RESOURCES
FOR THE FUTURE

January 21, 2016

US Environmental Protection Agency
1200 Pennsylvania Avenue NW
Washington, DC 20460

RE: Comments on EPA's Draft Evaluation, Measurement, and Verification Guidance for Demand-Side Energy Efficiency

As a Research Director and Senior Fellow at Resources for the Future (RFF), I am pleased to offer the attached comments to the United States Environmental Protection Agency (EPA) on its Draft Evaluation, Measurement, and Verification Guidance for Demand-Side Energy Efficiency.

As you know, RFF is a nonprofit and nonpartisan organization that conducts independent research—rooted primarily in economics and other social sciences—on environmental, energy, and natural resource policy issues. RFF neither lobbies nor takes positions on specific regulatory proposals, although individual researchers are encouraged to express their unique opinions—which may differ from those of other RFF experts, officers, and directors.

As always, the goal at RFF is to identify the most effective ways—from an economic perspective—to meet environmental objectives through regulation, policy, or market mechanisms. To that end, I have been working with my colleagues at RFF to analyze EPA's Clean Power Plan and I have a long-standing interest in advancing the state of the art for energy efficiency evaluation, measurement, and verification to enable the best use of this tool for reducing emissions and meeting state emissions reduction targets. The comments attached are my own and I hope they are useful for the development of the final guidance document.

Please feel free to contact me directly with any questions.

Sincerely,

Karen Palmer

Karen Palmer
Research Director and Senior Fellow
Resources for the Future
palmer@rff.org

Comments on EPA's Draft Evaluation, Measurement, and Verification (EM&V) Guidance for Demand-Side Energy Efficiency (EE)

Karen Palmer
Senior Fellow and Research Director, Resources for the Future
202.328.5106 | palmer@rff.org

End-use energy efficiency is expected to play an important role in the development of state plans to comply with the US Environmental Protection Agency's (EPA's) Clean Power Plan (CPP) and could be particularly important in those states that adopt an emissions rate-based approach to comply with EPA's carbon dioxide (CO_2) emissions guidelines. Under such an approach, entities including utilities, community groups, energy agencies, and others that take certain actions to increase investments in end-use energy efficiency (EE) are entitled to earn emissions reduction credits (ERCs) for every demonstrated MWh of electricity saved. In the aggregate, crediting energy savings in this way is expected to reduce demand for electricity and the need for generation from all sources, including emitting sources. ERCs can be used to augment the denominator in an affected generator's calculated emissions rate, thereby bringing the generator closer to compliance with the emissions rate standard. By supplementing the denominator in this way, efficiency-associated ERCs lessen the need for a generator to reduce its own emissions directly.

Under such a rate-based regime, the veracity of energy savings and associated ERCs are very important to the environmental integrity of the policy; thus, EPA rightly requires in the final CPP rule that entities wishing to earn ERCs for energy savings submit an ex-post evaluation conducted by an independent third party to verify that their actions or policies have yielded the claimed amount of energy savings. The draft evaluation, measurement, and verification (EM&V) guidance document issued in conjunction with the final rule and proposed model rule/federal plan offers a very useful description and guide for the states on how efficiency evaluations are currently conducted and provides a useful basis for moving forward. EPA is to be commended for putting the guidance document together and providing the opportunity to comment.

Summary of Comments

The comments below focus largely on the questions posed by EPA on the following:

- whether the guidance should specifically encourage greater use of comparison group approaches,
- the appropriate use of deemed savings, and
- potential intersections between the two approaches.

They also focus primarily on energy efficiency programs and measures implemented by utilities or third parties for electricity customers and not on policies such as appliance standards or building codes, although some of the studies cited below have been used to evaluate energy savings from the latter.

The primary conclusions of these comments are as follows:

- EM&V of energy efficiency measures, programs, and policies are both methods of measuring savings from different approaches to reducing energy use and means of identifying the most effective and cost-effective ways to encourage energy savings.
- The EPA EM&V guidance should be a living document that accommodates future advances in EE programs, policies, and technologies as well as new developments in program evaluation.
- EPA should advocate and provide incentives for greater use of comparison approaches (including randomized control trials, or RCTs, quasi-experiments and empirical approaches using state-of-the-art matching techniques) for evaluating EE measures, programs, and policies within the states.
- EPA should establish a centralized database of RCT and quasi-experimental evaluations of EE programs that includes information about the programs, the revealed energy savings, and, whenever possible, the underlying data used to conduct these analyses. This revealed savings database should be made available to the states and would form the basis of a collection of savings estimates that could be applied to similar measures and programs in other settings with appropriate adjustments as necessary.
- EPA, in conjunction with the US Department of Energy (DOE), should establish a task force or working group that includes representatives from states, utilities, and the efficiency community and experts in experimental and empirical methods to identify a set of high priority RCTs of EE programs and measures for evaluation. One priority for early evaluation would be EE programs in low-income neighborhoods, which are incentivized under the Clean Energy Incentive Program (CEIP).
- Once priorities are established, EPA and DOE should launch a series of EE RCTs and other research experiments to begin to populate the revealed savings database and help to facilitate a collection of savings estimates that can help inform future evaluations.

New Policy, Evolving Methods, and More Data

The development of EPA's Clean Power Plan, the associated federal plan/model rules and the 111(b) regulations for new electricity generators together set a new course for the future of the electricity sector—one where electricity generators must take into account the CO_2 consequences of their generation and investment actions. As this course has been and continues to be mapped out at EPA and in the states, two other revolutions are simultaneously unfolding in the areas of program evaluation and energy consumption data generation. This confluence of new developments, described in more detail below, suggests that the state of the art in both EE program design and EE evaluation is currently evolving. EPA's guidance to the states needs to be flexible enough to accommodate the insights from that ongoing evolution and to encourage methods that provide the best assessment of program-induced energy savings given available resources, with an eye toward making the best use of EE program funding.

On the program evaluation front, there have been important discoveries in recent years regarding the limitations of previously accepted approaches for measuring the causal relationships between EE policies or other market interventions and intended outcomes, as evidenced in the economics literature (Angrist and Pische 2009; Morgan and Winship 2014). As a consequence, the economic research community has shifted toward reliance on quasi-experimental techniques and RCTs for program evaluation studies that seek to draw conclusions about causal relationships. These methods are able to separate the effects of confounding factors (including differences between program participants and non-participants that might affect outcomes, differences in propensity to adopt policies, etc.) from the effects of the EE intervention itself through randomization or ways of limiting program participation that are independent of consumer responses to EE programs.

RCTs are used frequently in a number of other areas as well—including evaluation of economic development policies, education interventions, and the efficacy of new drugs—and they are increasingly making their way into academic work on energy efficiency (Allcott 2011; Ayres et al. 2013) and into guidance documents on how to evaluate energy efficiency programs (State and Local Energy Efficiency Action Network 2012; Stewart and Todd 2015). Because the random assignment of customers to particular EE interventions required under RCTs can be very difficult to implement, randomized encouragement (RE) methods are also being used to study the effectiveness of EE programs and policies (Fowlie et al. 2015). RE methods involve randomly selecting customers into a group that is more heavily encouraged to participate in an EE program that those in the control group. Quasi-experimental methods have also played a growing role in EE evaluation in recent years (Davis et al. 2014; Jacobsen and Kotchen 2013; Kotchen 2015).

On the data side, the rapid penetration of interval meters means high-frequency data on electricity consumption are being collected at a growing number of customer premises. Smart thermostats are also helping consumers and utilities to better understand energy use for heating and cooling and providing opportunities to help shape that consumption. The data collected by these devices are an important ingredient to advancing our understanding of the efficacy of efficiency measures, but the data alone are not sufficient. Another equally important step is to make these data available to the research community that has the interest and the tools to conduct the relevant analysis. And doing the right kind of analysis is crucial to understanding how well EE interventions are working to reduce energy use. In addition, these data are enabling new energy efficiency offerings including nudge programs, services such as no-touch audits based on energy data analytics, and detailed characterization of energy use at a customer’s premises that could help with targeting of particular EE programs and incentives.

These developments on the evaluation methodology and data fronts are ongoing and together they suggest changes in the types of opportunities for creating energy savings and associated policy interventions, as well as in approaches to EE evaluation moving forward. EPA’s guidance document should more explicitly recognize these opportunities and envision an expanded role in the future for experimental and quasi-experimental approaches.

In the following paragraphs I discuss the challenges and opportunities that affect the guidance document and the larger issue of incorporating energy efficiency into compliance planning for the CPP.

1616 P St. NW Washington, DC 20036-1400 Tel 202.328.5000 fax 202.939.3460 info@rff.org www.rff.org

Current EM&V Methods and Selected Issues

The EPA guidance document groups existing methods for evaluation into three categories:

1. project-based measurement and verification (PBMV),
2. deemed savings methods, and
3. comparison group methods.

The document also discusses how to determine which approach best fits which circumstance. Two of these approaches (PBMV and deemed savings) rely heavily on engineering methods for establishing baseline conditions and calculating energy savings. Comparison group methods include RCTs, randomized encouragement and quasi-experiments, but also encompass the use of matching methods such as propensity score matching and other approaches to define a comparison group against which changes in consumption associated with a program or measure can be measured. Within the collection of comparison group methods, RCTs, randomized encouragement and quasi-experiments provide cleaner identification of causal effects than matching methods due to the inability to control for unobservable factors.[1] However, all of the comparison group methods make use of actual consumption data thereby incorporating the effects of an EE measure or program on energy service consumption behavior and reducing the need to rely on assumptions. This and other features of comparison approaches are explored in greater detail below.

Better Accounting for Behavior

When engineering methods are used for evaluation, the implications of an EE measure for electricity customer *behavior* is largely driven by assumptions, often informed by surveys or metering studies for a small sample of customers but typically subject to a large amount of uncertainty. Potentially important questions arise about how well these assumptions represent actual behavior, under baseline and post-adoption conditions, but also the technology adoption decision itself. For example, some energy efficient technologies, such as more efficient forms of lighting, may result in changes in the quality of the light produced that could affect use of lights in important ways that are not factored into engineering calculations.[2] Also, because engineering approaches focus on how new technologies affect energy consumption and not on policy impacts on technology adoption itself, evaluators often rely on estimates of net to gross ratios to isolate that portion of the energy savings from technology adoption that is attributable to an efficiency intervention as opposed to adoption that would have occurred without the EE program. This attribution issue is particularly important in voluntary programs, which most EE programs are. Information about net to gross ratios typically comes from surveys of program participants that ask whether they would have adopted the technology in the absence of the efficiency program or measure. Such survey methods are generally not as reliable as methods based on observed behavior, especially when survey sample sizes are small.

[1] For more information on the range of comparison group methods and their properties see State and Local Energy Efficiency Action Network (2012) and Stewart and Todd (2015).

[2] The Uniform Methods Project document on measuring savings from residential lighting programs (Dimetrosky et al. 2015) indicates that more research is needed on how changes in lighting technology affect hours of use, particularly since most evaluations do not assess lighting use prior to program implementation.

1616 P St. NW Washington, DC 20036-1400 Tel 202.328.5000 fax 202.939.3460 info@rff.org www.rff.org

Comparison group methods bypass both of these shortcomings by assessing electricity savings through direct comparisons of changes in observed consumption between participating customers and comparison group customers. Any changes in usage in response to changes in energy service quality are captured directly, as is the effect of the policy or intervention on technology adoption. Because comparison methods are more reliable, there is a higher degree of certainty that energy savings identified by these approaches are additional to baseline energy use and thus, within the context of the CPP, are arguably more entitled ERCs that can be credited to CPP compliance. [3]

Behavioral and Operational Interventions

RCTs are the method of choice for evaluating the effectiveness of behavioral interventions such as OPower nudges that compare a customer's energy use to that of her neighbors. The OPower experience has shown that these techniques can be used to identify even the small percentages (typically between 2 and 3 percent) of energy savings that result from these interventions. Because these programs are not focused on specific technologies, engineering approaches to estimating savings are not feasible. Given the vast amounts of frequent energy consumption data being collected by smart meters and the findings by at least one data analytics company that substantial energy savings in commercial buildings could be achieved with changes in operations[4], future opportunities for savings from EE could lie increasingly in the operational realm, calling for greater reliance on comparison group methods to evaluation.

Linking Program Design and Evaluation

One aspect of comparison approaches that could be better communicated in the guidance document is that RCTs and, to a certain extent, quasi-experimental methods are not approaches to evaluation that can be applied after the fact. Both RCTs and quasi-experimental methods affect the initial design of an EE program and thus evaluators need to be involved in the development of these programs from the beginning. In the case of RCTs, the level of up-front involvement is likely greater than for a quasi-experiment due to the need for a clear articulation of the features of the EE program to be tested and assurances that randomization is adhered to and sample sizes are sufficient. While it is possible to do a quasi-experimental evaluation after an EE program has been executed, having a deliberate strategy in mind from the beginning will help produce better outcomes and more useful lessons for future policies.

Redeeming Deemed Savings by Building a Database of Revealed Energy Savings

Deemed savings provide energy savings values associated with a range of technology-focused EE measures and are typically most relevant where behavioral changes induced by the measure are less of a concern, such as replacing an appliance that runs all the time. The guidance document also cautions against using deemed savings measures when savings are expected to be large. Besides transferability, an additional concern about deemed savings approaches is the extent to which they are based on the outcome of an evaluation method that provides a robust

[3] For this reason, utilities that have faced challenges in regulatory proceedings to their ex post estimates of energy savings may prefer greater reliance on comparison methods that provide a more defensible result.
[4] See http://www.firstfuel.com/resources/library/infographic/low-no-cost-operational-changes-could-double-energy-efficiency-in-commercial-buildings/ for more information.

1616 P St. NW Washington, DC 20036-1400 Tel 202.328.5000 fax 202.939.3460 info@rff.org www.rff.org

way of identifying the causal effects of a particular measure or program. On page 17 in section 2.4, the draft guidance states that "deemed savings values should be developed by independent third parties and, whenever possible, should be based on empirical techniques such as RCTs and quasi-experimental designs."

This recommendation points to the importance of having disinterested third parties play a role in EE evaluation and of using state-of-the-art techniques whenever possible. It also suggests making the results of those RCT and quasi-experimental studies available to the EE evaluation community for potential use in other settings. One way that this idea could take shape is through the establishment of a centralized database of RCT and quasi-experimental evaluations of EE programs. Such a database should include not only information about the programs and resulting energy savings estimates but also, whenever possible, the data used to inform these analyses, including both electricity consumption data and data on other relevant parameters such as weather, dates, and location that define the setting for the measure being evaluated.

Making these data available in an organized fashion will facilitate replicability, and therefore credibility, of the findings. It will also help to facilitate transferability under appropriate circumstances. An accumulation of studies with associated data could help to form the basis for a meta-analysis of the findings of existing studies that could ultimately enable the transfer of these findings to similar EE programs in neighboring utility regions or states or other contexts with appropriate adjustments for differences in relevant parameters.

Assembling such a database could also be a way of socializing lessons from various experiments about policy design and how to make EE policies more effective. Such an effort could be led by EPA, perhaps in cooperation with DOE, together with representatives from states, utilities, EE evaluators, and researchers who develop and use these techniques in the EE evaluation (or related energy) contexts.[5] EPA should also provide guidance to the states as to when the database is sufficiently developed to be usable as a source of "revealed savings" estimates that would be presumably transferable to similar efficiency programs in other settings.

Identifying and Kick-Starting Priority Experiments to Advance EE Policy and Evaluation

To help kick-start this effort, EPA, in conjunction with DOE, should establish a task force or working group that includes representatives from states, utilities, and the efficiency community and experts in experimental methods to identify a set of high-priority RCTs of EE programs and measures for evaluation. Once priorities are established, EPA and DOE should also launch a series of EE RCTs and other research experiments to begin to populate this database. High priority experiments could be focused on improving understanding about popular program areas that prior non-experimental evaluations suggest have large savings (to provide experimental evidence on these apparent effects). More experimental evidence would also be useful for EE programs that target low-income communities as such programs are singled out in the Clean Energy Incentive Program (CEIP) under the CPP. A recent analysis by the Lawrence Berkeley National Laboratory (Hoffman et al. 2015) of data from rate-payer funded programs suggests that low-income programs are among the least cost-effective of all EE programs and that there is

[5] This group could be constituted similarly to a SEEAction task force, but with a greater focus on technical expertise in the development and execution of RCTs or quasi-experimental methods among some portion of its membership.

a wide range in cost per kWh saved estimates among different low-income programs. Moreover, in a randomized encouragement experiment, Fowlie et al. (2015) find that realized energy savings in the Michigan WAP were roughly half those predicted by the *ex-ante* engineering models. Experimental research on low-income policies could help states as they develop their plans for making use of the CEIP and evaluating associated energy savings outcomes.

Another way that federal agencies (EPA, DOE, Department of Defense, etc.) could enhance the use of more robust methods of EM&V and the accumulated lessons learned through their application is by conducting more EE experiments within its own fleet of buildings. The federal government owns or leases more than 700,000 buildings or structures; government agencies spend a lot of money on building energy use, accounting for roughly 2 percent of total energy use in buildings in the United States in recent years. In an executive order issued in March 2015, President Obama announced his intention to reduce energy intensity in federal buildings (Btu/square foot) by 2.5 percent per year between 2015 and 2025. RCTs could be used as a way to track savings and identify the most effective strategies that could be undertaken for achieving those goals. Lessons from such an effort could be folded into the experimental EE database and used to inform program designs in government buildings at other levels of government, as well as in other institutional settings.

Limitations

Experimental techniques do have some limitations. RCTs are costly and require randomized assignment of electricity consumers to an EE measure, which can be difficult to achieve. Randomized encouragement provides a way to introduce randomization while preserving customer choice, but this approach requires much larger groups of customers be included in the experiment for statistical power. RCTs or randomized encouragement approaches typically focus on specific features of the EE program being studied and it may be difficult to know in advance which features (size of incentive payment, nature of information, nature of default choice) is the most fruitful to study. RCTs will likely underestimate the savings associated with EE programs or measures that have important spillover effects on the energy consumption behavior of nonparticipants (so-called market transformation methods). Quasi-experimental approaches require some sort of qualification scheme (building size, building age) or waiting list (limit on participants in current period) that allows for the creation of a set of non-participating customers to form a comparison group. Depending on how that group is created, the lessons from comparisons may be of limited applicability beyond a small set of participants deemed sufficiently similar to the comparison group. Despite these limitations, much remains to be learned from more scientific analysis of the effectiveness of EE measures and programs.

To a certain extent, these limitations could be addressed with a coordinated effort to advance the use of comparison methods in a range of settings, such as the one described above. The time between now and the beginning of the first CPP compliance period affords an opportunity to begin to accumulate a body of evidence through more use of RCTs in areas of particular interest and where the greatest opportunities for future savings are perceived to be. EPA could also provide training to evaluators in states on how to conduct experiments and how to employ other sorts of comparison approaches.

1616 P St. NW Washington, DC 20036-1400 Tel 202.328.5000 fax 202.939.3460 info@rff.org www.rff.org

Bottom Line

Not all energy efficiency programs can or need be evaluated using experimental approaches, but the time has come for greater reliance on RCTs and other experimental techniques for evaluating energy savings in order to improve the effectiveness and cost-effectiveness of policies and measures designed to produce energy savings (Vine et al. 2014). Putting these methods into practice will require cooperation between researchers or evaluators and program administrators beginning at the program design stage. It will also require providing researchers and evaluators with access to electricity consumption data not just for program participants but for members of comparison groups—both before and after programs are in place. The results of these efforts can be used for a variety of purposes including helping with calibration of engineering models designed to predict future energy savings, identifying more effective and cost-effective policies, and informing “revealed savings” databases that can form the basis for more streamlined evaluations of future EE initiatives. The EPA EM&V guidance document should be enhanced to focus more on the values of comparison group methods, and EPA and other federal agencies should direct resources to more experiments and quasi-experimental evaluations.

CPP compliance begins in 2022 and crediting of EE programs in low-income communities under the CEIP starts two years prior. The years between now and then provide an excellent window during which experimentation and research can begin to advance knowledge about revealed savings from priority EE policies and programs. Lessons during this interim phase can help to inform a comprehensive revealed energy savings and associated studies database, and an evolving EM&V guidance document that pushes evaluators and policymakers toward more robust evaluation whenever feasible. By making the guidance a living document, EPA provides incentives for EE service and program providers and researchers to improve the effectiveness and cost-effectiveness of this promising approach to reducing greenhouse gas emissions, both within the electricity sector and beyond.

1616 P St. NW Washington, DC 20036-1400 Tel 202.328.5000 fax 202.939.3460 info@rff.org www.rff.org

References

Allcott, Hunt. 2011. Social Norms and Energy Conservation. *Journal of Public Economics* 95: 1082–1095.

Angrist, J.D., and J.S. Pischke. 2009. *Mostly Harmless Econometrics: An Empiricist's Companion*. Princeton University Press.

Ayres, Ian, Sophie Raseman and Alice Shih. 2013. Evidence from Two Large Field Experiments that Peer Comparison Feedback Can Reduce Residential Energy Usage, *Journal of Law, Economics, and Organization* 29(5): 992–1022.

Davis, Lucas, Alan Fuchs, and Paul Gertler. 2014. Cash for Coolers: Evaluating a Large-Scale Appliance Replacement Program in Mexico, *American Economic Journal: Policy* 6(4): 207–238.

Dimetrosky, Scott, Katie Parkinson, and Noah Lieb. 2015. Chapter 21: Residential Lighting Evaluation Protocol. *The Uniform Methods Project: Methods for Determining Energy Efficiency Savings for Specific Measures*. Subcontract Report NREL/SR-7A40-63205, February.

Fowlie, Meredith, Michael Greenstone, and Catherine Wolfram. 2015. Do Energy Efficiency Investment Deliver? Evidence from the Weatherization Assistance Program. E2e working Paper 020, June.

Hoffman, Ian, Gregory Rybka, Greg Leventis, Charles A. Goldman, Lisa Schwartz, Megan Billingsley, and Steven Schiller. 2015. *The Total Cost of Saving Electricity through Utility Customer-Funded Energy Efficiency Programs: Estimates at the National, State, Sector and Program Level*. Electricity Markets and Policy Group, Lawrence Berkeley National Laboratory, April.

Jacobsen, Grant, and Matt Kotchen. 2013. Do Building Codes Save Energy? Evidence from Residential Billing Data in Florida. *The Review of Economics and Statistics* 95: 34-49.

Kotchen, Matt. 2015. Do Building Energy Codes Have a Lasting Effect on Energy Consumption? New Evidence from Residential Billing Data in Florida. NBER Working Paper 21398.

Morgan S.L., and C. Winship. 2014. *Counterfactuals and Causal Inference: Methods and Principles for Social Research*. Cambridge University Press.

State and Local Energy Efficiency Action Network. 2012. *Evaluation, Measurement, and Verification (EM&V) of Residential Behavior-Based Energy Efficiency Programs: Issues and Recommendations*. Report of the Customer Information and Behavior Working Group and Evaluation, Measurement, and Verification Working Group, May.

1616 P St. NW Washington, DC 20036-1400 Tel 202.328.5000 fax 202.939.3460 info@rff.org www.rff.org

Stewart, James, and Annika Todd. 2015. Chapter 17: Residential Behavior Protocol. *The Uniform Methods Project: Methods for Determining Energy Efficiency Savings for Specific Measures*. Subcontract Report NREL/SR-7A40-62497 January.

Vine, Edward, Michael Sullivan, Loren Lutzenhiser, Carl Blumstein, and Bill Miller. 2014. Experimentation and the Evaluation of Energy Efficiency Programs., *Energy Efficiency* 7(4): 627–640.

1616 P St. NW Washington, DC 20036-1400 Tel 202.328.5000 fax 202.939.3460 info@rff.org www.rff.org