



Not adding up: free ridership and spillover calculations in energy efficiency evaluations

Zachary Froio · Pranay Kumar  · Frank A. Felder

Received: 28 June 2019 / Accepted: 19 May 2020 / Published online: 5 June 2020
© Springer Nature B.V. 2020

Abstract A key element of evaluation, measurement, and verification (EM&V) studies for energy efficiency programs involves estimation of net energy savings that account for free ridership, spillover, and induced market effects. The existing literature recognizes these effects to be significant and provides detailed guidelines to estimate them. However, there appears to be a disconnect between these guidelines and field evaluations conducted in practice. Our meta-analysis of 120 studies from 2006 to 2018 indicates that most free ridership and spillover estimates are based on survey results and expressed in percentage terms. We note that simply adding these percentages numerically without converting them into a common unit is inaccurate and

obscures a program's true impact. Additionally, there exists wide variations in nomenclature, classification, and methodologies adopted to estimate these metrics across programs and jurisdictions. Our scatterplot analysis of the reviewed EM&V reports indicates that with few exceptions, free ridership and spillover do not necessarily offset each other. We propose an alternative approach to estimate free ridership and spillover in energy units with costs in dollar terms, e.g., as the difference between a program participant's total willingness-to-pay and the total financial impact of the program's existence. We also feel that a consistent, transparent, and reliable evaluation methodology to estimate free ridership and spillover effects across programs and jurisdictions based on randomized or quasi-experimental designs will not only improve accuracy but will also have better comparability for informed policy decisions in future.

Keywords Energy efficiency · Evaluation, measurement, and verification (EM&V) · Program design · Free ridership · Spillover · Market effects

Introduction

Often considered to be the lowest-cost energy resource available, many government organizations, regulators, and utilities alike support the establishment and expansion of energy efficiency (EE) programs as a “first fuel” of choice among competing alternatives to accommodate energy demand (Friedrich et al. 2009). Energy

Z. Froio
Applied Energy Group, 211 Broad Street, Ste. 206, Red Bank,
New Jersey, 07701, USA
e-mail: zachary.froio@rutgers.edu

P. Kumar (✉) · F. A. Felder
E.J.Bloustein School of Planning and Public Policy, Rutgers
University, 33, Livingston Avenue, New Brunswick, NJ 08901,
USA
e-mail: pranay.kumar@rutgers.edu

F. A. Felder
e-mail: ffelder@ejb.rutgers.edu

F. A. Felder
Center for Energy, Economic & Environmental Policy, Public
Informatics Program, Rutgers Energy Institute, Rutgers
University, 33, Livingston Avenue, New Brunswick, NJ 08901,
USA

efficiency policy options are considered win-win solutions not only in terms of energy resource planning but also as an important greenhouse gas emission reduction strategy. Not surprisingly, the worldwide investment in energy efficiency was estimated to be \$231 billion in 2016 with the building sector accounting for 58% of total investment dollars (IEA 2017). In the USA, states have invested approximately \$8 billion in energy efficiency and have reportedly saved 27.1 million megawatt hours in 2018 (ACEEE Scorecard 2019).

A key element of effective evaluation, measurement, and verification (EM&V) studies for EE programs is the estimation of net energy savings, defined by the US Department of Energy (DOE) Uniform Methods Project (UMP) protocol document as “changes in energy use that are attributable to a particular EE program,” which may “implicitly or explicitly include the effects of free ridership, spillover, and induced market effects” (FR, SO, and ME, respectively). According to the UMP, free riders are “program participants who would have implemented a program measure or practice in the absence of program,” and these participants have been divided into total, partial, and/or deferred free riders. Furthermore, program spillover refers to “additional reductions in energy consumption or demand that are due to program influences beyond those directly associated with program participation” broadly categorized into two types: participant spillover (further subcategorized into inside, outside, like, and unlike spillover) and non-participant spillover. Similarly, market effects include “long-lasting, substantive changes in market structures or participant behavior that is reflective of an increased adoption of EE measures which is causally related to market interventions” (Li et al. 2018). Most evaluation protocols—including DOE UMP guidelines—provide for estimating net savings by deducting the sum of the estimated values of FR, SO, and ME (including negative values) from gross savings. Furthermore, the ratio of a program’s net savings to gross savings is commonly referred to as the net-to-gross (NTG) ratio. This calculation is often employed in a variation of the following series of equations as demonstrated by the UMP:

- (1.) Net savings = (gross savings) – (free ridership) + (spillover) + (market effects savings not already captured by SO)
- (2.) Net-to-gross ratio = $1 - (\text{free ridership})/(\text{gross savings}) + (\text{spillover})/(\text{gross savings}) + (\text{market effects})/(\text{gross savings})$

$$(3.) \text{ Net savings} = (\text{net-to-gross ratio}) * (\text{gross savings})$$

(Li et al. 2018).

However, the addition of FR, SO, and ME is problematic for two major reasons. First, this approach omits essential information regarding the effectiveness of a program’s design. The varying methodologies to estimate free ridership and spillover—as well as the inconsistent usage of both spillover and market effects within net savings calculations—inhibit the meaningful comparison of net savings across different programs. Furthermore, prescriptive changes to these program designs—for example, to reduce free ridership or to increase spillover—will differ between programs with varying degrees of these effects, regardless of the reported net savings estimations for each respective program. Second, when evaluations express FR, SO, and ME in percentages (namely with respect to total program participants), the addition of these percentages often does not result in the same outcome if such calculations were performed in terms of units of energy, e.g., kilowatt-hours, given the inherent differences between consumption and saving potentials across participants. We deliberate upon these arguments that challenge the internal validity of the theoretical construct adopted for estimation of the net energy savings of EE programs. Furthermore, our review of the EM&V report brings out wide variations in nomenclature, classification, and methodologies adopted in practice for evaluation of EE programs across jurisdictions. We find a clear disconnect between available guidelines and actual field evaluations conducted in practice. As such, even if we overlook the inherent limitations of the measurement guidelines for a moment, the wider discrepancies and inconsistencies in the evaluation methods adopted in practice make the net savings values less accurate and difficult to compare across programs and jurisdictions.

The remainder of this paper explores these major findings in detail. The “[Overview of EM&V studies that estimate FR, SO, and ME to determine net savings](#)” section begins with an overview of EM&V studies that estimate FR, SO, and ME to determine net savings; the “[Adding FR and SO loses essential information about a program’s design](#)” section demonstrates how the addition of FR and SO loses important information about a program’s design; the “[Discussion and alternative approaches](#)” section discusses the implications of these findings and makes the case for alternative approaches

to estimate net savings; and the “[Conclusions](#)” section concludes with policy recommendations and areas for further research.

Overview of EM&V studies that estimate FR, SO, and ME to determine net savings

Estimation of net energy savings does not just present technical challenges associated with quality, effectiveness, and validity of EE programs but also have important long-term policy implications for different stakeholders and impact on GHG emissions at a global level (Vine et al. 2010). A comprehensive policy analysis would require assessing cost-benefit analysis from the perspectives of different stakeholders depending upon the policy context using standard tests (Felder and Athawale 2017). However, our current analysis is limited to the estimation of net energy savings for impact evaluation of EE programs from ratepayer’s perspective. Accordingly, we find that free ridership has important implications not just for the assessment of the cost-effectiveness of EE programs in terms of energy units saved per program dollar spent, but also from equity considerations involving potentially free-riding program participants and the non-participants who fund such programs. A fundamental and problematic issue in estimating free ridership is constructing a satisfactory baseline scenario, i.e., a counterfactual. In other words, to quantify the net cause-effect relationship between program participation and energy savings, it is important to isolate those participants who would have implemented the particular EE measure even without any intervention. Failure to account for such participants might overstate a program’s effects. Similarly, failure to assess spillover effects might result in biased estimates of program impacts and inappropriate policy recommendations (Angelucci and Maro 2015). For estimating net savings, it is also important to consider the program influence on all market actors over successive stages of market development. Evaluation of market transformation programs for EE product and services has significant policy implications and they focus more on markets rather than on programs (Vine et al. 2009). As such, estimating market effects requires careful considerations and more elaborate studies employing mix of qualitative and quantitative methods over a longer period of time (Violette and Rathbun 2017). However, current EE program evaluation practices are marked by varied terminologies and

inconsistent methodologies that are compounded by a lack of publicly available information to assess program analyses. For these reasons, the program evaluation field has been subjected to a long and continued debate eluding consensus (Jaccard 2010).

The American Council for an Energy-Efficient Economy (ACEEE) compiles and compares the energy saving performance of all large cities, states, and territories in the USA and ranks them based on their respective performance scorecards. An important metric of the ACEEE scorecard is the net annual incremental electricity savings as a percentage of annual state utility retail sales. Saving information is obtained from public service commission databases, and sales data are sourced from the US Energy Information Administration (EIA). From this report, it is observed that there is wide diversity among states in the usage of net versus gross energy savings for program evaluation purposes, with just 12 states reporting gross savings, 21 states reporting net savings, and another 9 states reporting both values (ACEEE Scorecard 2019, p. 30). States report energy savings as either net or gross, with net savings accounting for free riders, spillovers, and market effects, and gross savings not accounting for these metrics.

Our findings from the literature review of the EE program experience in the EU countries reveal similar pattern that is not much different from the USA. We observe that EM&V protocols, methodologies, and outcomes are not consistently applied across the European Union either. Recently, most EU countries have started estimating gross savings but only the United Kingdom and Denmark estimate net energy savings. Whereas the free rider effect (also referred to as deadweight) is factored into the estimation of net energy savings in a suppliers’ obligation program in the UK, the same is not accounted for in a building rehabilitation program in Germany (Rosenow and Ray 2012). The examination of the costs and benefits of a Norwegian energy conservation program that provided financial support to participants revealed that about 70% of those participants was free riders (Haugland 1996). In Italy, an incentive program for energy upgrades used a survey of 3,000 households to find the free rider effect up to 70% (Alberini and Bigano 2014). Furthermore, comprehensive EM&V protocols for EE projects are at various stages of development in China and India (Slote et al. 2014).

For this study, two separate datasets of comprehensive EM&V studies were compiled to conduct two distinct analyses: (1) a review of the methodologies

Table 1 List of 100 EM&V reports filed in the USA between 2007 and 2018, retrieved using the E Source *DSM Insights* tool and reviewed for net savings methodologies

State	Document description	Principal author
PA	Act 129 Statewide Evaluator Annual Report Program Year 1: June 1, 2009–May 31, 2010	GDS Associates, Inc., Nexant, Mondre Energy and Syntil, Inc.
AR	ADM - AEP-SWEPCO - 4.29.2016 - 2015 Weatherization Evaluation Report - 07-082-TF	ADM
NM	ADM - PNM - 3.1.2015 - 2014 Evaluation Report	ADM Associates, Inc
NM	ADM - PNM - 3.1.2016 - 2015 Evaluation Report	ADM Associates
NM	ADM - PNM - NM - 2013 EE Independent M V Part 1	ADM
NM	ADM Associates - 4.14.2017 - 2016 PNM Evaluation Report	ADM Associates
IN	ADM Associates - 8.26.2016 - 2015 C&I Evaluation Reports - 44841	ADM Associates
IN	ADM Associates - 8.26.2016 - 2015 Residential Evaluation Reports - 44841	ADM Associates
AR	AEP - SWEPCO - AR - 4.28.2017 - 2016 Evaluation Report - 07-082-TF	ADM Associates
OH	AEP Ohio - 5.15.2017 - 2016 Annual Report Part 2 - 17-1229-EL-EEC	AEP Ohio
OH	AEP Ohio - 5.15.2017 - 2016 Annual Report Part 3 - 17-1229-EL-EEC	AEP Ohio
AR	AEP-SWEPCO - 4.29.2016 - 2015 Portfolio Evaluation Report - 07-082-TF	AEP - SWEPCO
AZ	APS - 3.1.2018 - 2017 MER Reports - E-00000 U-18-0055	Arizona Public Service
AZ	Arizona Public Service - 3.1.2017 - Navigant Eval Report - E-00000 U-17-0057	Navigant
WA	Avista - WA - 5.25.2012 - PY2011 - Multi-Sector Gas Impact Evaluation Report - UE-100176	Cadmus
MO	Cadmus - 5.26.2016 - Revised 2015 EM&V Report of Ameren MO - E0-2012-0142	Cadmus
ID	Cadmus - RMP - ID - 10.23.2013 - see ya later, refrigerator evaluation report - 2011-2012	Cadmus Group
ID	Cadmus - RMP - ID - 2.3.2012 - Home Energy Savings Evaluation Report 2009-2010	Cadmus Group
ID	Cadmus - RMP - ID - 2.3.2012 - See ya later, refrigerator Evaluation Report - 2009-2010	Cadmus Group
WY	CADMUS - RMP - WY - 1.21.2014 - 2011-2012 HES Eval Report	Cadmus
LA	Cleco - 3.1.2017 - PY2 ADM Eval Report - R-31106	ADM Associates
IL	ComEd - 12.21.2010 - Navigant Evaluation Summary - PY2010 - 07-0540	Navigant Consulting
IL	ComEd - 12.23.2009 - Summit Blue Consulting Evaluation Summary - PY2009 - 07-0540	Summit Blue Consulting LLC
TX	CPS Energy - Measurement and Verification of CPS Energy's FY2013 DSM Program Offerings - 5.22.2013	Nexant
VA	Dominion - 10.3.2011 - DSM Program Evaluation, Measurement, Verification Reports - KEMA - PUE-2010-00084	KEMA
NC	Dominion - 4.1.2013 - KEMA - EM&V Report - Volume 1 - E-22 Sub 486 - PUE-2011-00093	KEMA
VA	Dominion - 4.2.2012 - DSM Program Evaluation, Measurement, Verification Reports - KEMA - PUE-2010-00084	KEMA
NC	Dominion - NC - 4.1.2013 - KEMA - EM&V Report - Volume 1 - E-22 Sub 486	KEMA
NC	Dominion - NC - 4.1.2013 - KEMA - EM&V Report - Volume 2 - E-22 Sub 486	KEMA
OH	Duke Energy - Ohio - 3.14.2016 - Appendices E-F - 16-0513-EL-EEC	Duke Energy Ohio
OH	Duke Energy - Ohio - 3.14.2016 - Appendices G-K - 16-0513-EL-EEC	Duke Energy Ohio

Table 1 (continued)

State	Document description	Principal author
OH	Duke Energy - Ohio - 3.14.2016 - Appendices K-O - 16-0513-EL-EEC	Duke Energy Ohio
OH	Duke Energy - Ohio - 3.14.2016 - Appendices O-Q - 16-0513-EL-EEC	Duke Energy Ohio
OH	Duke Energy - Ohio - 3.14.2016 - Appendix K Part 2 of 3 - 16-0513-EL-EEC	Duke Energy Ohio
NC	Duke Energy Carolinas - 3.8.2017 - Evaluation Reports Accompanying 2017 Filing - E-7 Sub 1130	Multiple Authors
LA	Energy LA - 3.1.2016 - PY1 ADM Evaluation Report - R-31106	ADM Associates
AR	Energy - AR - 4.18.2017 - 2016 Evaluation Report - 07-085-TF	Tetra Tech
LA	Energy Gulf States LA - 3.1.2016 - PY1 ADM Evaluation Report - R-31106	ADM Associates
LA	Energy Gulf States LA - 3.1.2017 - PY2 ADM Evaluation Report - R-31106	ADM Associates
LA	Energy LA - 3.1.2017 - PY2 ADM Evaluation Report - R-31106	ADM Associates
WI	Focus On Energy - 5.15.2014 - PY 2013 Evaluation Report	Cadmus Group
WI	Focus on Energy - 5.20.2016 - 2015 Evaluation Report Vol 1 & 2	Cadmus
WI	Focus on Energy - 5.27.2015 - 2014 Evaluation Report	Cadmus
WI	Focus on Energy - 5.27.2015 - 2014 Evaluation Report Part 2 and Appendix	Cadmus
WI	Focus On Energy - 8.28.2013 - 2012 Evaluation Report Volume I and II and Appendices	Cadmus Group
AR	KEMA - 3.4.2013 - 2012 Evaluation Report - 07-085-Tf_383_4	KEMA
NY	LIPA Efficiency Long Island 2010 Annual Report - Volume II - 4.1.2011	Opinion Dynamics
NY	LIPA Efficiency Long Island 2011 Annual Evaluation Report - 5.18.2012	Opinion Dynamics
NY	LIPA Efficiency Long Island and Renewables Portfolio 2012 Program Guidance Document - 5.28.2012	Opinion Dynamics
NY	LIPA Efficiency Long Island PY2009 Assessment - Volume II - 5.1.2010	Opinion Dynamics
IA	MidAmerican - 10.11.2017 - Agriculture Evaluation	Tetra Tech
IA	MidAmerican - 10.11.2017 - Residential New Construction Evaluation	Tetra Tech
IA	MidAmerican - 10.27.2017 - Commercial Solutions Evaluation	Tetra Tech
IA	MidAmerican - 11.16.2017 - Industrial Partners Evaluation	Tetra Tech
IA	MidAmerican - IA - 11.15.2017 - Residential HomeCheck Evaluation	Tetra Tech
IL	Navigant - 11.2.2016 - ComEd New Construction Small Buildings PY8 Evaluation Report DRAFT - ILSAG	Navigant
MO	Navigant - 12.6.2017 - EM&V Report Appendices of KCP&L GMO PY2016 - EO-2015-0241	Navigant
MO	Navigant - 12.6.2017 - EM&V Report Appendix of KCP&L PY 2016 - EO-2015-0240	Navigant
MO	Navigant - 12.6.2017 - EM&V Report of KCP&L GMO PY2016 - EO-2015-0241	Navigant
MO	Navigant - 12.6.2017 - EM&V Report of KCP&L PY 2016 - EO-2015-0240	Navigant
ID	Navigant - RMP - ID - 2.15.2013 - Energy FinAnswer Evaluation Report - 2009-2011	Navigant
ID	Navigant - RMP - ID - 3.17.2015 - Energy FinAnswer 2012-2013 Eval	Navigant
ID	Navigant - RMP - ID - 3.19.2015 - FinAnswer Express 2012-2013 Eval	Navigant

Table 1 (continued)

State	Document description	Principal author
ID	Navigant - RMP - ID - 3.4.2013 - FinAnswer Express Evaluation Report - 2009-2011	Navigant
ID	Navigant - RMP - ID - 5.10.2013 - Irrigation Energy Savers Evaluation Report - 2009-2011	Navigant
NV	Nevada Power - 6.30.17 - MV Reports - 1 of 3 - 17-06043	Nevada Power
NV	Nevada Power - 6.30.17 - MV Reports - 2 of 3 - 17-06043	Nevada Power
NV	Nevada Power - 6.30.17 - MV Reports - 3 of 3 - 17-06043	Nevada Power
MT	NWE - 2007-2011 Evaluation Report	Northwestern Energy
AR	OGE - AR - 4.1.2017 - 2016 Evaluation Report - 07-075-TF	Tetra Tech and ADM Associates
HI	Opinion Dynamics - 1.18.2018 - EM&V Report of Hawaii Energy PY2016	Opinion Dynamics
HI	Opinion Dynamics - 11.28.2016 - EM&V Report of Hawaii Energy PY2015	Opinion Dynamics
HI	Opinion Dynamics - 3.16.2016 - EM&V Report of Hawaii Energy PY2014	Opinion Dynamics
PA	Pennsylvania Act 129 Statewide Evaluator Annual Report Program Year 2: June 1, 2010 – May 31, 2011	GDS Associates, Inc., Nexant, and Mondre Energy
NM	PNM - NM - 3.1.2012 - Evaluation Report Part 1 - PY 2011	ADM Associates
NM	PNM - NM - 3.16.2010 - MandV Repot PY 2009 - 2 of 2	KEMA
NM	PNM - NM - 5.1.2011 - Evaluation Report Part 1 - PY 2010	ADM Associates
NM	PNM - NM - 6.1.2010 - MandV Report PY 2009 - 1 of 2	ADM Associates
NC	Progress Energy Carolinas - NC - 1.26.2016 - Small Business Energy Saver Program 2014 EMV Report - E-2 Sub 1022	Navigant
NC	Progress Energy Carolinas - NC - 12.16.2014 - Small Business Energy Saver Program 2013 EMV Report - E-2 Sub 1022	Navigant
NY	PSEG LI - 5.30.2017 - 2016 Annual Eval Report Vol 2	Opinion Dynamics
NY	PSEG LI - 7.12.2016 - 2015 Annual Eval Report Vol 2	Opinion Dynamics
UT	RMP - UT - 10.23.2013 See ya later Refrigerator Program Evaluation Report 2011-2012	Cadmus Group
UT	Rocky Mountain Power - 10.25.2013 - 2011-2012 Cool Cash Program Impact Evaluation Report	Rocky Mountain Power
ID	S&L - RMP - ID - 1.26.2015 - Idaho LI Weatherization 2010-2012 Eval	Smith & Lehmann Consulting
SC	SCE&G - 5.29.2015 - PY 4 EMV Report - 2013-208-E	Opinion Dynamics
SC	South Carolina Electric & Gas - 5.30.2014 - PY 2013 EM&V Report - 2013-208-E	Opinion Dynamics
SC	South Carolina Electric & Gas - 5.31.2012 - Evaluation Report - PY 2011 - 2012-55-E	Opinion Dynamics
SC	South Carolina Electric & Gas - 5.31.2013 - Evaluation Report - PY 2012 - 2013-50-E	Opinion Dynamics
AZ	SRP - 8.1.2009 - PY 2009 Evaluation Report	SRP
AZ	SRP - 8.1.2010 - FY2010 Energy Efficiency Portfolio Evaluation Summary	Cadmus Group
IN	Statewide Indiana - 6.23.2013 - Core Programs Evaluation Report	DSMCC
CO	Tetra Tech - 3.30.2018 - 2017 EM&V Report of Black Hills CO - 15A-0424E	Tetra Tech
CO	Tetra Tech - 3.31.2017 - 2016 EMV Report of Black Hills CO - 15A-0424E	Tetra Tech

Table 1 (continued)

State	Document description	Principal author
CO	Tetra Tech - 4.1.2014 - 2012-3 EMV Report of Black Hills CO - 12A-100E	Tetra Tech
CO	Tetra Tech - 4.1.2016 - 2015 EMV Report of Black Hills CO - 12A-100E	Tetra Tech
IA	Tetra Tech - 5.1.2017 - 2016 Eval of BHE - EEP-2013-0001	Tetra Tech
WI	WI FOE - 4.19.2017 - Evaluation Report Appendices	Cadmus
WI	WI FOE - 4.19.2017 - Evaluation Report Vol 1	Cadmus
WI	WI FOE - 4.19.2017 - Evaluation Report Vol 2	Cadmus

and standard practices of program evaluations to arrive at net savings and (2) an examination of reported gross and net savings, NTG ratios, free ridership, spillover, and market effects estimations. The first analysis examines EM&V studies published from 2007 to 2018 using a set of criteria to understand the current state of practices for program evaluation methodologies in the USA (see Table 1 following the “[Conclusions](#)” section). These studies were obtained through the E Source *DSM Insights* tool, which contains a comprehensive catalog of documents and dockets for energy efficiency, load management, and demand response programs sourced from public utility commission filings across the country. For the purpose of this analysis, the *DSM Insights* search criteria included program evaluation report documents filed in any state in the USA between January 2005 and April 2018 for electric-only saving programs, resulting in a total of 100 evaluation reports in 24 states.

The vast majority of these 100 evaluations report gross savings, with 91.1% of those also reporting net savings estimations. Of the evaluations that report net savings, 89.3% directly report free ridership estimations, and all of which that detail their methodologies use survey techniques to estimate FR values. The evaluations that report free ridership do so in percentage terms of participants, energy savings, or both at the measure level (12.7%), the program level (10.1%), both (64.6%), or do not explicitly specify (12.7%). By comparison, only 66.7% of net saving evaluations estimates spillover effects. These studies report spillover in percentage terms of participants and/or energy savings at the measure level (13.1%), program level (16.4%), both (52.5%), or do not specify (18%). Like free ridership, all reports that detail their methodologies estimate spillover effects using survey techniques. A notable example is the Rocky Mountain Power Wyoming Home Energy Saving evaluation, which aggregated individual measures under six broad categories to estimate FR and SO with a statistically significant sample population. Of the net saving studies, less than 5% reported market effects, of which all estimated ME separately from spillover. In all, the majority of net saving studies—64.8%—weighted net savings by the percentage of energy savings.

The lack of ME estimations is noteworthy, and many reports acknowledge the need for further research to quantify these effects. Notable examples that assess market effects include Rocky Mountain Power Idaho’s

“See ya later, refrigerator” estimation of secondary market effects through participant and non-participant telephone surveys, Ameren Missouri’s Efficient Products estimation of market effects at 5% using a demand elasticity modeling method, and Wisconsin’s Focus on Energy estimation of ME at 0.9% for CFLs based on lighting saturations over a 5-year timeframe. Although repeated observations of target groups over time and across jurisdictions are considered helpful in establishing the persistence of program interventions, only 16.9% of the evaluations estimated FR, SO, or ME over time (Keay 2011). The majority of these reports either did not specify time intervals or did not appear to carry out analyses on a temporal basis. Certain obstacles—namely the resources required to obtain longitudinal data over a given timeframe—inhibit the engagement in such analyses. As a result, ME estimations within program net savings calculations are often based on tacit assumptions rather than comprehensive field results.

Of the 82 studies that report net savings, there is wide variation in the categories and terminology used to describe and thus estimate net saving components. Studies that estimate free ridership categorize the effect in terms of full-only (4.1%), full and partial (14.9%), full, partial, and deferred (9.5%), or otherwise do not explicitly specify (71.6%). Little consistency exists across the categorization of spillover effects as well, with studies estimating SO as like (4.1%), participant (29.7%), participant and non-participant (12.2%), or otherwise do not specify (51.4%). Only two studies—2.7% of those that report spillover—categorize all types: like, unlike, participant, and non-participant spillover.

The majority of the studies that estimate both FR and SO employ some variation of the NTG ratio saving calculation illustrated above, in which percentage estimates of free ridership and spillover are deducted and added, respectively, from gross saving estimates. However, the methodologies used to estimate FR and SO values vary across these reports. Most of the studies arrive at these estimates using a combination of yes/no responses to a series of participant survey questions. Of the 100 reviewed studies, none employed randomized controlled trials or quasi-experiments to estimate FR and SO. Others, however, use novel methods such as demand response modeling, trade ally interviews, or standard deemed values for specific measures, e.g., from statewide technical reference manuals. For example, Wisconsin’s Focus on Energy program evaluation estimates FR and SO values for lighting measures using a

combination of four methods: demand elasticity modeling, national sales modeling, corporate retailer interviews, and manufacturer interviews. In the same program, net savings for select HVAC measures were estimated using the Standard Market Practice (SMP) approach. Under this method, the evaluation team first established the average market baseline consumption with available market data for equipment sold outside the program. Ignoring spillover, saving net of free ridership was estimated as the difference between the market baseline and the average energy consumed by the measures installed under the program. Other jurisdictions—in Indiana and Delaware, for example—also recognize net savings estimations that employ the SMP methodology. SMP approaches typically assume that the standard market values can act as the baseline for program evaluation avoiding the need to account for free ridership separately. However, this may or may not be the case depending upon the difference between the actual energy consumption levels of program participants with respect to the standard market baseline and therefore, needs to be tested further. (Ridge et al. 2013). Another example of non-survey based net saving estimations is CPS Energy’s Home Efficiency Program, which uses the Texas Public Utility Commission’s approved deemed values and engineering calculations for individual measures.

The second analysis for this report draws from 20 comprehensive, publicly available EM&V studies published on public utility commission websites or found through search engine results using key energy efficiency EM&V terms (see Table 2 following the “Conclusions” section). These studies comprise of programs offered from 2006 through 2015 and span across 11 states (California, Connecticut, Delaware, Illinois, Indiana, Maine, Maryland, Massachusetts, New Hampshire, New York, and Rhode Island). All study programs are further organized into subprograms by both industry (i.e., residential, commercial, and industrial) and the specific energy efficiency measures administered within each subprogram.

These reports consist of 83 total subprograms from which data was obtained for the following variables: reported and evaluated program gross savings, program net savings, subprogram and measure-level NTG ratios, free ridership, spillover, and market effects. Of these 83 subprograms, all reported their estimated NTG ratios and 75 estimated their reported and evaluated gross savings and the realization rate of program gross

Table 2 Summary of free ridership and spillover values from 20 publicly-available EM&V studies published in the USA between 2006 and 2015

State	Year	Program name	Sector	NTGR	FR	SO	Reports at measure level?
CA	2015	Upstream HVAC Programs	C, I	0.64	0.36	N/A	N
CA	2015	Multifamily Whole Building Program	R	0.446	0.554	N/A	N
CA	2015	Agricultural Low Pressure Sprinkler Nozzles, Drip Irrigation and Micro Conversion Sprinklers	C, I	0.47	0.53	N/A	N
CA	2015	Residential Lighting Programs	R	0.41 - 0.63	N/A	N/A	Y
CA	2015	Nonresidential Custom Lighting	C, I	0.46 - 0.54	0.46 - 0.54	N/A	N
CA	2015	Nonresidential Pipe Insulation	C, I	0.44	0.56	N/A	Y
CA	2015	Multifamily Energy Efficiency Rebate	R	0.55 - 0.85	0.15 - 0.45	N/A	N
CA	2015	Commercial Pool Covers	C, I	0.68	0.32	N/A	Y
CA	2015	Custom Industrial, Agricultural, and Large Commercial Programs	C, I	0.50 - 0.57	0.43 - 0.50	N/A	N
CT	2009	Residential Central AC	R	0.58 - 0.74	0.26 - 0.42	N/A	Y
DE	2011	DNREC SEU and State Energy Efficiency Programs	R	0.49 - 0.78	0.22 - 0.51	N/A*	Y
IL	2014	Commonwealth Edison Company Energy Efficiency / Demand Response	R, C, I	0.52 - 1.04	N/A	N/A	N
IL	2016	Ameren Illinois Company Electric & Gas Residential and Commercial and Industrial Energy Efficiency Portfolios	R, C	0.47 - 1.01	N/A	N/A	N
IN	2012	Energizing Indiana	R, C, I	0.57 - 1.10	0.00 - 0.43	N/A	N
MA	2009	Western Massachusetts Electric Company: Overall Small Business Services Program	C, I	0.74 - 1.053	0.036 - 0.257	0.000 - 0.089	Y
MA, NH, RI	2009	National Grid USA: Commercial and Industrial	C, I	0.82 - 0.965	0.057 - 0.217	0.022 - 0.074*	Y
MA, NH, RI	2008	EnergyWise 2008 Program	R	0.73 - 0.74	0.15 - 0.16	0.42	Y
MD	2013	EmPOWER Maryland Energy Efficiency Residential HVAC Program	R	0.36 - 0.4	0.63 - 0.67	0.03	N
ME	2014	Efficiency Maine Retail Lighting Program	R	0.347 - 1	0 - 0.653	0	Y
NY	2006-2009	NYSEERDA Program	C, I	1.16 - 1.28	0.31 - 0.66	0.59 - 0.82	N

*Of the 20 evaluated studies, the *National Grid USA: Commercial and Industrial* report remains the only to have incorporated market effects in its net savings calculations, which it considers to be “non-participant spillover effects”

savings. A total of 46 subprograms reported values for subprogram free ridership, while 21—approximately 25% of all compiled subprograms—reported both free ridership and spillover effects to arrive at their respective net-to-gross ratios. Of the 20 total reports, only one—the National Grid USA 2009 *Commercial and Industrial Programs Free-ridership and Spillover Study*—attempted to estimate program market effects, under what it considers to be “non-participant spillover effects” (Kraft et al. 2010). The remaining evaluations either mention but do not attempt to estimate ME or simply overlook these effects. Furthermore, only nine reports—less than half of the evaluated studies—reported net saving metrics at the measure level.

Using the FR and SO percentage estimates from the 20 reviewed studies, a scatterplot was drawn as shown in Fig. 1 above. For ease of analysis, the plot area was further subdivided into four quadrants by intersecting the two axes at 50% FR and SO values. One of the obvious observations from the scatterplot is that most programs reflect only free ridership percentage values overlooking spillover effects. Using the same format, measure-level FR and SO values sourced from the Massachusetts Technical Resource Manual (TRM) were drawn as shown in Fig. 2 below. In the measure-level plot, most of the observations were found to be in the bottom left quadrant, representing low SO and FR effects.

From the figures above, it will be incorrect to draw any general conclusions regarding the program effects. However, we feel that not assessing all components of the net savings does not capture the true program effects and might lead to incorrect policy decisions. Furthermore, we also observe that the FR and SO percentages are not uniformly distributed across all four quadrants, suggesting that free ridership and spillover percentages vary significantly across the measures and might not offset each other. These findings do not support the argument that NTG ratios are likely to be estimated as 1 in the majority of existing evaluation studies (Haeri and Khawaja 2012) except perhaps for low income, hard-to-reach small business, and new programs, in which values for FR and SO are estimated to be negligible (PWP, Inc. 2017). Numerical values for market effects were not separately available in the compiled studies, but NTG ratios vary widely from 0.49 to 1.98 across reviewed programs and measures in the Massachusetts TRM. The review of these evaluation reports indicates that NTG ratios vary significantly across measures and programs across states, suggesting that free

ridership and spillover effects do not necessarily offset each other, which is what some of the states are implicitly assuming if they do not require the adjustment of gross savings for FR and SO. As such, assuming an NTG ratio equal to 1 may not capture the true program effects and either over- or understate the actual savings attributable to a given program.

Adding FR and SO loses essential information about a program’s design

Based on our review of the free ridership and spillover values estimated in the EM&V reports, we discuss the policy implications from the perspective of cost-benefit analysis of program impacts on ratepayers. Figure 3 illustrates four possible scenarios for a hypothetical EE program, with combinations of free ridership percentage scores of either “low” or “high” values. The scenario depicted in Quadrant III indicates that a program’s free ridership percentage is low, and that spillover is high. This may be considered a relatively cost-effective scenario, as little program expenditure will incur energy savings from program participants, while non-participants will accrue higher benefits relative to programs falling within the other quadrants. Alternatively, the scenario in Quadrant I is suboptimal, as free ridership outweighs spillover, indicating a lower level of cost-effectiveness with a relatively lower level of positive spillover and market effects to provide offsetting benefits. Given this inefficient allocation of resources, the program will likely invite criticism, potentially requiring it to be redesigned, more effectively implemented, or discontinued.

From a cursory review, the scenarios in Quadrants II and IV appear to have similar outcomes, as the effects of free ridership and spillover appear to offset each other. However, in practice, the two scenarios are dissimilar, suggesting different outcomes and implications about the effectiveness of each program. Low free ridership and spillover percentages (Quadrant II) indicate a well-designed and implemented program compared with those illustrated in the remaining three quadrants. The low percentage values of free ridership and spillover in Quadrant II are not likely to have significant financial or distributional implications, even if these effects do not offset each other completely. However, the presence of high free ridership and spillover in Quadrant IV may have significant implications for actual savings—even if the difference is small in percentage terms—if gross

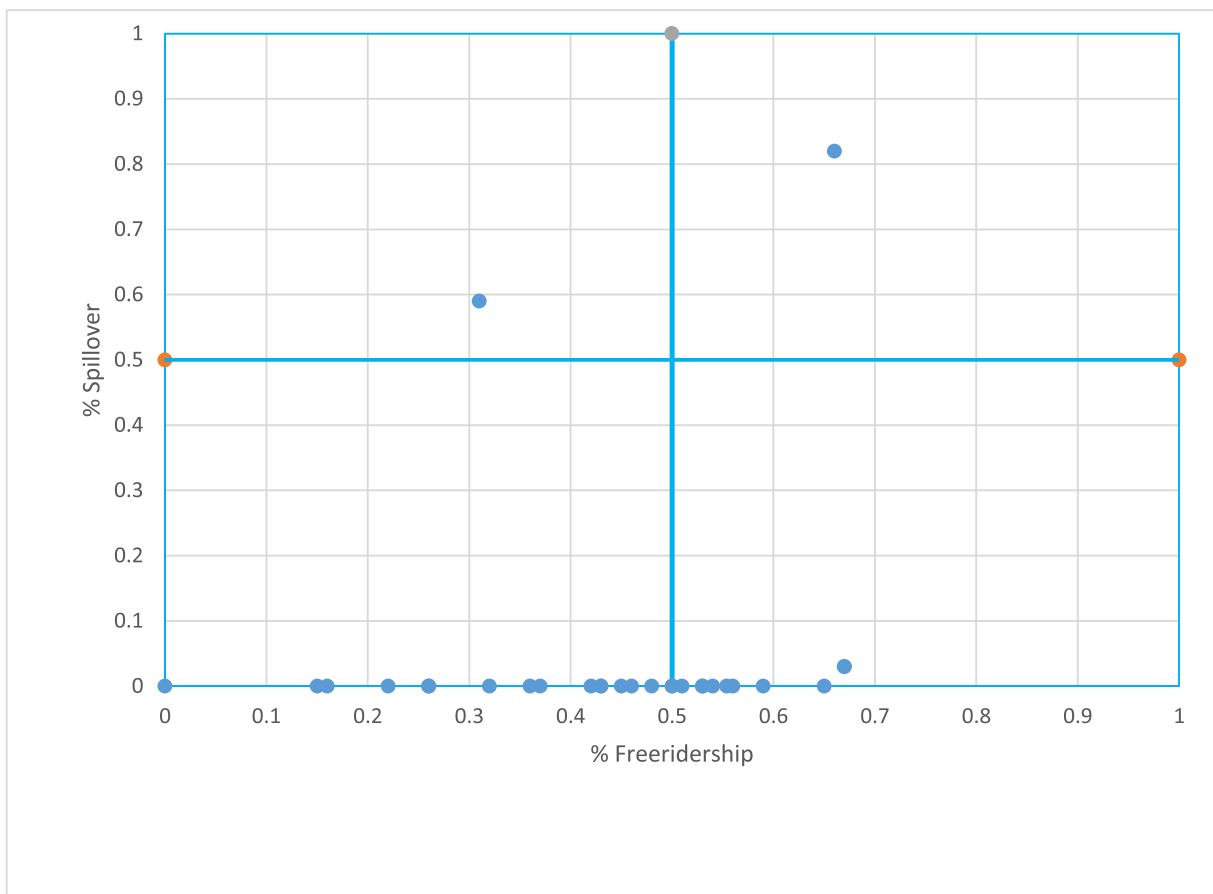


Fig. 1 Free ridership vs. spillover scatterplot (program-reported values)

energy savings are considerably high. Secondly, high free ridership and spillover percentages indicate a mature market for a given technology, positioned further up along a typical s-curve illustration of market penetration. Thus, for technologies that already achieve a sufficient market saturation without additional incentives, it is most practical and cost-effective to allow these to compete naturally in the market without incurring the costs of program intervention.

Additionally, the existing literature acknowledges with broad consensus that FR and SO exist within established EE programs and are significant enough to impact program savings, few studies to date acknowledge or attempt to measure the overlap between free ridership and spillover effects. A notable exception is Mahone and Hall's acknowledgement that EE programs with aggressive efficiency goals create an almost unavoidable overlap with similar programs within a given market (Mahone and Hall 2010). In such a case, one program's participants contributing positively to

net savings may be considered free riders detracting from another program's own savings. For example, Program A, a hypothetical residential lighting rebate program that induces its participants to enroll in hypothetical Program B, an existing smart thermostat rebate program, must recognize these savings as attributable to Program A, which are also assumed to be accounted for in Program B's tracking system. However, in this scenario, because Program A induces the additional spillover savings, Program B should not attribute these savings to its own program at the risk of double-counting net savings for each program. This scenario also suggests that spillover savings attributable to Program A may in fact cause free ridership in Program B, given that a participant would have adopted Program B's efficiency measure in the absence of Program B. Furthermore, even if properly attributed as Program B free ridership and Program A spillover savings, the realized savings will not be offset under each program's respective evaluation method, given that energy savings are not equivalent across unlike measures. As such, simply adding

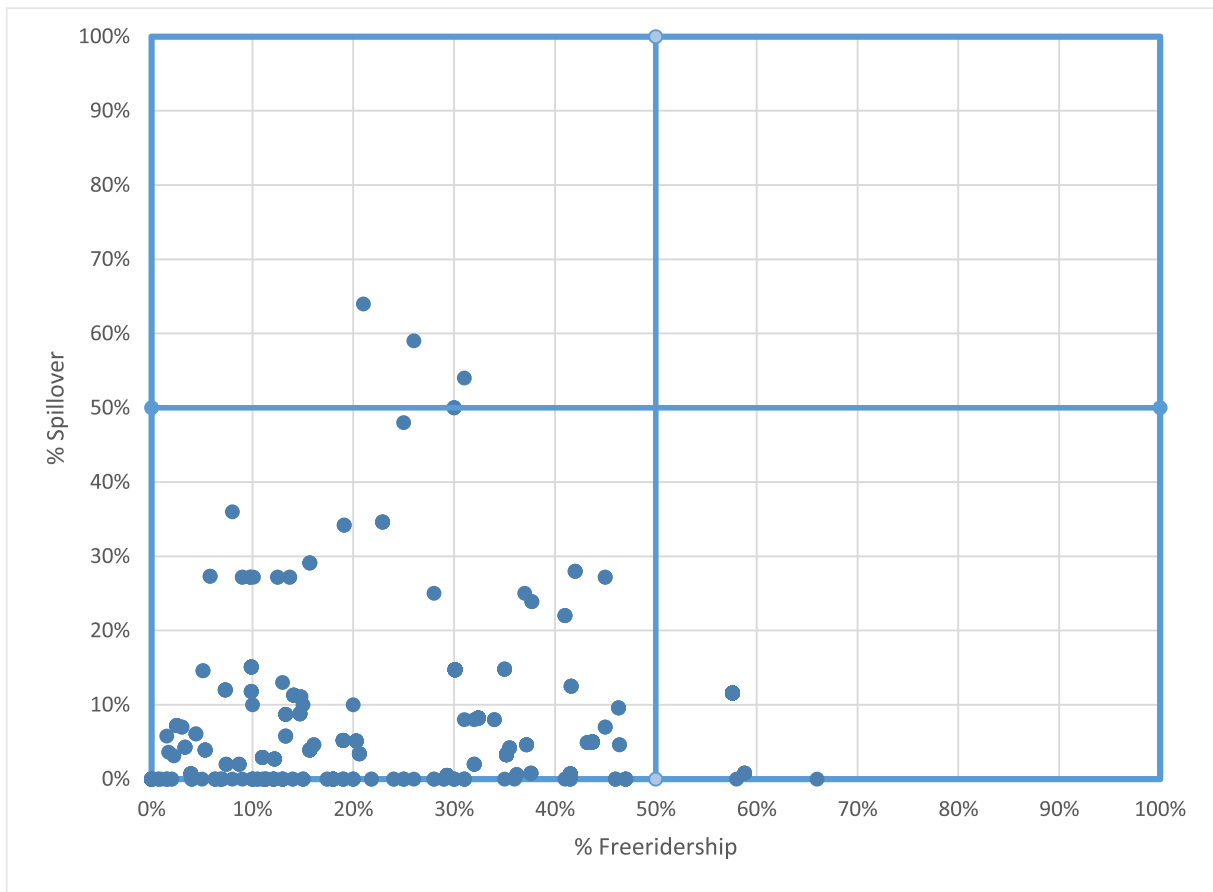


Fig. 2 Measure-level free ridership vs. spillover scatterplot (Massachusetts TRM)

together estimated values may not capture the dynamic overlaps between these effects and potentially result in inaccurate estimations of actual savings.

Given these considerations, EM&V practices must account for these overlaps in their respective evaluation and survey techniques, with similar logic extended to

Fig. 3 Free ridership vs. spillover matrix

		% Free Ridership	
		Low	High
% Spillover Savings	Low	Q. II (FR Low, SO Low)	Q. I (FR High, SO Low)
	High	Q. III (FR Low, SO High)	Q. IV (FR High, SO High)

emphasize the overlap between the various categories of FR, SO, and ME across the measure, program, and portfolio levels over time. For example, Itron's EmPOWER Maryland evaluation does not attribute spillover savings estimates to its residential appliance recycling subprogram to avoid potentially double-counting these savings, given that survey respondents indicated that they had adopted measures eligible for incentives through other EE programs (Itron, Inc. 2014).

Discussion and alternative approaches

Another major finding of our review is that most of the evaluation reports do not explicitly estimate FR and SO values in energy units. Instead, they express FR and SO in percentage terms and erroneously add within and between these categories to arrive at NTG estimates. This analysis argues against this approach, given that percentages must correspond to physical quantities in the same units in order to be added numerically. Thus, for FR and SO to be combined congruently, these values must be expressed in common energy units (i.e., megawatt-hours, therms, or megawatts for demand reductions) or in percentages of the same units. Furthermore, given their disparate functions, energy and demand savings cannot be added across unlike EE measures because they consume energy and/or reduce its usage in inherently different ways. Because unlike EE measures vary across the type of energy resource saved (e.g., natural gas, coal, etc.) and the timing of savings (e.g. hourly, seasonally), these savings are inherently unequal and cannot be appropriately added together to holistically express total net savings.

In most of the evaluated studies for this report, FR and SO are estimated based on assigned numerical scores to a set of survey questions and expressed in percentage terms. These numbers are then added together to estimate total energy savings, with a program's cost-effectiveness based upon the incremental costs incurred on portfolio of measures. Survey-based methods are among the most commonly used approaches to estimate FR, SO, and NTG ratios due to their cost-effectiveness and flexibility. However, survey results are susceptible to social desirability bias and the arbitrary assignment of FR scores based on the subjectivity of the surveyor or evaluator (Li et al. 2018). Furthermore, it is observed from those reports that detail their

respective survey methodologies and publish sample questionnaires, it is not possible for a reader to answer these questions and arrive at specific free ridership or spillover percentages because the underlying assumptions and functions to calculate these values are not made publicly-available. Thus, a transparent comparison of FR and SO across programs is not possible by reviewing published EM&V studies alone due to the differences between their respective methodologies, terminologies, and irreproducible survey results.

An alternate method to estimate FR could be in terms of the difference between a program participant's total willingness-to-pay for a given measure and the program's observed investment cost expressed in dollar terms (Grosche and Vance 2009). This methodology may also provide a more accurate illustration of a program's cost-effectiveness in terms of dollars per unit of reduced energy consumption or in terms of dollars per ton of carbon dioxide abated. Another more accurate and reliable approach to estimate net savings is to measure the difference in energy usage between program participants "treatment group" and a similar comparison group of non-participants "control group" at the same time based on randomized controlled trials or quasi-experimental methods (State and Local Energy Efficiency Action Network 2012). For example, by comparing inframarginal program participants and non-participants who did not meet program eligibility thresholds, a study of Mexican subsidies for replacement of refrigerators and air conditioners found that approximately half of all participants would have purchased the energy-efficient appliances in absence of the program (Boomhower and Davis 2014). In a recent working paper on cross-program spillover, Jessoe et al. employed randomized controlled trials to test the effects of social norm messaging about residential water usage on electricity consumption (Jessoe et al. 2017). Their results provide experimental evidence that behavioral interventions spill over to untreated sectors by altering consumer choice. Such techniques—especially randomized controlled trials—are often considered the highest standard of practice in the social sciences to establish causality (Li et al. 2018). These methods may have higher costs and time requirements and other limitations pertaining to ethical issues and applicability for large-scale evaluations. However, they may—with appropriate application—significantly limit bias and enhance the comparability and reliability of evaluation results.

Conclusions

Based on our review of the impact evaluation reports of EE programs for the cost-benefit analysis from rate-payers' perspective, we find there is a clear disconnect between the available guidelines and the actual practices employed in reality. Our review of the existing EM&V reports indicates that the majority of FR and SO estimates are based on survey results and expressed in percentage terms. Although percentages serve as a simple and useful indicator, the analysis in the “[Adding FR and SO loses essential information about a program's design](#)” section and the “[Discussion and alternative approaches](#)” section suggests that adding these percentages without converting the individual components into a common unit is inaccurate and obscures a program's true effects. Such methodologies may also fail to capture the dynamic overlaps between these effects and potentially result in inaccurate estimations of true program savings. Furthermore, a wide variation exists regarding the methodologies and estimated values of FR and SO across different states, thus inhibiting meaningful comparisons with be made across various program designs. The scatterplot analysis of the reviewed EM&V reports indicates that FR and SO do not necessarily offset each other, except perhaps for low income, hard-to-reach small business, and new programs, for which values for these effects are estimated to be close to zero. As such, simply adding the percentages without converting them into common units ignoring the overlaps between FR, SO, and ME, and assuming an NTG ratio equal to 1 may over- or understate the actual savings attributable to a given program.

An alternative approach for the analysis of FR and SO is to estimate and report these values in terms of energy units and analyze them in dollar terms to more accurately determine program effectiveness. Based on the dynamic behavior of FR as a function of a program's presence in a given market, another method would be to estimate free ridership in terms of participant's total willingness-to-pay for EE measures in the absence of program. This method may provide a more accurate illustration of a program's cost-effectiveness in terms of dollars per unit of reduced energy consumption or in terms of dollars per ton of carbon dioxide abated. As such, FR, SO, and ME should instead be reported in terms of a program's gross energy savings and associated dollar savings for a more transparent program evaluation design. Additionally, for a meaningful comparison across the measure, program, portfolio, and utility

levels, it is recommended that a consistent and reliable methodology be uniformly adopted across all EM&V studies. As a step forward, randomized or quasi-experimental designs can be tried for more accurate impact evaluations and for better comparability of EE programs across jurisdictions.

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

References

- Alberini, A., & Bigano, A. (2014). How effective are energy efficiency incentive programs? Evidence from Italian Homeowners. *SSRN*.
- Angelucci, M., & Maro, V. D. (2015). *Program evaluation and spillover effects*. Retrieved from World Bank Group Policy Research Working Paper: <http://documents.worldbank.org/curated/en/137731468171858310/pdf/WPS7243.pdf>.
- Berg, W., Nowak, S., Kelly, M., Vaidyanathan, S., Shoemaker, M., Chittum, A., et al. (2019). *ACEEE Scorecard*. Washington, DC: ACEEE.
- Boomhower, J., & Davis, L. W. (2014). A credible approach for measuring inframarginal participation in energy efficiency programs. *Journal of Public Economics*, 113, 67–79.
- Felder, F. A., & Athawale, R. (2017). "PACT-a-Mole": the case against using the Program Administrator Test for energy efficiency programs. *Energy Efficiency*.
- Friedrich, K., Eldridge, M., York, D., Witte, P., & Kushler, M. (2009). *Saving energy cost-effectively: a national review of the cost of energy saved through utility-sector energy efficiency programs*. Washington, DC: American Council for an Energy Efficient Economy.
- Grosche, P., & Vance, C. (2009). Willingness to pay for energy conservation and free-ridership on subsidization: evidence from Germany. *The Energy Journal*, 135–153.
- Haeri, H., & Khawaja, M. S. (2012). *The trouble with freeriders* (p. 40). *Public Utilities Fortnightly*.
- Haugland, T. (1996). Social benefits of financial investment support in energy conservation policy. *The Energy Journal*, 79–102.
- IEA. (2017). *International Energy Agency, Energy Efficiency 2017*.
- Itron, Inc. (2014). *Verification of reported program impacts from 2013 empower maryland energy efficiency programs with recommendations to improve future evaluation research*. Silver Spring: Itron, Inc..
- Jaccard, M. (2010). Paradigms of energy efficiency's cost and their policy implications; déjà vu all over again. *Modelling the Economics of Greenhouse Gas Mitigation: Summary of a Workshop* (pp. 1-19). The National Academic Press.
- Jessoe, K., Lade, G. E., Loge, F., & Spang, E. (2017). *Spillovers from behavioral interventions: experimental evidence from water and energy use*. Berkeley: E2e Project.

- Keay, M. (2011). *Energy efficiency- should we take it seriously?* The Oxford Institute for Energy Studies.
- Kraft, J., Rathbun, P., & Le Beau, L. (2010). *2009 Commercial and Industrial Programs Free-ridership and Spillover Study, Final Report*. Madison: PA Knowledge Limited.
- Li, M., Haeri, H., & Reynolds, A. (2018). *The uniform methods project: methods for determining energy efficiency savings for specific measures*. Golden: National Renewable Energy Laboratory.
- Mahone, D., & Hall, N. (2010). *Pursuit of aggressive energy targets: aligning program design and evaluation*. Washington: American Council for an Energy Efficient Economy.
- Massachusetts Executive office of energy and environmental affairs. (2015). *Massachusetts Clean Energy and Climate Plan for 2020: 2015 Update*. Boston: Massachusetts Executive Office of Energy and Environmental Affairs.
- Navigant Consulting, Inc. (2016). *Plan Year 6 Summary Evaluation Report*. Chicago: Navigant Consulting, Inc..
- New York State Joint Utilities. (2018). *Department of Public Service New York*. Retrieved from Department of Public Service New York: <http://www3.dps.ny.gov/W/PSCWeb.nsf/A11/72C23DECF52920A85257F1100671BDD?OpenDocument>.
- NMR Group, Inc. (2015). *Efficiency Maine Retail Lighting Program Overall Evaluation Report*. Somerville: NMR Group, Inc..
- Opinion Dynamics. (2017). *Ameren Illinois Company Electric & Gas Residential, and Commercial and Industrial Energy Efficiency Portfolios – PY8 Summary Evaluation Report*. Waltham: Opinion Dynamics.
- PWP, Inc. (2017). *Current Methods in Free Ridership and Spillover Policy and Estimation Final*. North Potomac: PWP, Inc..
- Ridge, R., Baker, M., Hall, N., Pahl, R., & Saxonis, W. (2013). Gross is gross and net is net: simple right? *Progress in Industrial Ecology- An international Journal*, 145-155.
- Rosenow, J., & Ray, G. (2012). Evaluating the evaluations: evidence from energy-efficiency programmes in Germany and the UK.
- Slote, S., Sherman, M., & Crossley, D. (2014). *Energy Efficiency Evaluation, Measurement, and Verification*. Regulatory Assistance Project.
- State and Local Energy Efficiency Action Network. (2012). *Energy Efficiency Program Impact Evaluation Guide*. Chicago: Schiller Consulting, Inc..
- Vine, E., Pahl, R., Meyers, S., & Turiel, I. (2009). An approach for evaluating the market effects of energy efficiency programs. *Energy Efficiency*.
- Vine, E., Hall, N., Keating, K. M., Kushler, M., & Pahl, R. (2010). Emerging issues in the evaluation of energy efficiency programs: the US experience. *Energy Efficiency*.
- Violette, D. M., & Rathbun, P. (2017). *Chapter 21: Estimating Net Savings – Common Practices The Uniform Methods Project: Methods for Determining Energy Efficiency Savings for Specific Measures*. Retrieved from NREL Report: <https://www.nrel.gov/docs/fy18osti/70472.pdf>.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.