the robustness of the results under a designed set of alternative specifications that are also consistent with the theory and appropriate econometric methods).⁹⁵

It seems unlikely that bottom-up studies would be entirely replaced by these top-down methods- (see Table 9 for pros and cons of these methods). As discussed earlier, there is likely a need to have program-level (and some measure-level) assessments to ensure that a program’s design will result in a program meeting its specified targets. As a result, evaluatorsEvaluators should ask, “Does the incremental value of the information produced by the top-down methods exceed the cost of the work?” At the national level, data from an adequate number of cross-sectional

⁹² ———Both pilot studies ran into data problems that would have to be overcome in future work and ~~there would~~could be a decent price tag associated with this workcostly to address. If the alternative were to build up statewide estimates by doing measure-specific engineering analyses, this aggregate Top-Down approach ~~would~~might be cheaperless expensive; however, bottom-up methods performed cost-effectively are probably needed for program support, design, and verification of savings at the program level. The issue is whether the incremental information provided by these aggregate studies has a value greater than its cost. That may vary by jurisdiction.

⁹³ This is a conclusion from the Cadmus (2012a) top-down applications; however, bottom-up approaches also routinely calculate confidence and precision levels for program and portfolio estimates of net savings. The advantage with the top-down approach might be that the confidence and precision levels can be calculated more easily at the aggregate level, because different values for confidence and precision across programs do not have to be combined using assumptions about the covariance across the different distributions from which these values are calculated for each program.

⁹⁴ ———Violette et al. (2012) discusses the importance of the data platform on which these top-down models are estimated.

⁹⁵ ———This sensitivity analysis might examine the stability of the estimates under alternative functional forms, inclusion of one or two variables, testing of interaction terms, and tests on subsets of the data.

observations ~~is~~are more easily available. For state-level studies, more work will be involved in setting up the databases and disaggregating the data into the number of needed cross-sections, which may introduce ~~a certain amount of~~some error into these observations.⁹⁶

Table 9. Top-Down Evaluations (Macroeconomic Models)—Summary View of Pros and Cons

Pros	<ul style="list-style-type: none"> Estimates net effects of all programs cumulatively No need to adjust for free ridershipfree ridership, spillover, or market effects at the aggregate level
Cons	<ul style="list-style-type: none"> Methods are not fully developed at the state or regional levels Relies on high-quality energy consumption data and on data regarding EE efforts within each cross-section analyzed Cannot provide savings at the measure, technology, or program levelslevel Does not provide information on how to improve program design and implementation processes

3.6 Structured Expert Judgment Approaches

Structured expert judgment approaches involve assembling a panel of experts who have a good working knowledge of the technology, infrastructure systems, markets, and political environments. This approach is one alternative for addressing market effects in different end-use markets. These experts are asked to estimate baseline market share for a measure or behavior. In some cases, they are also asked to forecast market share with and without the program in place. Structured expert judgment processes use a variety of specific techniques to ensure that the panel of experts specify and take into account key known facts about the program, the technologies supported, and the development of other influences over time (Tetra Tech et al., 2011).

The Delphi process is the most widely known technique (NMR ~~et al.~~Group, Inc. and Research Into Action 2010). ~~Using this process, each~~Each panelist is asked to make a judgment on the topic—based on the provided information and on ~~their~~his or her experience—and submit the information ~~back~~to the evaluators. The evaluators compile the information from the panelists and ~~resend~~return it to the panelists for another review. The panelists are asked whether they stand by their original judgments or whether the assessments of their peers have caused them to alter their judgments. At least two rounds of judgment are required for a Delphi panel, although more rounds can be used.

~~Some of the~~ advantages of the structured expert judgment approach are:

- The estimate is based on feedback from a group of experts, which can be particularly useful for programs with complex end-uses.
- It is a useful tool for consolidating results from multiple methods to develop a consensus estimate (see ~~Example~~example 2 below).

⁹⁶ ———Violette ~~and Provencher~~ (2012) ~~in Chapter 13 of these DOE Uniform Method Protocols discusses~~discuss attenuation bias where the coefficients on independent variable can be biased ~~toward~~toward zero due to errors in the measurement of variables. A similar effect is shown in Ridge (1997).

As with other approaches (such as market sales data analysis), the structured expert judgment method relies on high-quality data to inform the panel, so ~~a lack of these~~ sparse data can result in inaccurate estimates of net savings (NMR ~~et al.~~, Group, Inc. and Research Into Action 2010).

Two examples of using the structured expert judgment approach to estimate net savings are presented here. The first example describes how Delphi panels were used to estimate net savings for a residential new construction program in California. The second example describes the development a final estimate through the use of a Delphi panel's review of estimates.⁹⁷

Example 1: Residential New Construction Delphi Panel

~~A report~~In a study prepared for the California Public Utilities Commission Energy Division ~~describes in detail how~~, evaluators used two Delphi panels of Title 24 consultants and building industry experts to convert the gross savings estimates. The panel converted estimates from ~~investor-owned utility~~ IOU programs targeting the residential new construction sector to net savings estimates (NMR Hoefgen et al., 2011).

The panelists received detailed data pertaining to code compliance, compliance margins, and estimates of annual gross energy savings in ~~non-program~~ nonprogram homes at the state level and by climate region. After reviewing these data, panelists were asked to:

- Estimate the proportion of the electricity and natural gas savings attributable to the IOU programs targeting the residential new construction sector and other factors (non-IOU ~~RN~~ residential new construction programs, the economy/housing market, energy prices, and climate change).
- Estimate the percentage of net savings in ~~non-program~~ nonprogram homes attributable to different IOU program elements (builder trainings, incentives, and design assistance).
- Assess the extent to which the market effects were likely to persist in the absence or reduction of the IOU programs.
- Estimate the percentage of homes that would have been below-code in the absence of the IOUs' programs and other factors, and estimate the compliance margin of the below-code homes in the absence of each factor.

Each panelist completed two rounds of detailed surveys. In the second round, they were provided ~~with~~ a comparison with other panelists' responses and logic and allowed to change their answers. The evaluation team analyzed the Title 24 consultant responses (both weighted and unweighted) using the building industry experts' responses as a qualitative check. The Delphi panel provided estimates on gross electricity and gross natural gas savings ~~due to~~ from above-code homes. Both

⁹⁷ ———An application of the Delphi technique as applied outside of EE may be informative. Navigant (20132013b) conducted an evaluation of the Wind Power America program. The goal was to assess the impacts attributable to the program. The unique aspect of this Delphi exercise was ~~that~~ the use of range estimates; that is, experts were asked about lower and upper bounds to the effects as well as a best estimate. This approach allowed ~~for~~ the experts to provide their own insights into the uncertainty of the estimates. Gauging uncertainty and then using that in probabilistic and scenario analyses ~~is~~ are consistent with other utility resource planning activities. Adapting these methods to EE resource assessment may increase the usefulness of the information.

panels identified the various elements of training (builders, subcontractors, and Title 24 and code officials) as the most important elements of the IOUs’ programs.

Example 2: Lighting Program Delphi Panel

Another way to use a Delphi panel is to have the panel review estimates derived through other methods to develop a final estimate. As part of the evaluation of the Massachusetts ENERGY STAR Lighting Program (KEMA [et al., 2010](#)), evaluators used a Delphi panel of lighting and EE experts across the United States and Canada. The panelists were asked to integrate results from five methodologies that yielded NTG estimates (conjoint analysis, multistate modeling, revealed preference study, supplier interviews, and a willingness-to-pay study). Evaluators then used the Delphi panel’s review ~~results~~ in developing recommendations for the final NTG estimate. ([See Table 10 for pros and cons of this approach.](#))

Table 10: Structured Expert Judgment Approaches—Summary View of Pros and Cons

Pros	<ul style="list-style-type: none"> • The resulting estimate is the independent, professional judgment of a group of technology and/or market experts • It is a useful approach for programs with diverse and complex end-uses or practices • Is a useful tool for consolidating results from multiple methods to develop a consensus estimate • Panel members can provide levels of confidence and procedures using appropriate elicitation procedures methods
Cons	<ul style="list-style-type: none"> • The approach relies on high-quality data to inform the panel, leading to reasonable estimates of net savings • Sampling-based calculations of confidence and precision are not available

3.7 Deemed or Stipulated NTG Net-to-Gross Ratios

Deemed or stipulated NTG ratios are predetermined values and do not rely on a calculation-based approach. Deemed values are often based on previous NTG research that was conducted using at least one of the other methods described in this chapter.

NTG ratios are often stipulated when the expense of conducting NTG ratio analyses cannot be justified or when the uncertainty of the potential results is too great to warrant a study. A recent review of 42 jurisdictions in the United States and Canada (which represented ~~the vast majority of nearly all~~ jurisdictions with ratepayer-funded ~~energy efficiency~~EE programs) found that only 14% use a deemed approach to NTG [for C&I programs compared to 50% of the jurisdictions using an active research approach to developing estimates of net savings factors](#) (Navigant, [2013](#)).⁹⁸

⁹⁸ [Approximately one third of the jurisdictions did not adjust gross savings for either free ridership or spillover; however, many of those states conducted some NTG research to inform future program design. This reflects policy decisions in each state. Several states that did not adjust gross savings for net savings factors at the time of this study have changed or are contemplating changing to approaches that do estimate net savings. Pennsylvania and Maryland fall into this category.](#)

DRAFT #4 | Not for Citation

Deemed or stipulated NTG ratios are typically either set by a regulatory agency or negotiated between regulators and program administrators. These ratios may be determined at the portfolio level (for example, Michigan and Arkansas)⁹⁹ or on a measure-by-measure basis (for example, California and Vermont).¹⁰⁰ Typically, evaluators base the ratios on NTG studies from past evaluations and/or reviews of other similar programs in which ~~aan~~ NTG ratio was estimated. For example, it is not unusual in a multiyear portfolio cycle to estimate ~~aan~~ NTG ratio for an initial year (or possibly every other year), with deemed values used in the subsequent or intervening years. This multiyear estimation of NTG ratios is a compromise between performing net savings estimation studies every year and the use of deemed values based on that research for a selected time period. Massachusetts has recently moved to this approach.¹⁰¹

In other cases, evaluators use historical data or other information from a wide range of sources to develop a “weight of evidence” conclusion ~~regarding about~~ the program’s influence (SEE Action; 2012b). As discussed earlier, one common approach for developing a stipulated value is to use a panel of experts who have the relevant experience to make that judgment (Delphi panel).

~~While Although~~ using deemed or stipulated values is a relatively simple and low-cost approach, there are ~~a number of several~~ disadvantages. NTG values are variable across time and space, and strongly linked to program design ~~and implementation making~~. This makes deemed values or assumptions potentially unreliable when transferred from a program in one jurisdiction to a similar program in another jurisdiction.¹⁰² NTG values based on primary research efforts can produce estimates that are based on program-specific information (NMR ~~et al., Group, Inc. and Research Into Action~~ 2010). As a result, these values provide useful information for the ~~future program~~ design and implementation of programs¹⁰³ and may mitigate the risk to ratepayers from utilities receiving performance incentive payments on savings not actually attributable to the program (as well as the risk to ratepayers of making performance incentive payments that are too large). NTG values are also critical from a resource planning perspective and having better data on the actual energy savings achieved from energy efficiency programs can help the planning

⁹⁹ ———Arkansas: NTG deemed at 0.8—, http://www.apscservices.info/pdf/07/07-085-tf_286_44.pdf; Michigan—: NTG is deemed at 0.9 for all programs except pilot, education, and low-income programs, which are deemed at 1.0. <http://efile.mpsc.state.mi.us/efile/docs/17138/0009.pdf>. Note that most low-income programs are not subject to NTG analysis (that is, are deemed at 1.0).

¹⁰⁰ ———California—, <http://www.energy.ca.gov/deer/>; Vermont— http://www.encyvermont.com/docs/about_efficiency_vermont/annual_reports/2011_Gross_to_Net_Report_EfficiencyVermont.pdf; Vermont, see: www.encyvermont.com/docs/about_efficiency_vermont/annual_reports/2011_Gross_to_Net_Report_EfficiencyVermont.pdf

¹⁰¹ Massachusetts has been conducting extensive NTG research, but has moved to deemed/stipulated values for the next 3-year plan. Any NTG variances from the stipulated values have no effect on current cost recovery or incentive payments. Yet the extensive program- and measure-level NTG research continues where appropriate, and the state is benefiting from improved program designs without major controversy involving cost recovery and incentives for current programs.

¹⁰² ———Another issue raised by a reviewer was that the use of deemed NTG values can remove the incentive for the program administrator to reduce ~~freeridership~~ free ridership and maximize spillover and market effects to yield greater net savings values.

¹⁰³ ———For example, ~~freeridership~~ free ridership can inform decisions to discontinue incenting certain measures, increase incentive amounts, or increase the efficiency level being incented.

process (Navigant ~~Consulting, 2013~~2013a). Deemed or stipulated NTG values do not provide these benefits.

The following example illustrates how one agency uses deemed savings for program planning.

Example 1: California Public Utilities Commission ~~DEER database~~ Database for Energy Efficient Resources

The California Public Utilities Commission uses deemed savings (listed in its Database for Energy Efficient Resources) for planning purposes and interim savings estimates for its programs. These deemed savings are updated based on results of NTG studies. NTG savings values are presented for ~~kWh and kW~~ kilowatt-hours and kilowatts. (See Table 11 for pros and cons of this approach.)

Table 11: Deemed or Stipulated Approaches—Summary View of Pros and Cons

Pros	<ul style="list-style-type: none"> • This approach can reduce contentious after-implementation adjustments to estimated program savings because agreed-upon net savings factors are developed in advance of program implementation
Cons	<ul style="list-style-type: none"> • An incorrect estimate can be deemed • It is not based on program-specific information • The evaluator cannot assign sample-based statistical precision to the estimate • Developing deemed savings net values at the measure and technology levels can be time consuming and expensive • The process for developing deemed net savings can be contentious

3.8 Historical Tracing (or Case Study) Method

This method involves reconstructing the events (such as the launch of a product or the passage of legislation) that led to the outcome of interest. An example of this is developing a “weight of evidence” conclusion ~~regarding about~~ the specific influence a program had on the outcome.

Historical tracing relies on logical devices typically found in historical studies, journalism, and legal arguments (Rosenberg ~~et al.,~~ and Hoefgen 2009). These include:

- Compiling, comparing, and weighing the merits of narratives of the same set of events provided by individuals who have different points of view and interests in the outcome;
- Compiling detailed chronological narratives of the events in question to validate hypotheses regarding patterns of influence;
- Positing a number of alternative causal hypotheses and examining their consistency with the narrative fact pattern;
- Assessing the consistency of the observed fact pattern with linkages predicted by the program logic model; ~~and~~
- Using information from a wide range of sources (including public and private documents, personal interviews, and surveys) to inform historical tracing analyses.

The historical tracing method traces chronologically a series of interrelated events either going forward from the research point of interest to downstream outcomes, or working backward from an outcome along a path that is expected to lead to precursor events. If all likely paths are followed, forward tracing can capture a relatively comprehensive view of project or program effects. Because the path leads from a program event, the connection to the event is assured. Backward tracing usually focuses on a single outcome of importance and follows the trail back through developments that seem to have been critical to reaching the identified outcome. These developments may or may not link back to the research program of interest (see Ruegg and Jordan 2007).

Weiss (1997) suggests historical tracing is similar to theory-driven evaluation and can be viewed as an alternative to classical experimental design. This approach suggests that if the predicted steps between an activity and an outcome can be confirmed in implementation, this matching of the theory to the observed outcomes will lend a strong argument for causality. In other words, if the evaluation can show a series of microsteps that lead from inputs to outcomes, causal attribution, for all practical purposes, is supported by this approach.

Scriven (2009) argues that some researchers have been entranced by the paragon of experimental design—the RCT—and have generalized this into a virtual standard for good causal investigation. This view can be contrasted to the way that “epidemiology, engineering, geology, field biology, and many other sciences establish causal conclusions to the highest standards of scientific (and legal) credibility” (p. 151).

This method is best suited to an attribution analysis of major events, such as adoption of new building codes or policies. It is not typically applicable to EE programs. However, various elements of this approach may be used in the analysis of very large custom projects that essentially require case study approaches.

WhileBecause this method draws from multiple information sources, it is difficult or impossible to determine the magnitude of the effects, so the evaluator cannot assign statistical precision to the estimate (NMR ~~et al.~~, Group, Inc. and Research Into Action 2010). However, as part of making a persuasive case for attribution and providing evidence supporting a statistically derived net savings estimate, this method can be very important. Statistics alone often ~~are~~do not constitute a complete attribution assessment. They often require context using supporting logic to enhance the validity of the statistical estimates, as illustrated in the following example.

Example 1. Historical Tracing for a Residential New Construction Program

Keneipp et al., (2011) used historical tracing in conjunction with Delphi panels to develop energy savings for new homes. (see Table 12 for pros and cons of this approach). This study used historical tracing spanning 14 years of regulatory documents to create timelines of the residential new construction program presence and activities for Arizona Public Service Company. ~~Using these data, the~~The evaluators ~~created~~used these data to create an influence diagram of market influences on specific building practices. This information was then shared with two in-person Delphi panels of market experts who estimated the percentage of homes built in 2010 using specific building practices. These Delphi panels also developed the counterfactual scenarios used to show the net impact of the residential program on the percentage of homes that were built to standards, but would not have met these standards in the absence of the program. The Delphi outputs were then used to develop inputs for an engineering simulation model to

calculate energy savings per home. [This example illustrates how historical tracing can be used in combination with other methods to develop actual quantitative net savings estimates from an EE program.](#)

Table 12. Historical Tracing (or Case Study) Method—Summary View of Pros and Cons

Pros	Draws from multiple information sources Can be used at a market level for upstream EE programs Can be useful for making a persuasive case for attribution and provide evidence to support a statistically derived net savings estimate
Cons	It can be difficult to translate the influence factors into estimates of impacts without additional modeling The evaluator cannot calculate sample-based statistical confidence and precision levels for the estimate

DRAFT: NOT FOR CITATION

4 Conclusions and Recommendations

A central theme in this chapter is that all decisions have an implicit counterfactual scenario—what would have happened if the decision had not been made. In the context of EE program investments, net savings are ~~the savings~~those that are attributable to the program. In other words, they would not have occurred if the program had not been offered. This chapter presents a number of approaches for ~~addressing~~assessing attribution and the net impacts ~~resulting from~~of EE programs. ~~The~~This section discusses issues affecting the choice of a net savings approach within an evaluation context.

4.1 A Layered Evaluation Approach

It is important that the selected approach be appropriate for the intended audience and ~~that it presents~~present analyses supported by evidence. A well-executed statistical analysis ~~will~~may be ~~a central piece of the evaluation, but it still may~~ not be persuasive to many decision-makers and stakeholders on its own. All approaches should be supported by a narrative discussing why a specific approach was taken, the appropriate interpretation of the findings, and the context for identifying net savings- (see historical tracing above). The narrative and analysis should also recognize and indicate the uncertainty in net savings determination. Developing an appropriate narrative often leads to the application of layered methods of analyses.

Studies examining net savings from EE programs may contain both sophisticated quantitative analyses as well as intuitive analyses that show ~~that savings~~ that are attributable to the program exist. A compelling part of the narrative can be a simple case study of one or two market participants. A case study can show with a very high degree of internal validity that net savings were obtained, and/or provide examples of NTG factors including ~~freeridership~~free ridership, spillover, and market effects. An intuitive case study often is a useful first step in ~~ana~~ a two-part analysis framework to address estimates of net savings. ~~A framework can include two parts.~~ For example:

- **Part 1:** Establish the existence of the effect, possibly using a case study approach. This can include establishing the existence of savings that are attributable to the program. If the focus of the research is on estimating ~~freeridership~~free ridership or spillover, the first step can involve establishing the existence of these effects.¹⁰⁴ Once existence of an effect is established, ~~the next step involves determining the~~ magnitude of the effect- needs to be determined. This can be easier when the audience is convinced that the effect exists (*i.e.*, the effect is ~~non-zero~~nonzero), and the logic behind the attribution of the effect is set out.
- **Part 2:** This involves the extrapolation of the findings of the case studies to the more general participant population. Once the logic of the case studies is established, it is often possible to define and apply a statistical model consistent with this logic, or to develop an alternative approach to extrapolate the effect. This approach could include any of the methods discussed in this chapter—survey methods, common practice baselines, market

¹⁰⁴ ~~If the study is attempting to estimate the amount of spillover resulting from a program, the first step might be to isolate one or two case studies that compellingly show that spillover exists at participating sites.~~

data analyses and comparisons, structured expert surveys, or historical tracing to examine the influence of a program over time.

The framework above for analyzing net savings can be extended to three steps:

1. Perform an initial high internal validity case study to prove the existence of effects.
2. Establish an estimate range (using discussed methods).—see footnote 52 above). In other words, what is determine a reasonable lower bound for the impacts and what is the highest reasonable bound from the evaluation analyses. This provides information on about the importance of the studied effect and whether it is a part of net savings or an NTG factor (freeridershipfree ridership, spillover, or market effect).
3. Perform analyses using the methods presented in this chapter to develop the best estimate of impacts within the established range.¹⁰⁵

4.2 Selecting the Primary Estimation Method

The selection of appropriate net savings analysesanalysis methods will depend in part on the questions that need to be answered by a net savings study. Research questionsissues that have implications for the net savings approach include:

- Random control trialsRCTs and **quasi-experimental designs** employing DiD and regression methods along with RDD and RED designs (discussed in sectionSection 3.1 of this chapter). These approaches will captureproduce estimates of net savings that address freeridershipfree ridership and participant spillover. Nonparticipant spillover is not directly addressed but can be addressed through surveys of nonparticipants and market effects studies with trade allies.
- **Survey methods**. Survey results can be used to adjust engineering based gross savings estimates for freeridershipfree ridership and participant spillover (discussed in sectionSection 3.2). Nonparticipant spillover can be addressed through surveys of nonparticipants and market effects studies using trade allies.
- **Broader-based methods such as market sales, structured judgment, and historical tracing analyses** can all be used to provide program-specific net savings estimates and address spillover and market effects (discussed in sectionsSections 3.4, 3.6, and 3.8).
- **Common practice baseline methods** can produce estimates by developing baselines on a program basis (discussed in sectionSection 3.3). This approach may not fully address freeridershipfree ridership or participant spillover as, because it does not account for self-selection bias. Also, it does not directly address nonparticipant spillover. However, as previously noted, nonparticipant spillover can be addressed through surveys of

¹⁰⁵ —In a survey setting, this approach can help the survey respondent consider first the behavior that might result in lower, and then the higher impacts that might have been achieved if the program had not existed. The thought process developed by this three-step construct can help survey respondents produce better estimates of their most likely behavior: by first thinking through a construct where the respondent is first asked about factors that would result in a low-range value and then factors that would result in a high-range value.

nonparticipants and market effects studies with trade allies. Common practice baseline methods might be viewed as a compromise that balances out over- and ~~under-~~underestimated NTG factors in the net savings estimate.

- **Deemed or stipulated methods** can be set at the program level (discussed in ~~section 3.7~~Section 3.7); however, the applicability from one jurisdiction to another should be considered.
- **Top-down analyses** use aggregate data that represent the overall level of EE effort across all programs, but cannot isolate the effects of a single program (~~discussed in section 3.5~~or measure (discussed in Section 3.5). Top-down models conceptually address all of the NTG factors—free ridership, spillover, and market effects.

How can estimates of net savings on a program basis be combined with information ~~on~~about program implementation effectiveness? Approaches that provide estimates of net savings but also include elements that involve gathering information directly from participants, nonparticipants, and trade allies can be useful for improving program performance. For example, some programs are designed to minimize ~~freeridership~~free ridership to improve overall resource effectiveness ~~while other programs~~and others focus on expanding the magnitude of spillover and market effects. For these programs, specific estimates of ~~freeridership~~free ridership, spillover, and market effects—particularly if they are provided over a longer time period (every ~~two~~2 years)—can be used to assess overall program effectiveness.

Can evaluators estimate aggregate net savings from a portfolio of programs? ~~All of the~~ estimation approaches presented here, except the top-down analyses, can produce program-specific estimates that evaluators can aggregate up to the portfolio level. ~~However, top~~Top-down methods are designed to work with aggregate data, particularly at the regional level. ~~Top-down models conceptually address all of the NTG factors—freeridership, spillover, and market effects.~~

Other factors that influence the selection of appropriate methods will vary by program type, delivery, sector, and maturity. A recent ~~freeridership~~free ridership and spillover methodology study for the Massachusetts Program Administrators describes the key elements evaluators should consider when choosing a method (Tetra Tech et al., 2011). This study addressed the following factors:

- **Availability of market sales data with a meaningful comparison group.** If market sales data are available on the total sales of both efficient and standard equipment over time, these data are available for the program area, and there is an appropriate comparison area for the appropriate time period, total program effects may be estimated based on these data.

The ideal strategy is to compare the magnitude of the change in sales of energy-efficient equipment relative to the sales of standard equipment in the program area and the comparison area. However, the program ~~itself~~ tends to produce systematic differences between the program and ~~control~~comparison areas. Therefore, where a program has been operating for a long period of time, it is very difficult to find a comparable comparison area.

- **Homogeneity of the measure and the consumers.** ~~Random control trials~~RCTs and quasi-experimental designs work best when there are a large number of similar consumer types and measures. ~~Since large~~Large custom programs are likely to have fewer projects, ~~it is possible that~~so a few (or even one) very large project(s) can have a significant influence on ~~freeridership~~free ridership or spillover. Therefore, the evaluator should use multiple approaches that allow for a greater focus on ~~the~~ consumers that drive the overall impacts to confirm the findings for that program. Methods based on market data or samples of consumers who are making similar purchase decisions may not apply to programs with custom measures.
- **Likelihood of substantial upstream effects unknown to end-use participants.**¹⁰⁶ If there is a reasonable likelihood of substantial upstream effects that an end-use participant would not know about, then conducting an evaluation by using participating end-user surveys alone will tend to understate the effect of the program (even if consumers answer accurately from their ~~perspective~~perspectives). These situations require either information for the market as a whole (if the market sales-based approach is viable) or a combination of participant end-user and vendor surveys. ~~For example, the participating customer would not know that the program influence has changed what options are available, lowered the price of the efficient options, and/or increased the sales staff's knowledge and interest in promoting the efficient option.~~
- **Cost/value ~~trade-off~~tradeoffs.** Some methods that provide more credible results are ~~more costly~~costlier. This cost may be justified for ~~t~~ program components that are important to the portfolio, but not for all components. Importance to the portfolio is typically related to the level of spending or savings associated with a program component. However, a component's importance can also depend on future program plans or other "visibility" factors. The systematic assessment of the value of information gained by net savings estimation approaches ~~ass~~ compared to the cost of the research is needed to better balance the requests to meet confidence and precision levels for estimates. A target of 90% confidence at $\pm 10\%$ precision simply may not be reasonable for all but the largest programs in a portfolio. This systematic approach can examine the impacts on ratepayers from incorrectly attributing savings to a program. If it is a small program, the impacts on ratepayers will be small as measured with 90% confidence and 15% or 20% precision using a one-~~t~~tailed test. This can substantively reduce evaluation costs with little impact on the overall equity tradeoffs between ratepayers and utilities.
- **Data quality.** Data quality is a critical factor for all methods. Typical examples of potential limitations to good data quality are: (1) insufficient information in program tracking databases; (2) lack of clear definitions of what is contained in tracking systems (that is, a data dictionary); (3) limitations on the availability of nonparticipant data (including billing data); ~~and~~; (4) insufficient number of years of available billing data for participants; ~~and~~ (5) limitations on the availability of market sales data.

¹⁰⁶ ~~For example, the participating customer may not know that the program influence has changed what options are available, lowered the price of the efficient options, and/or increased the sales staff's knowledge and interest in promoting the efficient option.~~

4.3 Methods Applicable for Different Conditions

Table 13 lists methods that are suitable for programs with particular features:¹⁰⁷ (based on Tetra Tech et al. [2011]). Programs operate in a particular context and choosing the appropriate evaluation methods requires balancing the advantages and disadvantages of each method. Thus, this table does not list recommendations for a preferred method for a given situation. Rather, it indicates which of the available methods are applicable to programs with specific features. The scales (i.e., low to high) represented in the table for typical cost and complexity are meant to provide an indication of applicability and cost or complexity relative to other methods in the table.

DRAFT: NOT FOR CITATION

¹⁰⁷ ~~This is based on Tetra Tech et al. (2011) prepared for the Massachusetts program administrators.~~

Table 13. Summary of ~~methods applicable~~ Methods Applicable to ~~different conditions~~ Different Conditions

Net Savings Method	Surveyed Group	Applicability				Typical Cost or Complexity	Special Requirements
		Custom Measures	Measures With Few, Diverse Participants	Large Numbers of Similar Participants	Measures With Substantial Upstream Influence Invisible to Consumers		
Randomized control trials (RCT) and Quasi-Experimental Design using Differences in Differences (DiD) RCTs using DiD	None <u>None necessary, but could be conducted to help validate the baseline as an appropriate counterfactual scenario</u>	Poor	Poor	Good	Poor	Low	Random assignment of participants and controls or matched nonparticipant comparison group
Quasi-experimental design	None necessary but could be conducted to validate or develop better baselines	<u>Poor</u>	<u>Poor</u>	<u>Good</u>	<u>Poor</u>	<u>Low</u>	<u>Matched nonparticipant comparison group</u>
Regression models—Billing data analyses with control variables and Linear Fixed Effects Regression (LFER)	Participating consumers and comparison group consumers	Poor	Poor	Good if there is a valid comparison group	Good if there is a valid comparison group	Low	Need control variables that influence energy use across participants and nonparticipants
Survey based—participants, nonparticipants, and market actors	Participating end-end users	Good	Good	Good	Poor unless combined with retailer or contractor surveys	Medium	Counterfactual baseline based on survey responses
	Participating and Nonparticipating <u>nonparticipating</u> end_users	Poor	Poor	Good	Poor unless combined with retailer or contractor surveys	Medium-High	Nonparticipants must be representative of participants

DRAFT #4 | Not for Citation

Net Savings Method	Surveyed Group	Applicability				Typical Cost or Complexity	Special Requirements
		Custom Measures	Measures With Few, Diverse Participants	Large Numbers of Similar Participants	Measures With Substantial Upstream Influence Invisible to Consumers		
	Retail store managers and contractors	Good	Good	Medium	Good	Medium	
Survey based - qualitative sales and Counterfactual Scenario counterfactual scenario	Retail store managers and contractors	Poor	Poor	Good	Good	Low	
Structured expert judgment	Experts	Depends on quality of input methods				Low	
Market sales data (cross-sectional studies)	None	Poor	Poor	Good	Good	Low if data are available; high or not possible if data must be developed	Defined market segment
	Manufacturers and regional buyers and distributors	Poor	Poor	Good	Good	Low	
	Retail store managers and contractors	Good	Good	Medium	Good	Medium	
Common practice baseline	Participating and Nonparticipating end-user surveys or market sales data are used	Poor	Poor	Good	Good	Medium to high	Defined market segment
Top-down methods for regional application	None	Requires data on aggregate energy consumption and information on EE effort (expenditures or related program)				Depends on the cost of	Aggregate data available on

DRAFT #4 | Not for Citation

Net Savings Method	Surveyed Group	Applicability			Typical Cost or Complexity	Special Requirements
		Custom Measures	Measures With Few, Diverse Participants	Large Numbers of Similar Participants		
		variable) for a large number of cross-sectional observations over a period of time-			compiling the initial data set dataset	geographic cross-sections

DRAFT: NOT FOR CITATION

4.4 Planning Net Savings Evaluations—~~Issues to be~~ To Be Considered

Evaluation planners ~~must~~should consider a number of practical issues when planning a net savings evaluation. These include the use of the information, maturity of the program, timing of the study, frequency of net savings estimation, and whether to use multiple approaches. The following bullets summarize ~~provide direction when considering~~ these issues:

- **Use of the information.** It is important to consider how the results of the net savings evaluation will be used and the audience for which the evaluation is intended. This can include shareholder incentives, resource plans, program design, and environmental targets (for example, carbon emissions), among other policy goals.¹⁰⁸
- **Maturity of the program.** Almost all programs are assumed to have some ~~freeridership~~free ridership. The conventional wisdom is that as the program matures (all else equal), ~~freeridership~~observed free ridership will increase during the study period, but so will spillover and market effects. As a result, it becomes important to test for ~~the existence of~~ spillover and market effects as a program matures.
- **Timing of data collection.** To estimate ~~freeridership~~free ridership, the ~~timing of the data collection~~should occur be collected as soon as possible after program participation. This timely measurement minimizes recall bias (Baumgartner, 2013), provides apt feedback on program design, and reduces the possibility that the key decision-maker or market actor is no longer available. However, if the objective is to estimate spillover, the ideal time to collect data is at least ~~one to two~~1–2 years after program participation, as this allows sufficient time for spillover to occur. Finally, if the objective is to estimate market effects, ~~then~~ regular data collection over a period of time is required.
- **Frequency of net savings estimation.** The frequency of net savings or NTG analyses depends on the use of the information. If it is a component of financial incentives for a program administrator, evaluators may need to conduct these studies more frequently. Usually, there is no need to perform detailed net savings studies more than every other year. But, it also depends on the methods used. A statistical analysis of a residential behavioral program can be estimated every year ~~since, because~~ persistence is an important issue and ~~the study~~ costs ~~of the study~~ are low. ~~The Northeast Energy Efficiency Partnerships~~NEEP recommends that net savings estimates be made every ~~two to five~~2–5 years (Titus ~~et al., and Michals~~ 2008) ~~as there are~~because a number of factors ~~that~~ can ~~make~~cause estimates of net savings ~~can~~to change over time.
- **Triangulation of NTG approaches.** Using data from multiple sources limits the effects of self-report bias and measurement error (Baumgartner, 2013). Using an in-depth methodology with multiple sources also allows evaluators to weight the value of

¹⁰⁸ ——— For example, ~~the regional net savings research project (NEEP,~~ (2012) showed that “compared to New England and New York, states in the Mid-Atlantic more commonly use evaluated gross savings for utility regulatory compliance and net savings for program planning and measurement of cost effectiveness. In contrast, New England and New York are more likely to use evaluated net savings; in doing so, they apply NTG values prospectively rather than retrospectively.”

responses from different decision-makers (Megdal et al., 2009). Other data sources often used are: (1) interviews with key decision-makers at the site; (2) project file reviews or project analysis that looks at barriers to project installation, how the project addressed those barriers, and documentation on the participant's decision to go forward with the project; and (3) market data collection, which might include analyses of market sales and shipping data and surveys of market actors (GDS Associates, Inc. et al., 2010; SEE Action, 2012b).

- **Some evaluation issues are best addressed prior to rolling out a new or revised EE program.** Program design personnel and evaluators should work together in advance of implementing a program design that includes random assignment to discuss the data needed for evaluation that must be collected as part of program implementation.

4.5 Trends and Recommendations in Estimating Net Savings

As discussed in ~~the preceding section~~ Section 4.4, the choice of approach for estimating net savings will vary depending on the questions asked, the characteristics of the program(s) evaluated, and the ultimate use of the data. However, there are trends in the application of methods:

- The expanded use of informational and behavioral EE programs is leading to a greater use of ~~random control trials~~ RCTs and quasi-experimental designs that employ some form of randomization (RDD or RED) to help address self-selection.
- The complexity of programs and the need for assessing market effects is leading to a greater use of informed expert panels and Delphi-types of analyses.
- The need to examine trends in program performance over time and impacts on markets over time is resulting in long-term planning for net savings and NTG factor analyses (for example, regular studies conducted with panel data).
- Net savings studies are increasingly embedded in survey analyses that are also designed to gather information ~~on~~ about program implementation effectiveness.
- The value of information from net savings studies is being considered in a more structured manner to help manage evaluation costs. Achieving 90% confidence and 10% precision may be important for a very large EE program, but for a program that is one tenth of the size of the largest program, precision levels are being generated that represent only 1% of the large program. Also, one-tailed tests should be ~~more commonly~~ considered, ~~as~~ because for some applications, it ~~is~~ may be more important to attain a threshold level of net savings ~~with a certain level of confidence~~ than it is ~~that a program may exceed to bound the net savings target by~~ estimate both above and below using a ~~specific amount~~ two-tailed test. A one-tailed targeted precision level still allows for the calculation of the upper end to the confidence interval (Navigant, Violette and Rogers (2012), and there is value in knowing if there was a high likelihood that the target was exceeded by a given amount. The appropriate level of confidence and precision targets are now often reviewed by ~~both~~ EE program administrators and regulators to provide fair

attribution estimates that ~~protect~~minimize risks to ~~both~~-ratepayers and to utilities receiving incentives.⁴⁰⁹ Navigant (2013a) discusses a loss function approach for assessing the value of information from net savings studies; and information on sampling and the tradeoffs between confidence and precision for EE evaluation can be found in Violette and Rogers (2012) and Khawaja et al. (2013).

It has always been important to consider evaluation options ~~prior to~~before implementing an EE program or portfolio of programs. However, the importance of planning the types of net savings studies that are needed and the frequency of this measurement prior to program implementation are becoming critically important. Net savings studies embedded in experimental designs that are established prior to consumers becoming program participants allow for:

- The consideration of randomized designs
- The development of the data platform for estimating consumption-based models (including top-down models)
- The collection of information needed for well-run structured expert panel studies.

In conclusion, net savings methodologies continue to evolve and improve over time. No ~~one~~single methodology is appropriate for all programs or measures, and a single methodology is often not the best choice for estimating program or measure net savings. In the end, jurisdictions should design evaluation plans to assess net savings in conjunction with the key stakeholders considering:

- The appropriate schedule for the evaluation effort over time, taking into account the expected value of the information produced versus the cost of the research effort
- Program design and maturity
- The contribution of the program to overall portfolio savings (past, current, planned)
- The evaluation budget, objectives, and value
- Observations and lessons learned from other jurisdictions.

Finally, adequately documenting the methods used and effectively communicating the results of any net savings study is~~are~~ important. The beginning of this chapter presents a framework for persuasive communication.

⁴⁰⁹ ~~——— Navigant (2013) discusses a loss function approach for assessing the value of information from net savings studies; and Navigant (2012) presents information on sampling and the tradeoffs between confidence and precision.~~

DRAFT: NOT FOR CITATION

References

Abadie, A.; Imbens, G.W. (2011). "Bias-Corrected Matching Estimators for Average Treatment Effects." *Journal of Business and Economic Statistics* 29(1).

AEP (2012). Appendix H, Evaluation of Home Energy Reports, prepared by Navigant. <http://dis.puc.state.oh.us/DocumentRecord.aspx?DocID=9f64b688-a24c-4be5-a7f4-b256403dbb3f>.

Agnew, K.; Goldberg, M. (2013). "Whole-Building Retrofit with Consumption Data Analysis Evaluation Protocol." Chapter 8 in *The Uniform Methods Project: Methods for Determining Energy Efficiency Savings for Specific Measures*. www1.eere.energy.gov/wip/pdfs/53827-8.pdf.

Angrist, J.D.; Pischke, J.-S. (2008). *Mostly Harmless Econometrics: An Empiricist's Companion*. Princeton, NJ: Princeton University Press.

Arimura, T.H.; Li, S.; Newell, R.G.; Palmer, K. (2011). *Cost Effectiveness of Electricity Energy Efficiency Programs*. Resources for the Future Discussion Paper 09-48-Rev.

Auffhammer, M.; Blumstein, C.; Fowlie, M. (2008). "Demand-Side Management and Energy Efficiency Revisited." *The Energy Journal* 29(3): 91–103.

Baumgartner, R. (2013). "Survey Design and Implementation Cross-Cutting Protocols for Estimating Gross Savings." Chapter 12 in *The Uniform Methods Project: Methods for Determining Energy Efficiency Savings for Specific Measures*. www.nrel.gov/docs/fy13osti/53827.pdf.

BC Hydro (2012). *Review of a Top Down Evaluation Study: Rivers & Jaccard*. Prepared for BC Hydro, by Navigant Consulting, Inc., April.

Bodmann, S. (2013). Controlling for Program Participation Self-Selection Bias. Paper and Presentation at the International Energy Program Evaluation Conference.

Bradlow, E. (1998). "Encouragement Designs: An Approach to Self-Selected Samples in an Experimental Design." *Marketing Letters* 9(4), November.

Cadmus (2012a). *CPUC Macro Consumption Metric Pilot Study: Final Report*. Prepared for the California Public Utilities Commission, October.

Cadmus (2012b). *Efficiency Maine Trust Residential Lighting Program Evaluation: Final Report*. Prepared for the Efficiency Maine Trust. www.energymaine.com/docs/Efficiency-Maine-Residential-Lighting-Program-Final-Report_FINAL.pdf

Cadmus (2013). *Focus on Energy Calendar Year 2012 Evaluation Report, Volume II and Appendices A through O*. Prepared for the Public Service Commission of Wisconsin. https://focusonenergy.com/sites/default/files/FOC_XC_CY%2012%20Report%20Volume%20%20Final_05-3-2013.pdf.

Cadmus; Navigant (2012). *New York Energy Smartsm Products Program Market Characterization and Assessment Evaluation: Final Report*. Prepared for The New York State Energy Research and Development Authority, Victoria Engel-Fowles Project Manager, Project Number 9875, February. See bullet six study in the link:

<http://energyplan.ny.gov/Home/Publications/Program-Planning-Status-and-Evaluation-Reports/NYES-Evaluation-Contractor-Reports/2012-Reports/Market-Analysis.aspx>

Cadmus; Navigant Consulting; Opinion Dynamics Corporation (2012). *2012 Residential Heating, Water Heating, and Cooling Equipment Evaluation: Net-to-Gross, Market Effects, and Equipment Replacement Timing*. Prepared for the Electric and Gas Program Administrators of Massachusetts.

Castor, S. (2012). *Fast Feedback Results*. 2011 Final Report prepared for Energy Trust of Oregon. http://energytrust.org/library/reports/Fast_Feedback_-_20110.pdf

Commonwealth Edison (2012). *Energy Efficiency / Demand Response Plan: Plan Year 3 (6/1/2010-5/31/2011) Evaluation Report: Home Energy Reports*. Prepared by Navigant Consulting. www.icc.illinois.gov/downloads/public/edocket/323839.pdf

Cook, T.; Scriven, M.; Coryn, C.L.; Evergreen, S.D.H. (2010). "Contemporary Thinking About Causation in Evaluation." *American Journal of Evaluation* 31:105. <http://aje.sagepub.com/content/31/1/105>

Demand Research, LLC (2012). *Macro Consumption Metrics Pilot Study: Final Report*. Prepared for: California Public Utilities Commission Energy Division, November.

Diamond, A.; Haninmueller, J. (2007). *The Encouragement Design for Program Evaluation*. Harvard University and International Finance Corporation. See:

Dubin, J.; McFadden, D. (1984). "An Econometric Analysis of Residential Electric Appliance Holdings and Consumption." *Econometrica* 52(2):345–362

Duflo, E.; Glennerster, R.; Kremer, M. (2007). *Using Randomization in Development Economics Research: A Toolkit*. Centre for Economic Policy Research, Discussion Paper 6059.

Eto, J. (1988). "On Using Degree-days to Account for the Effects of Weather on Annual Energy Use in Office Buildings." *Energy and Building* 12, 113–127.

Eto, J.; Prahl, R.; Schlegal J. (1996). *A Scoping Study on Energy-efficiency Market Transformation by California Utility DSM Programs*. Lawrence Berkeley National Laboratory. <http://emp.lbl.gov/sites/all/files/lbnl%20-%2039058.pdf> .

Fagan, J.; Messenger, M.; Rufo, M.; Lai, P. (2009). "A Meta-Analysis of Net to Gross Estimates in California." Paper presented at the 2009 AESP conference.

Feng, W., Jun, Y.; Xu, R. (2006). *A Method/Macro Based on Propensity Score and Mahalanobis Distance to Reduce Bias in Treatment Comparison in Observational Study*. Public Health Research, paper pr05. www.lexjansen.com/pharmasug/2006/publichealthresearch/pr05.pdf

Fowlie, M.; Wolfram, C. (2009). *Evaluating the Federal Weatherization Assistance Program Using a Randomized Encouragement Design (RED)*. Presented to Environmental Energy Technologies Division (EETD), Lawrence Berkeley National Laboratory, September. http://eetd.lbl.gov/sites/all/files/lbl_09-11-09.pdf

Fowlie, M.; Wolfram, C. (undated). *An Experimental Evaluation of the Federal Weatherization Assistance Program*. Presentation prepared for the Michigan Public Service Commission. www.dleg.state.mi.us/mpsc/electric/workgroups/lowincome/fowlie_wolfram.pdf

GDS Associates, Inc. (2012) *GDS Analysis of Proposed Department of Energy Evaluation, Measurement & Verifications Protocols*. Final report prepared for the National Rural Electric Cooperative Association. <https://www.nreca.coop/wp-content/uploads/2013/12/EMVReportAugust2012.pdf>

GDS Associates, Inc.; Nexant; Mondre Energy (2010). *Net Savings: An Overview*. RFP 2009-prepared for the Statewide Evaluator.

Goldberg, M.; Kademan, E. (1995). *Is It Net or Not? A Simulation Study of Two Methods*. In *Energy Program Evaluation: Uses, Methods, and Results*, 459–465. Chicago, IL: National Energy Program Evaluation Conference.

Greene, W. (2011). *Econometric Analysis*, 7th Ed., Prentice Hall.

Guo, S.; Fraser, M. (2010). *Propensity Score Analysis: Statistical Methods and Applications*, SAGE Publications, Inc. (Note: Chapter 4 provides an updated discussion of the Heckman models to self-selection along with appropriate caveats. This discussion can be found at: www.sagepub.com/upm-data/30234_Chapter4.pdf

Gwet, K.L. (2010). *Inter-Rater Reliability Using SAS: A Practical Guide for Nominal, Ordinal, and Interval Data*. Gaithersburg, MD: Advanced Analytics, LLC.

Gwet, K.L. (2012). *Handbook of Inter-Rater Reliability: The Definitive Guide to Measuring the Extent of Agreement Among Multiple Raters*. Gaithersburg, MD: Advanced Analytics, LLC.

Haeri, H. (2013). *Uniform Methods Project: Methods for Determining Energy Efficiency Savings for Specific Measures*, Chapter 1: The Introduction. www.nrel.gov/docs/fy13osti/53827.pdf.

Haeri, H.; Khawaja, M.S. (2012). *The Trouble with Freeriders*. Public Utilities Fortnightly. www.cadmusgroup.com/wp-content/uploads/2012/11/Haeri-Khawaja-PUF-TroublewithFreeriders.pdf.

Hall, N; Ladd, D.; Khawaja, M.S. (2013). “Setting Net Energy Impact Baselines: Building Reliable Evaluation Approaches.” Paper presented at the 2013 International Energy Program Evaluation Conference, Chicago, IL.

Heckman, J.J. (1979). “Sample Selection Bias as a Specification Error” *Econometrica* 47(1): 153–161.

Ho, D.; Imai, K.; King, G; Stuart, E. (2007). “Matching as Nonparametric Preprocessing for Reducing Model Dependence in Parametric Causal Inference.” *Policy Analysis* 15(3):199–236.

Hoefgen, L.; Clendenning, G.; Osman, A.; Keating, K.; Vine, E.; Lee, A.; Stewart, J.; Stoops, J. (2011). “Finding and Counting Market Effects: A New Construction Program Example.” Paper presented at the 2011 International Energy Program Evaluation Conference, Chicago, IL.

Imbens, G.; Lemieux, T. (2010). “Regression Discontinuity Designs: A Guide to Practice.” *Journal of Economic Literature* 48:281–355.

Itron, Inc. (2010). 2006-2008 Evaluation Report for PG&E Fabrication, Process and Manufacturing Contract Group. Prepared for the California Public Utilities Commission, Energy Division. [www.calmac.org/publications/PG%26E Fab 06-08 Eval Final Report.pdf](http://www.calmac.org/publications/PG%26E%20Fab%2006-08%20Eval%20Final%20Report.pdf); [www.calmac.org/publications/PG%26E Fab 06-08 Eval Final Report Appendices.pdf](http://www.calmac.org/publications/PG%26E%20Fab%2006-08%20Eval%20Final%20Report%20Appendices.pdf)

Keating, K. (2009). “Freeridership Borscht: Don’t Salt the Soup.” Paper presented at the 2009 International Energy Program Evaluation Conference.

KEMA, Inc. (2010). *Final Evaluation Report: Upstream Lighting Program, Volume 1*. CALMAC Study ID: CPU0015.01; Prepared for: California Public Utilities Commission, Energy Division. Prepared by: KEMA, Inc., Prime Contractor: The Cadmus Group, Inc. [www.calmac.org/publications/FinalUpstreamLightingEvaluationReport_Vol1 CALMAC 3.pdf](http://www.calmac.org/publications/FinalUpstreamLightingEvaluationReport_Vol1_CALMAC_3.pdf)

Keneipp, M., Meurice, J.; Alspector, D.; Sutter, M.; Krause, R.; Hines, T. (2011). “Getting MIF’ed: Accounting for Market Effects in Residential New Construction Programs.” Paper presented at the 2011 International Energy Program Evaluation Conference. Boston, MA.

Kennedy, P. (2008). *A Guide to Econometrics*, 6th Edition. Wiley-Blackwell, April.

Khawaja, M.S.; Rushton, J.; Keeling, J. (2013). “Sample Design Cross-Cutting Protocols.” Chapter 11 in *The Uniform Methods Project: Methods for Determining Energy Efficiency Savings for Specific Measures*, www.nrel.gov/docs/fy13osti/53827.pdf.

Loughran David S.; Kulick, J. “Demand-Side Management and Energy Efficiency in the United States”. *The Energy Journal* 25 (1), 19-41. 2004.

Mahalanobis, P. (1936). “On the Generalised Distance in Statistics.” *Proceedings of the National Institute of Sciences of India* 2 (1):49–55. www.new.dli.ernet.in/rawdataupload/upload/insa/INSA_1/20006193_49.pdf.

McMenamin, J.S. (2008). *Defining Normal Weather for Energy and Peak Normalization*. Itron, Inc. <https://www.itron.com/PublishedContent/Defining%20Normal%20Weather%20for%20Energy%20and%20Peak%20Normalization.pdf>.

McKenzie, D (2009). “Impact Assessments in Finance and Private Sector Development -- What Have We Learned and What Should We Learn?” The World Bank

Development Research Group, Policy Research Working Paper 4944, May. See: <https://openknowledge.worldbank.org/bitstream/handle/10986/4137/WPS4944.pdf?sequence=1>

Megdal, L.; Patil, Y.; Gregoire, C.; Meissner, J.; Parlin, K. (2009). “Feasting at the Ultimate Enhanced Freeridership Salad Bar.” Paper presented at the International Energy Program Evaluation Conference, Portland, OR.

www.anevaluation.com/pubs/Salad%20Bar%202009%20IEPEC%20paper%205-12-09.pdf.

Messenger, M.; Bharvirkar, R.; Golemboski, B.; Goldman, C.; Schiller, S. (2010). *Review of Evaluation, Measurement and Verification Approaches Used to Estimate the Load Impacts and Effectiveness of Energy Efficiency Programs*. Lawrence Berkeley National Laboratory. <http://emp.lbl.gov/sites/all/files/lbnl-3277e.pdf>.

Miller, K. (2011). “Cognitive Interviewing.” In *Question Evaluation Methods: Contributing to the Science of Data Quality*, pp. 51–76. Jennifer Madans, Kristen Miller, Aaron Maitland, and Gordon Willis (Eds.) Hoboken, NJ: John Wiley & Sons, Inc.

Mort, D. (2013). “Metering Cross-Cutting Protocols.” Chapter 9 of *The Uniform Methods Project: Methods for Determining Energy Efficiency Savings for Specific Measures*.” www.nrel.gov/docs/fy13osti/53827.pdf.

Navigant (2013). *Custom Free Ridership and Participant Spillover Jurisdictional Review*. Prepared for the Sub-Committee of the Ontario Technical Evaluation Committee, May. www.ontarioenergyboard.ca/documents/TEC/Evaluation%20Studies%20and%20Other%20Reports/Ontario%20NTG%20Jurisdictional%20Review%20-%20Final%20Report.pdf.

Navigant (2013). *Impact and Process Evaluation of the U.S. Department of Energy’s Wind Powering America Initiative*. Prepared for: Department of Energy Office of Energy Efficiency and Renewable Energy, Final Report, DOE/EE-0897 May. http://www1.eere.energy.gov/analysis/pdfs/wind_powering_america_evaluation_2013.pdf

NEEP (2012). “Regional Net Savings Research, Phase 2: Definitions and Treatment of Net and Gross Savings in Energy and Environmental Policy.” Submitted to the Northeast Energy Efficiency Partnerships: Evaluation, Measurement, and Verification Forum, by NMR Group, Inc. and Research Into Action, December.

<https://www.neep.org/Assets/uploads/files/emv/NEEP%20-%20Regional%20Net%20Savings%20Report%2012-05-12.pdf>.

New York Department of Public Service (2012). *Evaluation Plan Guidance for EEPS Program Administrators*, Update #3, Appendix F. Albany, New York.

New York Department of Public Service (2013a). *Guidelines for Calculating the Relative Precision of Program Net Savings Estimates*. Appendix I.

New York Department of Public Service (2013b). *Guidelines for Estimating Net-To-Gross Ratios Using the Self-Report Approach*. Appendix H.

NMR Group, Inc. and Research Into Action, Inc. (2010). *Net Savings Scoping Paper*. Revised Draft prepared for Northeast Energy Efficiency Partnerships: Evaluation, Measurement, and

[Verification Forum. www.neep.org/Assets/uploads/files/emv/emv-products/FINAL Net Savings Scoping Paper 11-13-10.pdf.](http://www.neep.org/Assets/uploads/files/emv/emv-products/FINAL_Net_Savings_Scoping_Paper_11-13-10.pdf)

[NMR Group, Inc. and Tetra Tech \(2011\). *Cross-Cutting Net to Gross Methodology Study for Residential Programs – Suggested Approaches*. Final report prepared for the Massachusetts Program Administrators. www.ma-eeac.org/Docs/8.1_EMV%20Page/2011/2011%20Residential%20Studies/Residential%20MA%20NTG%20Methods%20Final%20072011.pdf.](http://www.ma-eeac.org/Docs/8.1_EMV%20Page/2011/2011%20Residential%20Studies/Residential%20MA%20NTG%20Methods%20Final%20072011.pdf)

[NMR Group, Inc.; KEMA; Cadmus Group, Inc.; Tetra Tech \(2011\). *Massachusetts ENERGY STAR Lighting Program: 2010 Annual Report*. Final Report prepared for Energy Efficiency Advisory Council Consultants, Cape Light Compact, NSTAR, National Grid, Until, and Western Massachusetts Electric. https://www.efis.psc.mo.gov/mpsc/commoncomponents/viewdocument.asp?DocId=935690223.](https://www.efis.psc.mo.gov/mpsc/commoncomponents/viewdocument.asp?DocId=935690223)

[Nonresidential Net-To-Gross Ratio Working Group \(2012\). *Methodological Framework for Using the Self-Report Approach to Estimating Net-to-Gross Ratios for Nonresidential Customers*. Prepared for the Energy Division, California Public Utilities Commission.](#)

[Oak Ridge National Laboratory \(1991\). *Handbook to DSM Program Evaluation*. Eric Hirst and John Reed, eds., NTIS Pubs., Washington, DC, # ORNL/CON -336, December.](#)

[Parfomak, P.; Lave, L. \(1996\). “How Many Kilowatts Are in a Negawatt? Verifying the Ex-Post Estimates of Utility Conservation Impacts at a Regional Level.” *Energy Journal* 17 \(4\).](#)

[Peters, J.; McRae, M. \(2008\). “Freeridership Measurement Is Out of Sync with Program Logic...or, We’ve Got the Structure Built, but What’s Its Foundation.” In *Proceedings of the 2008 ACEEE Summer Study on Energy Efficiency in Buildings, Washington, DC.* www.aceee.org/files/proceedings/2008/data/papers/5_491.pdf.](http://www.aceee.org/files/proceedings/2008/data/papers/5_491.pdf)

[PG&E \(2013\). *Evaluation of Pacific Gas and Electric Company’s Home Energy Report Initiative for the 2010-2012 Program*. Prepared by Freeman, Sullivan Co. Available on the CALMAC.org website: www.calmac.org/publications/2012_PGE_OPOWER_Home_Energy_Reports_4-25-2013_CALMAC_ID_PGE0329.01.pdf.](http://www.calmac.org/publications/2012_PGE_OPOWER_Home_Energy_Reports_4-25-2013_CALMAC_ID_PGE0329.01.pdf)

[Prahl, R.; Ridge, R.; Hall, N.; Saxonis, W. \(2013\). “The Estimation of Spillover: EM&V’s Orphan Gets a Home.” In *Proceedings of the 2013 International Energy Program Evaluation Conference, August.*](#)

[Provencher, B; Vittetoe-Glinsmann, B.; Dougherty, A.; Randazzo, K.; Moffitt, P., Prahl, R. \(2013\). *Some Insights on Matching Methods in Estimating Energy Savings for an Opt-In, Behavioral-Based Energy Efficiency Program*. 2013 International Energy Program Evaluation Conference, Chicago.](#)

[Provencher, B.; Glinsmann, B. \(2013\). *Evaluation Report: Home Energy Reports – Plan Year 4*. Prepared for Commonwealth Edison Company. February.](#)

[Puget Sound Energy \(2012\). *Home Energy Reports Program: Three Year Impact, Behavioral and Process Evaluation*. Prepared for: Puget Sound Energy, Prepared by: KEMA, Inc. https://conduitnw.org/_layouts/Conduit/FileHandler.ashx?RID=849.](https://conduitnw.org/_layouts/Conduit/FileHandler.ashx?RID=849)

[Ridge, R. \(1997\). *Errors in Variables: A Close Encounter of the Third Kind*. In Proceedings of the 1997 International Energy Program Evaluation Conference. August, Chicago, IL.](#)

[Ridge, R.; Baker, M.; Hall, N.; Prah, R.; Saxonis, W. \(2013\). “Gross Is Gross and Net Is Net: Simple, Right?” Paper presented at the 2013 International Energy Program Evaluation Conference, Chicago, IL.](#)

[Ridge, R.; Willems, P.; Fagan, J.; Randazzo, K. \(2009\). “The Origins of the Misunderstood and Occasionally Maligned Self-Report Approach to Estimating Net-to-Gross Ratio.” Paper presented at the 2009 Energy Program Evaluation Conference, Portland, OR.](#)

[Rosenberg, M.; Hoefgen, L. \(2009\). *Market Effects and Market Transformation: Their Role in Energy Efficiency Program Design and Evaluation*. Prepared for the California Institute for Energy and the Environment and the California Public Utilities Commission Energy Division.](#)

[RTF \(2012\). *Guidelines for the Development and Maintenance of RTF Savings Estimation Methods*. NW Council, Released December 4. http://rtf.nwcouncil.org/subcommittees/deemed/.](http://rtf.nwcouncil.org/subcommittees/deemed/)

[Ruegg, R.; Jordan, G. \(2007\). *Overview of Evaluation Methods for R&D Programs: A Directory of Evaluation Methods Relevant to Technology Development Programs*. Prepared for the U.S. Department of Energy: Office of Energy Efficiency and Renewable Energy. https://www1.eere.energy.gov/analysis/pdfs/evaluation_methods_r_and_d.pdf.](https://www1.eere.energy.gov/analysis/pdfs/evaluation_methods_r_and_d.pdf)

[Rufo, M. \(2009\). “Evaluation and Performance Incentives: Seeking Paths to \(Relatively\) Peaceful Coexistence.” In Proceedings of the International Energy Program Evaluation Conference, Portland, OR, August.](#)

[Sacramento Municipal Utilities District \(2011\). *Evaluation Report: OPOWER SMUD Pilot Year2*. Prepared by Navigant. February. http://opower.com/company/library/verification-reports?year=2011.](http://opower.com/company/library/verification-reports?year=2011)

[Sacramento Municipal Utilities District \(2013\). *Load Impact Results from SMUD’s Smart Pricing Options Pilot*. Prepared by Freeman Sullivan & Co. for Sacramento Municipal Utility District – SMUD contact Ms. Lupe Jimenez](#)

[Scriven, M. \(2009\). “Demythologizing Causation and Evidence.” In *What Counts as Credible Evidence in Applied Research and Evaluation Practice*. Stewart I Donaldson, Christina A. Christie, and Melvin Mark \(Eds.\). Los Angeles, CA: SAGE Publications.](#)

[Sebold, F.D.; Fields, A.; Skumatz, L.; Feldman, S.; Goldberg, M.; Keating, K.; Peters, J. \(2001\). *A Framework for Planning and Assessing Publicly Funded Energy Efficiency*. www.calmac.org/events/20010301PGE0023ME.pdf .](http://www.calmac.org/events/20010301PGE0023ME.pdf)

[SEE Action \(2012a\). *Evaluation, Measurement, and Verification \(EM&V\) of Residential Behavior-Based Energy Efficiency Programs: Issues and Recommendations*. Prepared by A. Todd, E. Stuart, S. Schiller, and C. Goldman, Lawrence Berkley National Laboratory. <http://behavioranalytics.lbl.gov/reports/behavior-based-emv.pdf>.](http://behavioranalytics.lbl.gov/reports/behavior-based-emv.pdf)

[SEE Action \(2012b\). *Energy Efficiency Program Impact Evaluation Guide*. Prepared by Steven R. Schiller, Schiller Consulting, Inc. See: \[www1.eere.energy.gov/seeaction/pdfs/emv_ee_program_impact_guide.pdf\]\(http://www1.eere.energy.gov/seeaction/pdfs/emv_ee_program_impact_guide.pdf\).](http://www1.eere.energy.gov/seeaction/pdfs/emv_ee_program_impact_guide.pdf)

[Shadish, W.R., Cook, T.D.; Campbell, D.T. \(2002\). *Experimental and Quasi-Experimental Designs for Generalized Causal Inference*. New York: Houghton Mifflin Company.](#)

[Southern California Edison \(2012\). *Edison SmartConnect Demand Response and Energy Conservation Annual Report*, prepared by David Hanna et al., Itron, Inc. for Eric Bell, SCE project manager. \[https://www.pge.com/regulation/DemandResponseOIR/Pleadings/SCE/2012/DemandResponseOIR_Plea_SCE_20120430_237124.pdf\]\(https://www.pge.com/regulation/DemandResponseOIR/Pleadings/SCE/2012/DemandResponseOIR_Plea_SCE_20120430_237124.pdf\)](https://www.pge.com/regulation/DemandResponseOIR/Pleadings/SCE/2012/DemandResponseOIR_Plea_SCE_20120430_237124.pdf)

[Stryker, A.; Gaffney, K. \(2013\). “Why the Light Bulb Is No Longer a Textbook Example for Price Elasticity: Results from Choice Experiments and Demand Modeling Research.” In *Proceedings of the International Energy Program Evaluation Conference*, Chicago, IL, August.](#)

[Stuart, E.A. \(2010\). “Matching Methods for Causal Inference: A Review and a Look Forward.” *Statistical Science* 25\(1\):1–21.](#)

[TecMarket Works; The Cadmus Group; Opinion Dynamics Corporation; Integral Analytics; Building Metrics; Energy Efficient Homes Midwest \(2012\). *Indiana Evaluation Framework*. Prepared for the Indiana Demand Side Management Coordination Committee.](#)

[Tetra Tech, Inc.; KEMA; NMR Group, Inc. \(2011\). *Cross-Cutting \(C&I\) Free Ridership and Spillover Methodology Study Final Report*. Massachusetts Program Administrators. \[www.ma-eaac.org/Docs/8.1_EMV%20Page/2011/2011%20Commercial%20%20Industrial%20Studies/MA%20FR_SO%20CI%20%20Study%20w%20Exec%20Summary%205-26-2011%20v11.pdf\]\(http://www.ma-eaac.org/Docs/8.1_EMV%20Page/2011/2011%20Commercial%20%20Industrial%20Studies/MA%20FR_SO%20CI%20%20Study%20w%20Exec%20Summary%205-26-2011%20v11.pdf\)](http://www.ma-eaac.org/Docs/8.1_EMV%20Page/2011/2011%20Commercial%20%20Industrial%20Studies/MA%20FR_SO%20CI%20%20Study%20w%20Exec%20Summary%205-26-2011%20v11.pdf)

[Titus, E.; Michals, J. \(2008\). “Debating Net Versus Gross Impacts in the Northeast: Policy and Program Perspectives.” *ACEEE Summer Study on Energy Efficiency in Buildings*, \(5\): 312–323. \[http://www.aceee.org/files/proceedings/2008/data/papers/5_429.pdf\]\(http://www.aceee.org/files/proceedings/2008/data/papers/5_429.pdf\).](http://www.aceee.org/files/proceedings/2008/data/papers/5_429.pdf)

[U.S. Department of Energy \(2010\). *Guidance Document #7: Topic: Design and Implementation of Program Evaluations that utilize Randomized Experimental Approaches*. Smart Grid Investment Grant Technical Advisory Group, November. \[https://www.smartgrid.gov/sites/default/files/pdfs/cbs_guidance_doc_7_randomized_experimental_approaches.pdf\]\(https://www.smartgrid.gov/sites/default/files/pdfs/cbs_guidance_doc_7_randomized_experimental_approaches.pdf\).](https://www.smartgrid.gov/sites/default/files/pdfs/cbs_guidance_doc_7_randomized_experimental_approaches.pdf)

[Violette, D. \(2013\). “Persistence and Other Evaluation Issues Cross-Cutting Protocols.” Chapter 13 in *Uniform Methods Project for Determining Energy Efficiency Program Savings for Specific Measures*. NREL/SR-7A30-53827, April. See: \[www1.eere.energy.gov/wip/pdfs/53827-13.pdf\]\(http://www1.eere.energy.gov/wip/pdfs/53827-13.pdf\)](http://www1.eere.energy.gov/wip/pdfs/53827-13.pdf)

Violette, D.; Barkett, B.; Schare, S.; Skumatz, L.; Dimetrosky, S. (2005). *Commercial/Industrial Performance Program (CIPP) Market Characterization, Market Assessment And Causality Evaluation*. Prepared for NYSERDA, Jennifer Ellefsen, Project Number 7721, March.

Violette, D.; Brakken, R.; Schon, A.; Greer, J. (1993). *Statistically-Adjusted Engineering Estimates: What Can The Evaluation Analyst Do About The Engineering Side Of The Analysis?*. Published in the *Proceedings of the 1993 International Energy Program Evaluation Conference (IEPEC)*.

Violette, D.; Keneipp, M.; Ozog, M. (1991). *Impact Evaluation of Demand-Side Management Programs — Volume 1: A Guide to Current Practice*. Electric Power Research Institute Pubs., Palo Alto, CA, #EPRI CU-7179, February.

Violette, D; Ozog M. Cooney, K. (2003), *Retrospective Assessment of the Northwest Energy Efficiency Alliance -- Findings and Report*. Prepared for: Northwest Energy Efficiency Alliance, Ad Hoc Retrospective Committee, December 8. See: http://www.theboc.info/pdf/Eval-BOC_SummittBlue_NEEA_2003.pdf

Violette, D.; Provencher, B. (2012). *Review of a Top Down Evaluation Study: Rivers & Jaccard (2011)*. Prepared for BC Hydro, Navigant Consulting, Inc., April.

Violette, D.; Rogers, B. (2012). *A Sampling Methodology for Custom C&I Programs*. Prepared by Navigant Consulting, Inc., prepared for the Ontario Natural Gas Technical Evaluation Committee, Ontario Energy Board, November. See: report available at Ontario Energy Board website: <http://www.ontarioenergyboard.ca/documents/TEC/Evaluation%20Studies%20and%20Other%20Reports/TEC%20SC%20-%20Sampling%20Method%20-%20Final%20Report%2020121112.pdf>

Violette, D.M.; Provencher, B.; Sulyma I. (2012). “Assessing Bottom-Up and Top-Down Approaches for Assessing DSM Programs and Efforts.” In *International Energy Program Evaluation Conference Proceedings*, Rome, June.

Weiss, C. (1997). “Theory-Based Evaluation: Past, Present, and Future” Special Issue: Progress and Future Directions in Evaluation: Perspectives on Theory, Practice, and Methods, *New Directions in Evaluation*, Volume 1997, Issue 76.

West, S. (2008). “Alternatives to the Randomized Controlled Trial.” *American Journal of Public Health* 98(8). <http://ajph.aphapublications.org/doi/abs/10.2105/AJPH.2007.124446>.

Wooldridge, J.M. (2010). *Econometric Analysis of Cross Section and Panel Data*. Cambridge, MA, MIT Press Ltd.

Appendix: Price Elasticity Studies as a Component of Upstream Lighting Net Savings Studies

Studies of upstream changes in the price for residential lighting products have received attention as a way to complement surveys with market actors, or even replace these surveys with econometric models. [The way in which price can be viewed as a driver of program savings and the importance of other program components is discussed in Stryker and Gaffney \(2013\).](#)

[Price elasticity studies are currently being applied in several jurisdictions.](#) To date, these studies have focused on residential lighting products and, within that category, mostly on CFL sales. [For example,](#) Cadmus (~~2012 and~~2012b, 2013) and KEMA (2010) tested several different methods for estimating the increase in CFL sales resulting from a program-induced price reduction ~~due~~ [caused by](#) program activities (markdowns negotiated with retailers and coupons).

Cadmus (~~2012~~2012b) examined Efficiency Maine's residential lighting program and Cadmus (2013) examined Wisconsin's Focus on Energy residential lighting program. Both studies used a price elasticity approach. These two studies estimated expected bulb purchases (and associated savings) at prices offered under the program and then the purchases that would have occurred at original retail prices. The difference between these two values was viewed as net savings in this study.

Cadmus (~~2012 and~~2012b, 2013) used a single equation regression model where the quantity of CFLs purchased ~~were~~was a function of the price of CFLs and a select set of other independent variables. The data used to estimate this equation included package and bulb sales for each retailer, by model number and by week. The ~~data set~~dataset does not include information ~~on~~about the ~~customers that~~consumers who purchased the CFLs, but does contain information ~~on~~about quantities of CFLs sold and retailer prices. ~~Customer~~Consumer variables desirable in a demand equation would include income and education, but often these variables are not available in the retailers' sales tracking systems.

A regression was estimated relating quantities of ~~CFLs~~CFLs sold by retailer to the price of CFLs that week for each retailer. Other factors such as promotional events were considered in determining consumer purchases. Programmatic factors such as labeling and information dissemination are pervasive throughout the lighting programs and, while potentially important, could not be addressed due to lack of variation across consumer purchases.

These two studies showed an increase in the sales of ~~CFL bulbs~~CFLs as prices decreased due to markdowns negotiated with retailers and discount coupons provided to consumers. The second step of the approach involved estimating what the sales would have been at the higher prices that would have prevailed without the program (that is, the counterfactual scenario).

Considerable effort was made in these price elasticity studies to control for ~~other~~factors [other than price](#) that might also affect CFL sales ~~other than price~~, but it is difficult to show that any method is free of bias. In the case of the Efficiency Maine lighting program, there were three components to the program. Two were linked to price (markdowns and coupons) and a third was linked to overall participation in the Appliance Rebate Program, "with Appliance Rebate Program participants electing to receive a free six-pack of CFL bulbs, via a check-off on the

Appliance Rebate Program application form.” The third part of the program would have provided CFLs at essentially no cost and it is not clear how this would have factored into the analysis.

Cadmus (~~2012 and~~2012b, 2013) present several general caveats to the demand equation approach used in the study. First, ~~they~~it acknowledged that “this estimation method has rarely been used in upstream lighting program evaluations as such data generally have been unavailable. As Efficiency Maine ... tracked these data and shared them for this evaluation, Cadmus found such econometric demand estimation provided the best method for estimating the program’s freeridership.” Second, Cadmus (2013) indicates that it “will continue to look for alternative methods to calculate net-to-gross,” and that “the model used for the ... 2012 evaluation does not account for spillover.”¹¹⁰

KEMA (2010) used price variables to estimate net savings in an upstream lighting study. This study had the benefit of a sizeable data collection effort that included consumer surveys. As part of the in-store consumer intercept research, brief interviews were conducted with shoppers who had just made a lighting purchase (revealed preference) as well as “stated preference” surveys with other consumers recruited randomly. Intercept surveys were conducted with 1,463 ~~customer~~customers across 378 stores.

KEMA (2010) used three primary types of methods for estimating net savings:

- Supplier and consumer self-report methods
- Econometric models
- Total sales (market-based) approach.

Among the econometric modeling efforts, four ~~different~~ econometric models were used:

- Pricing (price formation model)
- Conjoint ~~Elasticity~~elasticity
- Revealed ~~Preference Purchase~~preference purchase
- Stated ~~Preference Purchaser Elasticity~~preference purchaser elasticity.

The first two econometric methods—price formation and the conjoint elasticity model—were both needed to produce a net savings estimate. Revealed preference and stated preference models can produce net savings directly. As a result, there were four econometric models, but only three ~~different~~ approaches for estimating net savings.

The price formation model estimates the percentage reduction in CFL prices that resulted from program incentives. This is combined with the conjoint analysis, which estimated the

¹¹⁰ Cadmus (2012) indicates that spillover is not addressed in this study; however, looking at the overall change in sales in a market caused by price elasticity, has included spillover elements in other studies that use a similar price elasticity approach.

DRAFT #4 | Not for Citation

corresponding percentage increase in market share/sales that result from a price decrease. This allowed the net savings to be calculated by combining the findings from the pricing study with the conjoint demand elasticity study—in other words, the program induced reduction in prices from the pricing study multiplied by the estimate of change in sales ~~due to~~ caused by a lower price from the conjoint study.

KEMA (2010) revealed a preference for store intercepts to survey customers that made actual CFL purchases. These customers were asked to indicate how many CFLs they would have bought compared to their actual purchases at double the price they actually paid. Response categories were: (1) the same amount, (2) fewer, ~~or~~ and (3) none. ~~While~~ Although still based on hypothetical, self-reported responses, the revealed preference respondents may be a more reliable sample because they just made an active purchase decision. However, revealed preference respondents may be somewhat unlikely to indicate they would have paid more for what they just purchased. KEMA (2010) used a random survey of customers, including customers ~~that~~ who did not actually purchase a CFL. KEMA (2010) states that the magnitude of the potential bias across these two methods is unknown, “but it is likely that NTG ratio estimates from stated preference respondents are biased downward and NTG ratio estimates from revealed preference respondents are biased upward.”

The revealed preference model allowed KEMA to use the store-intercept survey data to model CFL purchase rates with and without program effects. This model was based on a logistic regression to model the probability of buying a CFL rather than an “equivalent” non-CFL as a function of price, displays, customer characteristics, and bulb characteristics, by channel. The fitted models were evaluated under program and ~~non-program~~ nonprogram conditions. For each channel, the difference between the probability of purchasing CFLs under the program condition and that under the ~~non-program~~ nonprogram condition was the program-attributable CFL sales share.

In summary, the price elasticity studies completed to date have been limited to residential lighting programs. Cadmus (~~2012 and~~ 2012b, 2013) developed a demand model specification based on an examination of alternative specifications. KEMA (2010) developed several ~~different~~ approaches for examining the change in CFLs sold as a function of program-induced lower prices. KEMA (2010) concluded that from the econometric approaches, the revealed preference model was the preferred approach. It should be noted that these approaches focus on ~~free ridership~~ free ridership and do not address spillover or longer term market effects. Currently, several evaluations are using the price-elasticity method to estimate net savings from residential lighting. An expanded literature will likely provide additional confidence in this method for addressing ~~free ridership~~ free ridership from upstream lighting programs, and possibly an expansion of this method to other residential product programs.

References

~~Abadie, A. and G.W. Imbens (2011). *Bias-Corrected Matching Estimators for Average Treatment Effects*. Journal of Business and Economic Statistics Vol. 29, No. 1.~~

~~AEP Ohio, (2012) Appendix H, Evaluation of Home Energy Reports, prepared by Navigant, May. <http://dis.puc.state.oh.us/TiffToPDF/A1001001A12E15B14941H23668.pdf>~~

~~Agnew, Ken, Goldberg, Mimi (2013). Chapter 8: "Whole-Building Retrofit with Consumption Data Analysis-Evaluation Protocol." *The Uniform Methods Project: Methods for Determining Energy Efficiency Savings for Specific Measures*. http://www1.eere.energy.gov/wip/pdfs/53827_8.pdf.~~

~~Angrist, Joshua D. & Jorn-Steffen Pischke (2008). *Mostly Harmless Econometrics: An Empiricist's Companion*. Princeton, NJ: Princeton University Press.~~

~~Arimura, Toshi H., Shanjun Li, Richard G. Newell, and Karen Palmer, (2011). *Cost Effectiveness of Electricity Energy Efficiency Programs*. Resources for the Future Discussion Paper 09-48 Rev.~~

~~Auffhammer, Maximilian, Carl Blumstein, and Meredith Fowlie (2008). *Demand Side Management and Energy Efficiency Revisited*. The Energy Journal 29 (3), 91-103~~

~~Baumgartner, R. (2013). *The Uniform Methods Project: Methods for Determining Energy Efficiency Savings for Specific Measures*, Chapter 12: Survey Design and Implementation Cross-Cutting Protocols for Estimating Gross Savings. <http://www.nrel.gov/docs/fy13osti/53827.pdf>.~~

~~BC Hydro (2012). *Review of a Top-Down Evaluation Study: Rivers & Jaccard*. Prepared for BC Hydro, by Navigant Consulting, Inc., April.~~

~~Bradlow, E. (1998). *Encouragement Designs: An Approach to Self-Selected Samples in an Experimental Design*. Marketing Letters, Vol. 9, No. 4, November.~~

~~Cadmus Group, Inc. (2012). *CPUC Macro Consumption Metric Pilot Study: Final Report*. Prepared for the California Public Utilities Commission, October.~~

~~Cadmus Group, Inc. (2012). *Efficiency Maine Trust Residential Lighting Program Evaluation: Final Report*. Prepared for the Efficiency Maine Trust. <http://www.efficiencymaine.com/docs/Efficiency-Maine-Residential-Lighting-Program-Final-Report-FINAL.pdf>~~

~~Cadmus Group, Inc. (2013). *Focus on Energy Calendar Year 2012 Evaluation Report, Volume II and Appendices A through O*. Prepared for the Public Service Commission of Wisconsin. http://www.focusonenergy.com/sites/default/files/FOC_XC_CY%2012%20Report%20Volume%20II%20Final_08-28-2013.pdf~~

DRAFT #4 | Not for Citation

~~Cadmus Group and Navigant (2012). *New York Energy SmartSM Products Program Market Characterization and Assessment Evaluation: Final Report*. Prepared for The New York State Energy Research and Development Authority, Victoria Engel-Fowles Project Manager, Project Number 9875, February. http://www.nyserda.ny.gov/BusinessAreas/Energy_Data_and_Prices_Planning_and_Policy/Program_Evaluation/NYES_Evaluation_Contractor_Reports/2012-Reports/Market_Analysis.aspx~~

~~Cadmus Group, Navigant, Opinion Dynamics Corporation (2012). *2012 Residential Heating, Water Heating, and Cooling Equipment Evaluation: Net-to-Gross, Market Effects, and Equipment Replacement Timing*. Prepared for the Electric and Gas Program Administrators of Massachusetts.~~

~~Castor, Sarah (2012). *Fast Feedback Results. 2011 Final Report* prepared for Energy Trust of Oregon. http://energytrust.org/library/reports/Fast_Feedback_20110.pdf~~

~~Commonwealth Edison Company (2012). *Home Energy Reports Evaluation*. Prepared by Navigant. <http://www.ice.illinois.gov/downloads/public/edocket/323839.pdf>~~

~~Cook, T. et. al. (2010). *Contemporary Thinking About Causation in Evaluation*, American Journal of Evaluation, 31:105. <http://aje.sagepub.com/content/31/1/105>~~

~~Demand Research, LLC, (2012). *Macro Consumption Metrics Pilot Study: Final Report*. Prepared for: California Public Utilities Commission Energy Division, November.~~

~~Diamond, A. and J. Haninmueller (2007). *The Encouragement Design for Program Evaluation*. Harvard University and International Finance Corporation. http://www.doestoc.com/docs/5419170/The_Encouragement_Design_for_Program_Evaluation_September_Alexis_Diamond~~

~~EPRI (1991). *Impact Evaluation of Demand Side Management Programs: A Guide to Current Practice*. EPRI CU 7179, February.~~

~~Energy Valuation Organization, International Performance Measurement and Verification Protocols (IPMVP) (January 2012). *Concepts and Options for Determining Water and Energy Savings*, Vol. 1.~~

~~Eto, J., (1988). "On Using Degree-days to Account for the Effects of Weather on Annual Energy Use in Office Buildings," *Energy and Buildings*, vol. 12, p. 113—127~~

~~Eto, Joe, Ralph Prahl, and Jeff Schlegal (1996). *A Scoping Study on Energy efficiency Market Transformation by California Utility DSM Programs*. Lawrence Berkeley National Laboratory. <http://eetd.lbl.gov/ea/ems/reports/39058.pdf>.~~

~~Fagan, Jennifer, Mike Messenger, Mike Rufo, Peter Lai (2009). *A Meta-Analysis of Net to Gross Estimates in California*. Paper presented at the 2009 AESP conference.~~

~~GDS Associates, Inc., Nexant, and Mondre Energy (2010). "Net Savings: An Overview." RFP 2009-1 prepared for the Statewide Evaluator.~~

~~Feng, W. (2006). *A Method/Macro Based on Propensity Score and Mahalanobis Distance to Reduce Bias in Treatment Comparison in Observational Study*. Public Health Research, paper pr05. <http://www.lexjansen.com/pharmasug/2006/publichealthresearch/pr05.pdf>~~

~~Fowlie, M and C. Wolfram (2009). *Evaluating the Federal Weatherization Assistance Program using a Randomized Encouragement Design (RED)*. presented to Environmental Energy Technologies Division (EETD), Lawrence Berkeley National Laboratory, September. http://eetd.lbl.gov/sites/all/files/lbl_09-11-09.pdf~~

~~Fowlie, M and C. Wolfram, (undated). *An Experimental Evaluation of the Federal Weatherization Assistance Program*. Presentation prepared for the Michigan Public Service Commission. http://www.dleg.state.mi.us/mpsc/electric/workgroups/lowincome/fowlie_wolfram.pdf~~

~~Fowlie, M and C. Wolfram, (undated). *Randomized Encouragement Design Analysis of the DOE Weatherization Assistance Program*. http://iber.berkeley.edu/research/researchabstracts/Wolfram_RandomizedEncouragement.pdf~~

~~GDS Associates, Inc. (2012). *Analysis of Proposed Department of Energy Evaluation, Measurement and Verification Protocols*, sponsored by the National Rural Electric Cooperative Association. <https://www.nreca.coop/wp-content/uploads/2013/12/EMVReportAugust2012.pdf>~~

~~GDS Associates, Inc., Nexant, and Mondre Energy. (2010). *Net Savings: An Overview*. RFP 2009 prepared for the Statewide Evaluator.~~

~~Greene, W., (2011). *Econometric Analysis*, 7th Ed., Prentice Hall.~~

~~Haeri, Hossein. (2013). *Uniform Methods Project: Methods for Determining Energy Efficiency Savings for Specific Measures*, Chapter 1: The Introduction. <http://www.nrel.gov/docs/fy13osti/53827.pdf>.~~

~~Haeri, Hossein, and M. Sami Khawaja (2012). *The Trouble with Freeriders*. Public Utilities Fortnightly. <http://www.cadmusgroup.com/wp-content/uploads/2012/11/Haeri-Khawaja-PUF-TroublewithFreeriders.pdf>.~~

~~Hall, N. et al. (2013). *Setting Net Energy Impact Baselines: Building Reliable Evaluation Approaches*. Paper presented at the 2013 International Energy Program Evaluation Conference, Chicago, Illinois.~~

~~Ho, D. et al. (2007). *Matching as Nonparametric Preprocessing for Reducing Model Dependence in Parametric Causal Inference*. *Policy Analysis* 15(3); pp. 199-236.~~

~~Imbens, G. and Lemicux, T. (2010). *Regression Discontinuity Designs: A Guide to Practice*. Journal of Economic Literature 48, 281-355.~~

~~Itron, Inc. (2010). 2006-2008 Evaluation Report for PG&E Fabrication, Process and Manufacturing Contract Group. Prepared for the California Public Utilities Commission, Energy Division. http://www.calmac.org/publications/PG%26E_Fab_06-08_Eval_Final_Report.pdf http://www.calmac.org/publications/PG%26E_Fab_06-08_Eval_Final_Report_Appendices.pdf~~

~~Keating, Ken (2009). *Freeridership Borscht: Don't Salt the Soup*. Paper presented at the 2009 International Energy Program Evaluation Conference.~~

~~KEMA (2010). *Final Evaluation Report: Upstream Lighting Program, Volume 1*. CALMAC Study ID: CPU0015.01; Prepared for: California Public Utilities Commission, Energy Division. Prepared by: KEMA, Inc., Prime Contractor: The Cadmus Group, Inc. http://www.calmac.org/publications/FinalUpstreamLightingEvaluationReport_VolI-CALMAC-3.pdf~~

~~Keneipp, M. et al. (2011). *Getting MIF'ed: Accounting for Market Effects in Residential New Construction Programs*. Paper presented at the 2011 International Energy Program Evaluation Conference, Boston, Massachusetts.~~

~~Kennedy, P. (2008). *A Guide to Econometrics*, 6th Edition. Wiley-Blackwell, April.~~

~~KEMA, NMR Group Inc., Itron, Inc., The Cadmus Group, Inc. (2010). *Residential New Construction (Single Family Home) Market Effects Study*. Phase II Report prepared for California Public Utilities Commission Energy Division. Study ID CPUC0051.01. http://calmac.org/publications/RNC_mkt_effects_Phase_2_report_final_120610-ID.pdf~~

~~Khawaja, M. Sami, Josh Rushton, Josh Keeling. (2013). *The Uniform Methods Project: Methods for Determining Energy Efficiency Savings for Specific Measures*, Chapter 11: "Sample Design Cross-Cutting Protocols. <http://www.nrel.gov/docs/fy13osti/53827.pdf>.~~

~~Mahalanobis, P. (1936). "On the generalised distance in statistics". *Proceedings of the National Institute of Sciences of India* 2 (1): 49-55. http://www.new.dli.ernet.in/rawdataupload/upload/insa/INSA_1/20006193_49.pdf~~

~~McMenamin, J.S., (2008). *Defining Normal Weather for Energy and Peak Normalization*. Itron, Inc. <https://www.itron.com/PublishedContent/Defining%20Normal%20Weather%20for%20Energy%20and%20Peak%20Normalization.pdf>~~

~~Megdal, L. et al. (2009). *Feasting at the Ultimate Enhanced Freeridership Salad Bar*. Paper presented at the International Energy Program Evaluation Conference, Portland, Oregon.~~

~~<http://www.anevaluation.com/pubs/Salad%20Bar%202009%20IEPEC%20paper%205-12-09.pdf>~~.

~~Messenger, Mike, Ranjet Bharvirkar, Bill Golemboski, Charles Goldman, and Steven Schiller. (2010). *Review of Evaluation, Measurement and Verification Approaches Used to Estimate the Load Impacts and Effectiveness of Energy Efficiency Programs*. Lawrence Berkeley National Laboratory. <http://emp.lbl.gov/sites/all/files/lbnl-3277e.pdf>~~

~~Mort, Dan. (2013). *The Uniform Methods Project: Methods for Determining Energy Efficiency Savings for Specific Measures*, Chapter 9: Metering Cross-Cutting Protocols. <http://www.nrel.gov/docs/fy13osti/53827.pdf>~~

~~Navigant (2012). *A Sampling Methodology for Custom C&I Programs*. Prepared by D. Violette, D., and B. Rogers for the Ontario Natural Gas Technical Evaluation Committee, Ontario Energy Board, November.~~

~~Navigant (2013). *Custom Free Ridership and Participant Spillover Jurisdictional Review*. Prepared for the Sub-Committee of the Ontario Technical Evaluation Committee. <http://www.ontarioenergyboard.ca/documents/TEC/Evaluation%20Studies%20and%20Other%20Reports/Ontario%20NTG%20Jurisdictional%20Review%20-%20Final%20Report.pdf>.~~

~~New York Department of Public Service (July 31, 2013). *Guidelines for Estimating Net-To-Gross Ratios Using the Self-Report Approach*. Appendix H.~~

~~New York Department of Public Service (July 31, 2013). *Guidelines for Calculating the Relative Precision of Program Net Savings Estimates*. Appendix I.~~

~~New York Department of Public Service (November 2012). *Evaluation Plan Guidance for EEPS Program Administrators*, Update #3, Appendix F. Albany, New York.~~

~~Navigant Consulting (May, 2013). *Custom Free Ridership and Participant Spillover Jurisdictional Review*. Prepared for the Sub-Committee of the Ontario Technical Evaluation Committee.~~

~~NEEP (2012). *Regional Net Savings Research, Phase 2: Definitions and Treatment of Net and Gross Savings in Energy and Environmental Policy*, submitted to the Northeast Energy Efficiency Partnerships: Evaluation, Measurement, and Verification Forum, by NMR Group and Research Into Action, December. <https://www.neep.org/Assets/uploads/files/emv/NEEP%20-%20Regional%20Net%20Savings%20Report%2012-05-12.pdf>~~

~~NMR Group, Inc., KEMA, Cadmus Group, Inc., Tetra Tech (2011). *Massachusetts ENERGY STAR Lighting Program: 2010 Annual Report*. Final Report prepared for Energy Efficiency Advisory Council Consultants, Cape Light Compact, NSTAR, National Grid, Unitil, and~~

DRAFT #4 | Not for Citation

~~Western Massachusetts Electric.~~

~~<https://www.efis.psc.mo.gov/mpsc/commoncomponents/viewdocument.asp?DocId=935690223>.~~

~~NMR Group, Inc. and Research Into Action, Inc. (2010). *Net Savings Scoping Paper*. Revised Draft prepared for Northeast Energy Efficiency Partnerships: Evaluation, Measurement, and Verification Forum. http://www.necp.org/Assets/uploads/files/emv/emv-products/FINAL_Net_Savings_Scoping_Paper_11-13-10.pdf.~~

~~NMR Group, Inc. and Tetra Tech (2011). *Cross-Cutting Net to Gross Methodology Study for Residential Programs—Suggested Approaches*. Final report prepared for the Massachusetts Program Administrators. http://www.ma-eeac.org/Docs/8.1_EMV%20Page/2011/2011%20Residential%20Studies/Residential%20MA%20NTG%20Methods%20Final%20072011.pdf~~

~~Oak Ridge National Laboratories (1991). *Handbook to DSM Program Evaluation*. Eric Hirst and John Reed, eds., NTIS Pubs., Washington, DC, # ORNL/CON-336, December.~~

~~Parfomak, P. and L. Lave (1996). *How Many Kilowatts Are in a Negawatt? Verifying the Ex-Post Estimates of Utility Conservation Impacts at a Regional Level*. *Energy Journal* 17 (4).~~

~~Peters, Jane, Marjorie McRae (2008). *Freeridership Measurement Is Out of Sync with Program Logic...or, We've Got the Structure Built, but What's Its Foundation*. In *Proceedings of the 2008 ACEEE Summer Study on Energy Efficiency in Buildings*, Washington, DC. http://www.aceee.org/files/proceedings/2008/data/papers/5_491.pdf.~~

~~Prahl, R. et al. (2013). *The Estimation of Spillover: EM&V's Orphan Gets a Home*. Proceedings of the 2013 International Energy Program Evaluation Conference, August.~~

~~Provencher, B. and B. Glinsmann (2013). *Evaluation Report: Home Energy Reports—Plan Year 4*. Prepared for Commonwealth Edison Company. February.~~

~~Provencher et al. (2013). *Some Insights on Matching Methods in Estimating Energy Savings for an Opt-In, Behavioral-Based Energy Efficiency Program*. 2013 International Energy Program Evaluation Conference, Chicago.~~

~~Regional Technical Forum. (2012). *Guidelines for the Development and Maintenance of RTF Savings Estimation Methods*. NWCouncil, Released December 4. Web page: <http://rtf.nwcouncil.org/subcommittees/deemed/> and document link: [http://www.google.co.uk/url?sa=t&rct=j&q=&esrc=s&frm=1&source=web&cd=3&ved=0CDwQFjAC&url=http%3A%2F%2Frtf.nwcouncil.org%2Fsubcommittees%2Fdeemed%2FDraftForReview_Guidelines%2520for%2520RTF%2520Savings%2520Estimation%2520Methods%2520\(12-04-2012\).docx&ei=epYGUs-](http://www.google.co.uk/url?sa=t&rct=j&q=&esrc=s&frm=1&source=web&cd=3&ved=0CDwQFjAC&url=http%3A%2F%2Frtf.nwcouncil.org%2Fsubcommittees%2Fdeemed%2FDraftForReview_Guidelines%2520for%2520RTF%2520Savings%2520Estimation%2520Methods%2520(12-04-2012).docx&ei=epYGUs-)~~

[SGOxy0gXo6YDAAg&usg=AFQjCNHDq5kHbCPklBvFfoobTohbVCoupA&sig2=Jh63EbYf2eFEmfzQ0nf9fw&bvm=bv.50500085,d.d2k](https://www.seeaction.org/SGOxy0gXo6YDAAg&usg=AFQjCNHDq5kHbCPklBvFfoobTohbVCoupA&sig2=Jh63EbYf2eFEmfzQ0nf9fw&bvm=bv.50500085,d.d2k)

Ridge, Richard, Ken Keating, Lori Megdal, and Nick Hall (2007). *Guidelines for Estimating Net To Gross Ratios Using the Self Report Approaches*. Prepared for the California Public Utilities Commission.

Ridge, R. (1997). *Errors in Variables: A Close Encounter if the Third Kind*. Proceedings of the 1997 International Energy Program Evaluation Conference. August, Chicago.

Ridge, R. et al. (2009). *The Origins of the Misunderstood and Occasionally Maligned Self Report Approach to Estimating Net to Gross Ratio*. Paper presented at the 2009 Energy Program Evaluation Conference, Portland, Oregon.

Ridge, R. et al. (2013). *Gross Is Gross and Net Is Net: Simple, Right?*. Paper presented at the 2013 International Energy Program Evaluation Conference, Chicago, Illinois

~~Rosenberg, M., and Hoefgen, L. (2009). *Market Effects and Market Transformation: Their Role in Energy Efficiency Program Design and Evaluation*. Prepared for the California Institute for Energy and the Environment and the California Public Utilities Commission Energy Division.~~

~~Rufo, Michael (2009). *Evaluation and Performance Incentives: Seeking Paths to (Relatively) Peaceful Coexistence*, Proceedings, the International Energy Program Evaluation Conference, Portland, Oregon, August.~~

~~Sacramento Municipal Utilities District (SMUD) (2011). *Evaluation Report: OPOWER SMUD Pilot Year 2*. Prepared by Navigant, February.~~
~~http://www.opower.com/uploads/library/file/6/opower_smud_yr2_eval_report_final1.pdf~~

~~Sebold, Frederick D, Alan Fields, Lisa Skumatz, Shel Feldman, Miriam Goldberg, Kenneth Keating and Jane Peters (2001). *A Framework for Planning and Assessing Publicly Funded Energy Efficiency*. <http://www.calmac.org/events/20010301PGE0023ME.pdf>;~~

~~SEE Action (2012a). *Evaluation, Measurement, and Verification (EM&V) of Residential Behavior Based Energy Efficiency Programs: Issues and Recommendations*. Prepared by A. Todd, E. Stuart, S. Schiller, and C. Goldman, Lawrence Berkley National Laboratory. <http://behavioranalytics.lbl.gov/reports/behavior-based-emv.pdf>. (2012 a).~~

~~SEE Action (2012b). *Energy Efficiency Program Impact Evaluation Guide*. Prepared by Steven R. Schiller, Schiller Consulting, Inc. http://www1.eere.energy.gov/seeaction/pdfs/emv_ee_program_impact_guide.pdf. (2012 b).~~

~~SMUD (2013). *Load Impact Results from SMUD's Smart Pricing Options Pilot*. Prepared by Freeman Sullivan & Co. for Sacramento Municipal Utility District — SMUD contact Ms. Lupe Jimenez.~~

~~Southern California Edison (2012). Edison SmartConnect[®] Demand Response and Energy Conservation Annual Report, prepared by David Hanna et al., Itron, Inc. for Eric Bell, SCE project manager.
https://www.pge.com/regulation/DemandResponseOIR/Pleadings/SCE/2012/DemandResponseOIR_Plea_SCE_20120430_237124.pdf~~

~~Stuart, Elizabeth A. (2010). Matching Methods for Causal Inference: A Review and a Look Forward, 25(1); pp. 1-21. Statistical Science.~~

~~TecMarket Works et al. (2012). Indiana Evaluation Framework. Prepared for the Indiana Demand Side Management Coordination Committee.~~

~~Tetra Tech, Inc., KEMA, NMR Group, Inc. (2011). Cross-Cutting (C&I) Free Ridership and Spillover Methodology Study Final Report. Massachusetts Program Administrators. <http://www.ma-eeac.org/Does/8.1-EMV%20Page/2011/2011%20Commercial%20&%20Industrial%20Studies/MA%20FR-SO%20CI%20%20Study%20w%20Exec%20Summary%205-26-2011%20v11.pdf>~~

~~The Nonresidential Net To Gross Ratio Working Group (October, 2012). Methodological Framework for Using the Self-Report Approach to Estimating Net to Gross Ratios for Nonresidential Customers. Prepared for the Energy Division, California Public Utilities Commission.~~

~~Titus, E. and Michals, J. (2008). Debating Net Versus Gross Impacts in the Northeast: Policy and Program Perspectives. ACEEE Summer Study on Energy Efficiency in Buildings, (5); pp. 312-323. <https://www.aceee.org/files/proceedings/2008/data/papers/5-429.pdf>~~

~~U.S. Department of Energy (2010). Guidance Document #7: Topic: Design and Implementation of Program Evaluations that utilize Randomized Experimental Approaches". Smart Grid Investment Grant Technical Advisory Group, November.
http://www.smartgrid.gov/sites/default/files/pdfs/cbs_guidance_doc_7_randomized_experimental_approaches.pdf~~

~~Violette et al. (1991). Impact Evaluation of Demand Side Management Programs—Volume 1: A Guide to Current Practice. Electric Power Research Institute Pubs., Palo Alto, CA, #EPRI CU-7179, February.~~

~~Violette, D. et al. (1993). Statistically Adjusted Engineering Estimates: What Can The Evaluation Analyst Do About The Engineering Side Of The Analysis?. Published in the Proceedings of the 1993 Energy Program Evaluation Conference.~~

~~Violette, D. et al. (2005). Commercial/Industrial Performance Program (CIPP)~~

~~Market Characterization, Market Assessment And Causality Evaluation. Prepared for NYSERDA, Jennifer Ellefsen, Project Number 7721, March.~~

~~Violette, D and B. Provencher (2012). *Review of a Top Down Evaluation Study: Rivers & Jaccard (2011)*. Prepared for BC Hydro, Navigant Consulting, Inc., April.~~

~~Violette, D. M., Provencher, B., and Sulyma, I. (2012). *Assessing Bottom Up and Top Down Approaches for Assessing DSM Programs and Efforts*. International Energy Program Evaluation Conference Proceedings, Rome, June.~~

~~Violette, D. (2013). *Uniform Methods Project for Determining Energy Efficiency Program Savings for Specific Measures*, Chapter 13: "Persistence and Other Evaluation Issues Cross-Cutting Protocols", **Subcontract Report**, NREL/SR-7A30-53827, April. See: http://www1.eere.energy.gov/wip/pdfs/53827_13.pdf~~

~~West, S, (2008). *Alternatives to the Randomized Controlled Trial*. American Journal of Public Health. Vol. 98, No. 8. <http://ajph.aphapublications.org/doi/abs/10.2105/AJPH.2007.124446>~~
~~Wooldridge, J. M. (2002). *Econometric Analysis of Cross Section and Panel Data*. Cambridge, MA: MIT Press.~~

~~Wooldridge, J.M. (2010). *Econometric Analysis of Cross Section and Panel Data*. MIT Press Ltd..~~