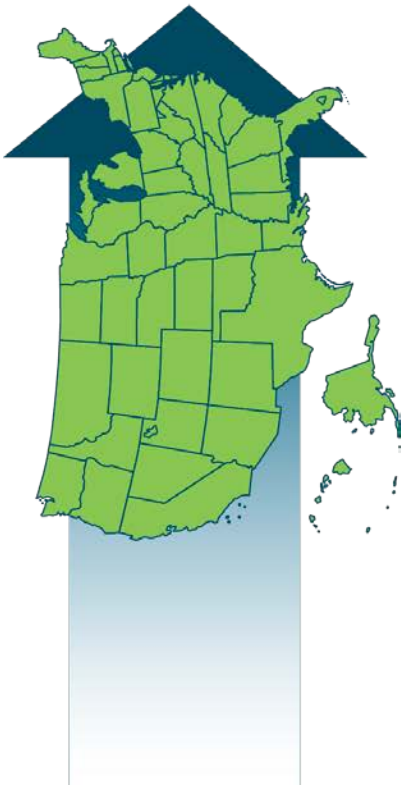STATE & LOCAL ENERGY EFFICIENCY ACTION NETWORK

SEE Action

ENERGY STAR® Retail Products Platform (RPP): Conditions and Considerations in Evaluating Market Transformation Programs and Evaluation Guidance for RPP

Evaluation, Measurement, and Verification Working Group

January 2018

The State and Local Energy Efficiency Action Network is a state and local effort facilitated by the federal government that helps states, utilities, and other local stakeholders take energy efficiency to scale and achieve all cost-effective energy efficiency by 2020.

Learn more at www.seeaction.energy.gov

ENERGY STAR® Retail Products Platform (RPP): Conditions and Considerations in Evaluating Market Transformation Programs and Evaluation Guidance for RPP is a product of the State and Local Energy Efficiency Action Network (SEE Action), facilitated by the U.S. Department of Energy and the U.S. Environmental Protection Agency. Content does not imply an endorsement by the individuals or organizations that are part of SEE Action working groups or reflect the views, policies, or otherwise of the federal government.

This document was final as of January 9, 2018.

If this document is referenced, it should be cited as:

Sara Conzemius and Alexandra Dunn. 2018. *ENERGY STAR® Retail Products Platform (RPP): Conditions and Considerations in Evaluating Market Transformation Programs and Evaluation Guidance for RPP*. Prepared by ILLUME Advising, LLC, for the State and Local Energy Efficiency Action Network.

FOR MORE INFORMATION

Regarding *ENERGY STAR® Retail Products Platform (RPP): Conditions and Considerations in Evaluating Market Transformation Programs and Evaluation Guidance for RPP,* please contact:

Michael Li   Peter Banwell
U.S. Department of Energy   U.S. Environmental Protection Agency
michael.li@ee.doe.gov   banwell.peter@epa.gov

Regarding the State and Local Energy Efficiency Action Network, please contact:

Johanna Zetterberg
U.S. Department of Energy
johanna.zetterberg@ee.doe.gov

Table of Contents

Acknowledgments

Evaluation, Measurement, and Verification (EM&V) of Market Transformation Programs: Issues and Recommendations for the ENERGY STAR® Retail Products Platform (RPP) is a product of the State and Local Energy Efficiency Action Network (SEE Action).

This report was prepared by Sara Conzemius and Alexandra Dunn of ILLUME Advising LLC under contract to the U.S. Environmental Protection Agency, and by Joe Van Clock of Research into Action. The SEE Action EM&V Working Group created a stakeholder group specifically for this report called the ESRPP Committee. The authors received direction and comments from the ENERGY STAR RPP Committee, and the following individuals from the committee provided specific input.

- Peter Banwell, U.S. Environmental Protection Agency

- Scott Dimetrosky, Apex Analytics

- Sue Hanson, TetraTech

- Katherine Johnson, Johnson Consulting Group

- Todd Malinick, Research Into Action

- Dulane Moran, Northwest Energy Efficiency Alliance

- Ralph Prahl, Independent Consultant

In addition to direction and comment by the ENERGY STAR RPP Committee, this report was prepared with highly valuable input from technical experts: Emily Clark (VEIC), Kevin Cooney (Navigant), Sam Dent (VEIC), Brian Hedman (Cadmus), Daniel Lawlor (Cadmus), Noah Lieb (Apex Analytics), Joe Loper (Itron), Monica Nevius (NMR Group), Jeremy Petersen (Xcel Energy), Rick Ridge (Ridge & Assoc.), Mitch Rosenberg (DNVGL), Jean Shelton (Xcel Energy), and Elizabeth Titus (Northeast Energy Efficiency Partnership).

Acronyms

BCE	business and consumer electronics
BPA	Bonneville Power Authority
CFL	compact fluorescent light
EM&V	evaluation, measurement, and verification
EPA	U.S. Environmental Protection Agency
LED	light emitting diode
MT	market transformation
NEEA	Northwest Energy Efficiency Alliance
OLS	ordinary least squares
PG&E	Pacific Gas and Electric Company
POP	point of purchase
RPP	Retail Products Platform
SEE Action	State and Local Energy Efficiency Action Network
UES	unit energy savings

Preface and Caveats

The ENERGY STAR Retail Products Platform (RPP) is a newer program model that offers participating retailers a financial incentive for the sale of targeted product categories that meet program-defined efficiency criteria. Deploying appropriate evaluation tools is critical to assessing the validity of long-term investment in the RPP.

The guidelines outlined in this document apply to expected outcomes of a relatively new program model. As discussed elsewhere in this report, no traditional impact evaluations were completed on RPP as of October 2017. Researchers have undertaken pilot evaluations, market-characterization research, baseline research, and assessments of market shifts but because of the newness of the program model, these efforts have not yet resulted in a prescriptive path for how to evaluate RPP. That noted, the completed research and many of the existing tools in the industry's evaluation toolbox do provide insight and strategies that can guide program sponsors as they begin to plan their program-evaluation activities.

Typically, a document such as this is drafted after many years of pilot programs and parallel evaluations of those programs. The most-successful approaches for determining impact then can be codified in publicly available white papers and often become industry "best practices." The RPP is too nascent to put forth such a document. Instead what is offered here are guidelines and key considerations to help design an evaluation of RPP. Over time, as the RPP model expands and as more evaluation work is completed, this guidance and thus this document will be revised or will evolve. In short, this is not a step-by-step plan for evaluating RPP. Instead, this document gives guidance to ensure that the model is evaluable, and includes descriptions of potential methodologies to compare RPP outputs and outcomes against program logic, and discussion of how any measurement of RPP must be considered over the accrual of savings that will not happen quickly but that will accur over a long time horizon.

Both program sponsors and jurisdictional policy drivers need to be involved in deciding on an evaluation approach. A sponsor with a market-transformation mandate could take a very different approach to evaluating RPP than a sponsor that requires more near-team resource acquisition objectives. Careful consideration of program logic in the context of jurisdictional policy drivers should be the first step undertaken by any potential sponsor.

Although there is no current evaluation recipe for RPP, RPP is unique in the volume of evaluation resources available for potential and current sponsors. The U.S. Environmental Protection Agency (EPA) facilitates monthly conference calls focused on evaluation topics specific to RPP. During these calls participants can tap into the expertise of more-seasoned participants, and can gather insights from other program managers and evaluation professionals. These calls will serve as a continuing source for updates on evaluation methodologies, learning, and results that can be invaluable to those involved in RPP, and which will be integrated into future evaluation guidance documents dedicated to RPP.

List of Figures

List of Tables

List of Terms

The following terms have specific meanings in this report on the ENERGY STAR Retail Products Program, but their definitions can differ in other contexts.

Assortment—The list of models of a product that retailers make available in their stores and online.

Baseline—The starting point of a market or product market share absent prior to the start of the RPP intervention. For the purposes of RPP, a baseline should be used as the point of comparison for product sales and market share prior to the program intervention. The baseline becomes the counterfactual for determining the lift, energy savings, and increased unit sales that are a result of the RPP.

Big-Box Chains—Big-box chains are national chains including general merchants, home-improvement stores, and club warehouses.

Buy-Down Programs—For the purposes of this document, buy-down program models are those that directly buy down (reduce) the cost of a product for the retailer with the intent that the retailer will pass those cost savings on to the end-use customers via a lower price at the retail outlet. Buy-down programs target either upstream levers (usually manufacturers), allowing them to offer a reduced cost to retailers and midstream levers, enabling them to lessen the cost to customers; or target distributers enabling them to offer a reduced cost to installers/contractors who, in turn, pass that savings onto their customers. Buy-down programs have the specific intent of affecting the first cost barrier by lowering the product cost at the point of purchase. In these models, customers typically are aware that the cost of the product has been reduced for them.

Counterfactual—A condition or state of the market prior to or absent any program intervention that is then compared to the condition or state of the market after or with the intervention.

Delta—The difference between the energy use of an energy-efficient product and a non-energy-efficient product.

Downstream Programs—For the purposes of this report, downstream programs are those that offer a direct incentive to the end-use customer to reduce the first cost barrier. In downstream models the customer must take action to receive the discount (send in a rebate, participate in an instant rebate). In these models the end-use customer is the target of the intervention.

Golden Carrot Programs—Programs that are used to accelerate the development of efficient products and technologies by providing significant incentives to manufacturers to directly influence performance standards.

Lift—The number of efficient units sold over time, attributable to a program intervention, between the baseline market conditions and the conditions after a period of program intervention.

Logic Model—A graphical depiction and description of the logical relationships between the inputs, activities, outputs, and outcomes of a program.

Market Effects—Changes 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-efficient products, services, or practices and is causally related to market intervention(s) (e.g., programs). Examples of market effects discussed in the *Scoping Study* include increased levels of awareness of energy-efficient technologies among customers and suppliers, increased availability of efficient technologies through retail channels, reduced prices for efficient models, build-out of efficient model lines, changes in design practices, and—the end-goal—increased market share for efficient goods and services.[1]

Market Transformation—A strategy that **intends to induce** long-lasting, sustainable changes in the structure or functioning of a market. This is achieved through a program or policy design that reduces barriers to the adoption of energy-efficient technologies, or markets that sell, distribute, install, or manufacture those technologies to the point where continuation of the same publicly-funded intervention is no longer appropriate in that specific market. These programs and policies also must move forward to bring the next generation of even more efficient technologies, processes, and design solutions to the market.

Midstream Programs—For the purposes of this report midstream program models are those in which program sponsors provide incentives to retailers to encourage assorting, promoting, and selling a greater percentage of energy-efficient products than they otherwise would. Retailers are not required to pass those incentives on to customers, instead they may use the money in whatever way they see fit. In these models the end-use customer typically is hat not aware a market intervention has taken place or that the intervention might have influenced their purchase decision.

Program Qualified Share (or Sales)—The number of products that qualify for a program and the corresponding market share of those products compared to non-qualified products in the same category.

Resource Acquisition—A program that seeks to purchase energy savings in the public interest, often through financial or technical assistance.[2] It is typically understood that resource-acquisition savings cannot be relied upon for lasting changes.[3]

Sales Elasticity—Models that use customer price-sensitivity to assess customers' movement toward a particular product. In these models, sales revenue serves as the dependent variable and price as the major independent variable.

Stock—The number of a given model item on hand in a store.

Theory-Driven Evaluation—An evaluation that is based on the conceptual framework (program theory and logic) of the program, taking into account contextual factors and causal mechanisms.

---

[1] Mitchell Rosenberg and Lynn Hoefgen. "CiEE Market Effects and Market Transformation: Their Role in Energy Efficiency Program Design and Evaluation." March 2009.

[2] http://aceee.org/sites/default/files/publications/researchreports/ie126.pdf.

[3] http://library.aesp.org/resources/Docuworks/file_display.cfm?id=103.

Unit Energy Savings (UES)—The per-product energy, demand, and (if applicable) natural gas savings, typically the delta in energy consumption between efficient or program-qualifying products and inefficient or non-program-qualifying products.

Upstream Programs—For the purposes of this report upstream programs are synonymous with buy-down programs; that is, they are program models that directly buy down (reduce) the cost of a product to the retailer with the intent that the retailer will pass the cost savings onto the end-use customers via a lower price at retail. The upstream lever target typically is the manufacturer of a targeted product and the purpose is to use the buy-down to lower the product cost to retailers who in turn pass that savings onto customers. Upstream buy-down programs have the specific intent of addressing first cost barriers by reducing the product cost at retail. In these models, customers typically are aware that their cost of the product has been reduced.

Executive Summary

This report provides guidance that can be used for estimating energy savings impacts resulting from midstream market transformation programs with specific focus on the ENERGY STAR Retail Products Platform (RPP). For the purposes of this report, market transformation is defined as a strategy that **intends to induce** long-lasting, sustainable changes in the structure or functioning of a market. This is achieved through a program or policy design that reduces barriers to the adoption of energy-efficient technologies, or markets that sell, distribute, install, or manufacture those technologies to the point where continuation of the same publicly-funded intervention is no longer appropriate in that specific market. These programs and policies must also move forward to bring the next generation of even more efficient technologies, processes, or design solutions to the market.

Retail Products Platform is a new program model that offers participating retailers—primarily national, big-box chains that control a large share of the market for the targeted products—a financial incentive for each unit sold within targeted product categories that meet program-defined efficiency criteria. Regulators, program administrators, and stakeholders can begin to assess the validity of the long-term investment in market transformation programs by deploying appropriate evaluation tools, similar to those outlined in this report. These guidelines, however, apply to expected outcomes of a relatively new program model. No traditional impact evaluations have been completed as of October 2017; pilot evaluations, market-characterization research, baseline research, and assessments of market shifts all have been undertaken. Effective evaluation of RPP requires program sponsors, regulators, and evaluators to accept that market transformation programs require a paradigm shift, relying on theory-driven evaluation and benefits that accrue over the long-term.

This report discusses qualitative and quantitative methods for ensuring that the estimated savings for RPP are valid and aligned with the program theory, and that any operational learnings are considered in an evaluation effort. It also discusses several evaluation design and analysis factors that affect the validity of the estimated savings impact: the evaluation design, the length of historical data collection, the estimation method, potential evaluator conflicts of interest, and the importance of sales data. Also included are several additional considerations that will affect RPP over time in including evolving baselines, cost-effectiveness, and the question of program attribution/free-ridership.

The following program evaluation design and analysis methods are recommended for RPP.

- Use a theory-driven evaluation approach based on a pre-established program theory and logic model—to validate the components in the program theory and to determine whether the activities, as implemented, are effective to meet the program goals.

- Set clear and measurable short-term, mid-term, and long-term indicators to be used as a measure of progress over time.

- Establish a baseline against which the program impacts can be measured.

A detailed discussion of these recommendations is included in this report.

1. Introduction

1.1. Overview

Program administrators have offered energy-efficiency programs to end-use customers for three decades to address the challenges of high energy prices; energy security, reliability, and independence; air pollution; and global climate change. These programs historically have used strategies such as subsidies, rebates, and other financial incentives to motivate consumers to install high-efficiency technologies.

Energy-efficiency programs have achieved significant energy savings, totaling up to 10% of the energy supplied in some regions of the country.[4] Simultaneously there have been significant changes in the efficiency marketplace, resulting in dramatic reductions in energy use for common household appliances and some home electronics due in part to changes in federal minimum efficiency standards and increasing efficiency through the availability of ENERGY STAR–labeled products. Because of this general increase in energy efficiency across numerous devices and appliances, the savings available to efficiency programs has decreased —putting pressure on program cost-effectiveness. This increased efficiency is a positive for the industry but the market shares for efficient technologies still have significant room to grow, a challenge that requires program administrators to look at new program models to bridge that gap.

A second challenge is the number of devices people have in their homes. Plug load is one of the fastest load-growth areas. The relatively low cost of many of these items makes traditional downstream or upstream models less effective, because the low cost makes any cost-effective rebate too small to motivate customer change. To effectively address plug-load efficiency, program managers must adapt to this new environment by streamlining program delivery, reducing transaction costs, and implementing new program models.

Market transformation programs like RPP offer a unique opportunity to address the challenge of capturing **all cost-effective** energy resources, within specific product categories and to reduce program costs related to products with smaller or less cost-effective savings opportunities. The increasing number of consumer products and the overall trend among many products of increasing efficiency have converged to create a compelling case for developing program models that can:

- Allow program sponsors to provide smaller per-unit incentives for energy-efficient versions of products

- Provide greater access to and insight from retail sales data sufficient to identify emerging trends in a variety of consumer products

- Link program activities to enhancements in ENERGY STAR specifications

- Provide scale sufficient to affect product assortment and retain retailer engagement.

---

[4] ACEEE Energy Efficiency in the United States: 35 Years and Counting. http://aceee.org/energy-efficiency-united-states-35-years-and.

Together, these features promise the potential for a more efficient and strategic intervention in the retail market for consumer products. The ENERGY STAR Retail Products Platform (RPP) is a nationally-coordinated program that engages multiple efficiency program administrators, major retailers, and staff at the Environmental Protection Agency (EPA) in an effort to increase the retail availability of energy-efficient versions of select product categories through incentives, analysis, and the ENERGY STAR label. The RPP works "midstream" between the consumer and the manufacturer, by targeting retail decision makers through the provision of small per-unit incentives on sales of products for which the market share for ENERGY STAR–labeled versions could be improved. The EPA, in collaboration with program administrators, provides facilitation assistance and develops tools and guidelines to support the program and encourage consistency in implementation.

Retail Products Platform incentives are provided to retailers for sales of products that have been screened and meet specific program requirements. Typically, products are at least ENERGY STAR qualified. It is not unusual, however, for program sponsors to establish tiered incentives that provide greater incentives for products that beat ENERGY STAR by a set percentage (such as 15% or 30% greater than base ENERGY STAR qualification). The long-term goal of ESRPP is to steadily shift an assortment of select product categories in such a way that the overall energy-consumption profile of the entire product category is reduced. In addition to the provision of midstream incentives, RPP provides a mechanism for identifying and verifying shifts in the market that indicate a product's ENERGY STAR specification should be reviewed. Steady improvement in the energy performance of ENERGY STAR–labeled products is an important long-term outcome of the overall RPP.

Through RPP, participants aim to take advantage of the market and competitive forces among big-box retailers to drive increasing lower energy profiles of product categories over time. RPP does not direct retailers to use incentive dollars in specific, predefined ways. Leveraging the expertise of retailers requires letting them decide how best to increase sales of qualified products. Retailers can choose to use their incentive dollars to increase their marketing activities, reduce the price of qualified products to make them more attractive, shift stocking or assortment practices to improve the prominence of and choices among efficient options, tap customer-loyalty programs, or simply improve the net margin earned on qualified products to encourage their selection by merchant staff.

Expected benefits of this approach include the following.

- **Increased market share for qualified products.** Increases in promotions, availability, and improved pricing should all result in steady improvement of overall market share for qualified products.

- **A shift in product assortment toward energy-efficient options.** Increasing and over time, the profitability of energy-efficient options should influence the decision-making process of merchandising staff.

- **Steady improvement in the specification process and efficiency associated with the ENERGY STAR label.** Access to full category sales data and longitudinal tracking should identify opportunities to improve the ENERGY STAR tiers and provide more strategic intervention in the ENERGY STAR specification setting process.

1.2. Evaluating the Retail Products Platform

Evaluating RPP requires a theory-driven approach based on a pre-established program theory and logic model. By establishing a sound program theory and logic, evaluators will be better able to validate whether the components outlined in the program theory are sound, and whether the activities, as implemented, are effectively meeting the program goals.

Clear and measurable short-term, mid-term, and long-term indicators should be used as a measure of progress over time. By establishing indicators upfront (e.g., number of participating retailers, pecent of customers in participating territories, changes in ENERGY STAR specifications) evalutors should be able to determine whether the model is moving in the right direction in terms of the porgram theory and logic. Finally, RPP provides access to full category sales data,[5] something that program sponsors have long sought to provide market context and enable more strategic use of program funds. Using this data, along with other market insights, an evaluation of RPP should be able to establish a first-year baseline against which the post–pilot period data can be compared and the impacts measured. This is discussed in greater detail in sections 3 and 4 of this report.

Alternatives to the theory-driven approach—which involve comparison areas, end-user self-report for determining free-ridership, and price elasticity models—are not applicable to midstream market transformation (MT) programs for several reasons. First, the scale of midstream programs requires action be taken at the retailer-merchant level, meaning that the effects of the intervention could spill into non-participating jurisdictions, and although this is not a bad thing, it is still a challenge when evaluating MT programs. Second, although survey questions for rebate, direct-install, and custom programs can be relatively straightforward (e.g., "Did you purchase and install an energy-efficient refrigerator in the last year?"), those cannot be asked of customers who do not recognize that what is assorted (and thus what they can purchase) has been influenced by the program. A well-designed midstream market transformation program should be invisible to the end-use customer. This also means that the framework to assess attribution needs to shift in the context of RPP and other midstream market transformation programs. Because the customer is purchasing absent awareness that an intervention has taken place, the traditional framing of free-ridership, the customers' opinion of what they would have done absent the program becomes a moot point. Rather, for RPP, the assessment of program-induced impact should be based on shifts in retailer decision making attributable to the intervention. In short, if the evaluation begins by asking about end-user free-ridership, then the evaluation began with the wrong question and the wrong framing. The question instead should be, "Would the retailer have acted absent the program?" making any discussion of free-ridership something that should happen at the retailer/merchant level. Third, because the intervention might have no impact on the price of the targeted products, price-elasticity models which measure customer price-sensitivity to assess customers' movement towards a particular product are not feasible because the price—which is the major independent variable in these studies—might or might not be impacted by the intervention.

---

[5] Full category sales data includes the total sales for all products in a category, both program qualified and non-program qualified.

Again, it is important to note that these guidelines are based on what we know about a new program model that has not yet undergone significant impact evaluation efforts across multiple utilities. Evaluating RPP will require program sponsors, regulators, and evaluators to accept that market transformation (MT) programs must use an approach that focuses on benefits that accrue over the long term. As more work is done to evaluate RPP efforts nationally, the guidance requires updating to address some of the current unknowns and to point to the real-world results and learnings from those efforts.

1.3. Report Scope

This report is specific to RPP, but the concepts covered can be applied to other midstream market-transformation programs similar in nature. Market transformation for the sake of this report is defined as a strategy that **intends to induce** long-lasting, sustainable changes in the structure or functioning of a market. This is achieved through a program or policy design that reduces barriers to the adoption of energy-efficient technologies, or markets that sell, distribute, install, or manufacture those technologies to the point where continuation of the same publicly-funded intervention no longer is appropriate in that specific market. These programs and policies also must move forward to bring the next generation of increasingly efficient technologies, processes, and design solutions to the market.

This document is not meant to apply to buy-down programs—which are designed to directly affect the price of a product at the point of sale. Buy-down programs often are confused or labeled as midstream; however, they are inherently different in that their primary aim is to reduce the first-cost barrier by passing a lower price onto the end user.

1.4. Report Roadmap

The primary objective of this document is to provide guidance on the analysis methods that can be used to estimate the effects from RPP and other similar market-transformation programs. The report is organized as follows. Section 1 provides a background on the theory behind RPP, and market-transformation programs more broadly. Understanding these programs and their impetus provides important context for any readers new to these types of initiatives.

Sections 2 and 3 contain specific recommendations on the design, execution, and reporting of evaluations as well as discussion on some additional considerations regarding RPP as well as the limits to this guidance given the newness of the program model.

1.5. Intended Audience

This report provides guidance for structuring evaluation activities for RPP and other similar programs, and describes the strengths and limitations of approaches that can be used when evaluating these types of programs. The report discusses the paradigm shift required to measure impact from programs that achieve savings over a long time-horizon and require scale at the regional or national level to be successful.

SEE Action developed this guidance document to ensure that the challenges of evaluating RPP are well understood and that, when conducted, evaluations are in alignment with the market transformation theory that underlies them, and with full awareness of the challenges involved. Rather than deploying methods appropriate to standard consumer rebate programs focused on resource acquisition, practitioners should consider the overall hypotheses that guide the program and adapt inquiries to explore evidence of success in short-term, mid-term, and long-term time frames, for multiple products.

The audience for this report is broad and includes program managers and designers looking for insight on how RPP should be planned and designed to be evaluable, to evaluation professionals who could be facing evaluating RPP or a similar program for the first time, to regulators who might need a deeper understanding of the program model, the validity of savings, and the concept of theory-driven evaluation.

2. Theory

2.1. Basic Logic of RPP—Main Levers

As the number of electronic devices in U.S. homes increases, plug load energy use is forecasted to be one of the fastest-growing energy end-uses over the next 20 years.[6] For many of these products, however, the per-unit energy savings of efficient models over non-efficient models are too small to justify incentives sufficient to influence an end-user's purchase decision. At the same time, technological changes and new product standards have increased the baseline energy efficiency of appliances such as refrigerators and clothes washers, reducing the energy savings that programs can claim from efficient models. As a result, it has become more difficult for programs to offer the types of downstream rebates on these products that have been a common part of energy-efficiency programs. In response to these challenges, program administrators have sought new lower-cost approaches to promote the uptake of energy-efficient products.

The Retail Products Platform is one such approach. RPP offers participating retailers—primarily national, big-box chains that control a large share of the market for the targeted products—a financial incentive for each unit sold within targeted product categories that meet program-defined efficiency criteria. Although the product categories RPP currently targets include soundbars, room air cleaners, clothes dryers, room air conditioners, and freezers, the program is designed to allow administrators to add and drop product categories as markets change and new opportunities arise.

The RPP is an adaptation and expansion of business and consumer electronics (BCE) programs implemented in California and four states in the Pacific Northwest between 2008 and 2014, which primarily focused on promoting energy-efficient televisions. The program administrators that offered BCE incentives, notably Pacific Gas & Electric and the Northwest Energy Efficiency Alliance, developed the RPP approach. The U.S. Environmental Protection Agency's (EPA) ENERGY STAR Program also became involved in RPP, serving in a coordinating role. The EPA, current utility sponsors such as Pacific

---

[6] https://www.eia.gov/.

Gas and Electric Company (PG&E), and the program-design consultant Navitas, are encouraging program administrators around the country to launch RPP programs so that program administrators can approach retailers in a coordinated way and increase the program's influence by offering incentives in a greater proportion of each retailer's market.

Instead of motivating an end-user to select an efficient unit, RPP seeks to change retailer decision making to favor energy efficient products in their merchandising decisions. Advocates see two advantages of this approach over a more traditional, downstream or midstream buy-down program design. First, RPP is expected to influence uptake of qualified efficient products and offer a lower per-unit incentive than would be necessary to influence an end-user's purchase decision. For example, although a $20 incentive might not be sufficient to influence an end-user's purchase decision on a refrigerator retailing for $800, that incentive might be sufficient to make that model more profitable for the retailer than a non-energy-efficient model, due the volume of refrigerators sold.

The second advantage advocates see to RPP relative to a traditional buy-down or downstream rebate is that RPP provides retailers an opportunity to draw on a wider range of strategies to increase uptake of efficient products. Although a buy-down's influence is pointedly focused on reducing the price of a qualified product, RPP is focused on altering assortments to favor efficient products. The RPP is designed to allow the retailer flexibility in selecting which strategies to use to sell the targeted products. In this view, retailers are closer to the market than program administrators and thus are likely to be more knowledgeable about which strategies will be most effective.

Figure 1 shows a schematic of direct and indirect sources of influence expected to flow from RPP incentives according to program logic. Under the program theory, RPP incentives will motivate retailers to take three types of actions that will influence both the demand for, and the supply of, efficient products, thus increasing their uptake.

1. Retailers influence sales for efficient products by assorting more energy-efficient models, thus leaving consumers with fewer non-efficient options from which to choose.

2. Retailers also influence the choice architecture in which consumers make decisions by modifying factors such as promotion—defined broadly to include product placement and sales associate training—and pricing.

3. Retailers influence the supply of efficient products by increasing the share of efficient products on the floor by specifically requesting that manufacturers increase the efficiency of their offerings.
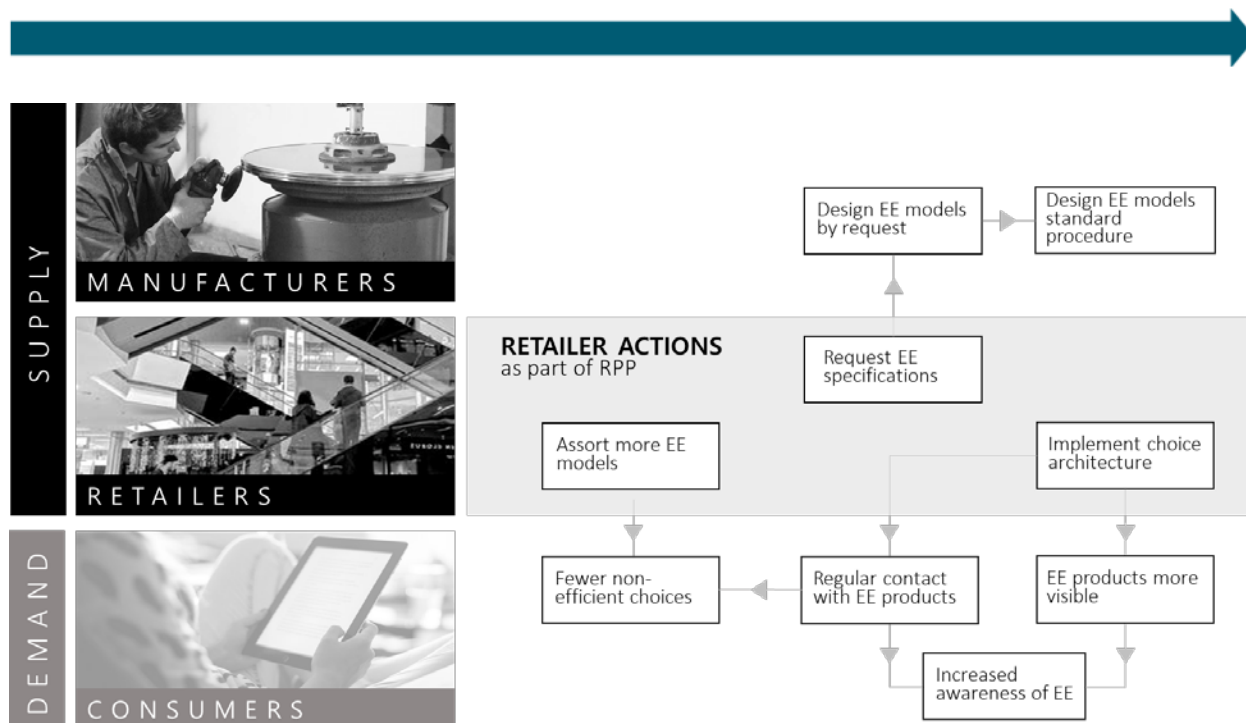
Figure 1. The direct and indirect influence of RPP[7]

2.2. What RPP Is and What It Is Not

The energy-efficiency industry typically views traditional program interventions (incentives and information) and behavior interventions as distinct program designs. Retail Products Platform uses incentives to drive change in practice and assorting at the retailer level. It also relies on merchants to change their ordering and stocking. Although not a traditional behavior-change program, it is useful to think about how RPP might change retailer practices to achieve the desired program outcomes. Specifically, RPP sets out to provide a sufficiently-large pointing mechanism to qualified product models and help increase a retail merchant's interest in these models. Because merchants are interested in increasing profit for their company or buying group, this type of incentive is expected to affect their decision making. Within an evaluation context, identifying the primary practices expected to be influenced through the program will, at a minimum, help to guide questions for merchants and retailer staff.

2.3. Program Stream and the Time It Takes to Observe Effects

RPP differs from traditional programs on two major dimensions, the stream it intervenes in, and the time it takes to observe effects. Traditional midstream and downstream programs tend to yield cost-effective savings almost immediately, and upstream "golden carrot"[8] programs tend to realize savings after a longer ramp-up period. RPP, in contrast, incents in the midstream but, due to long product

---

[7] Schematic derived from Research Into Action 2016; PG&E 2014; 2015a; 2015b; 2015c.
[8] Golden carrot programs are used to accelerate the development of efficient products and technologies by providing incentives to manufacturers. Discussion of a successful golden carrot program for super-efficient refrigerators can be found at https://www.cee1.org/content/golden-carrots-beginning.

buying cycles, the time it takes for program effects to trickle up to manufacturers, and the time it takes to influence ENERGY STAR specifications, program effects take longer to observe. Figure 1 plots several common program types across time and stream dimensions. Between the two extremes lie bounty programs such as refrigerator-recycling programs and golden carrot programs like the front-loading washer programs of the early 2000s which worked closely with manufacturers to produce front-loading clothes washers for the U.S. market. Bounty programs realize savings quickly by targeting end-users' non-efficient appliances. Conversely, golden-carrot programs do not realize savings for several years because the product must be designed, manufactured, and sold before savings can be measured.
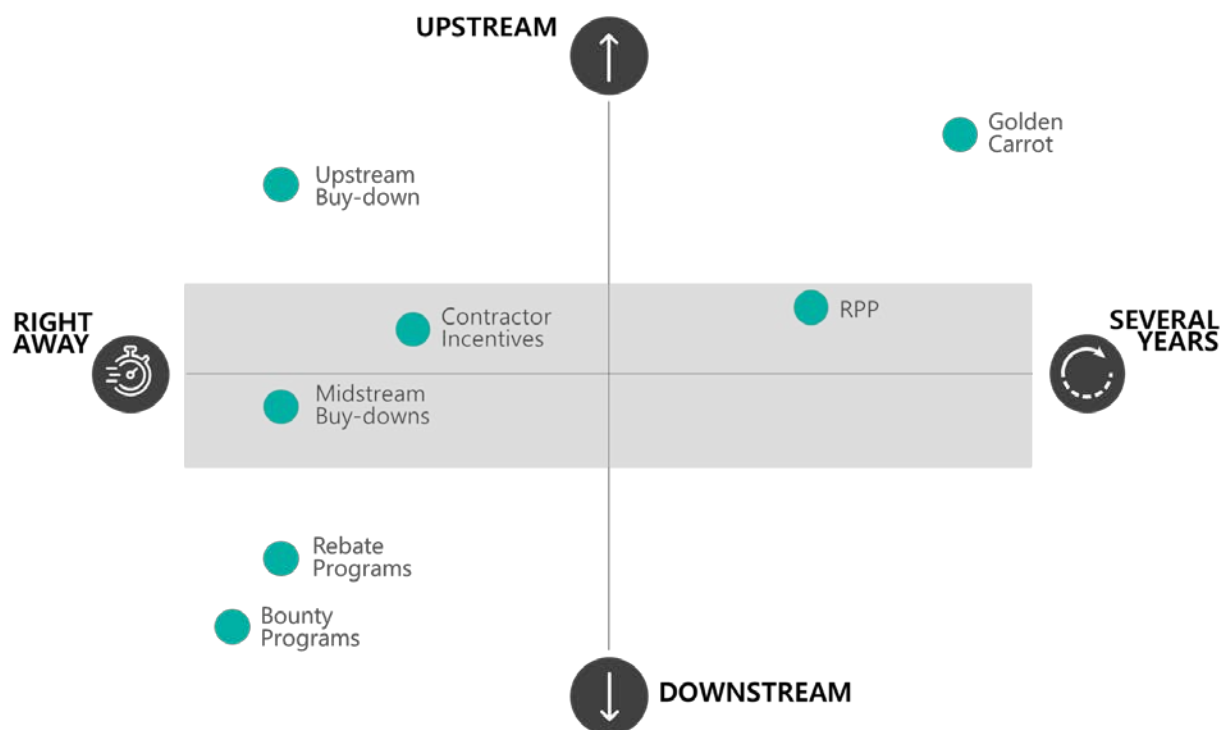

Figure 2. Program designs by stream and time to observe savings

Like midstream buy-down programs, RPP provides incentives for each unit of a qualified product sold by participating retailers. However, the logic of RPP is fundamentally different from the logic of a midstream buy-down. Ultimately, a midstream buy-down program seeks to increase sales of qualified products by influencing the end-user's purchase decision. By reducing or removing the incremental cost of the efficient option, a buy-down program seeks to influence end-users to select and purchase the efficient product over a less-efficient option. Thus, in a midstream buy-down, the retailer becomes a medium through which the program influences the end-user. Buy-down program designs seek to minimize the impact of participation on the retailer; they do not seek to alter the retailer's practices and are likely to have a relatively minor impact on the retailer's internal assessment of business performance.

Although its initial effects will take longer to appear than downstream and midstream buy-down programs, RPP's effects are expected to increase over time as the assortment of qualified products

steadily expands consistent with RPP's status as a market-transformation program.[9] As retailers gain experience with RPP incentives, they could incorporate consideration of a product's energy efficiency into a wider range of their business decisions, including asking manufacturers to design products to ENERGY STAR specifications.[10] As a result, manufacturers might recognize increased retailer demand and prioritize energy efficiency in their product design decisions. All of this further has the potential to lead to more-stringent ENERGY STAR specifications, and potentially energy codes, serving to transform the retail market for targeted products.

2.4. Assessing the Market Barriers RPP Can Address

Market-transformation programs like RPP could play a role in achieving long-term energy efficiency goals for program administrators, allowing consumers to recognize energy-efficiency benefits without the rebate hassle. But these programs are not without challenges. To help understand market transformation, we must understand the market barriers to such programs. Thus, it is important for market-transformation programs to clearly define the market barriers they will address and how. As evaluators, it is then necessary to conduct a market analysis or characterization to verify that the market barriers identified in program design exist for the products in the program. Based on prior research, evaluators identified three main market barriers—as described by Eto, Prahl, and Schlegel (1996)—that RPP's design addresses: information search costs, organizational practices, and product unavailability (Table 1).[11]

Table 1. Market Barriers and RPP

Eto, Prahl, and Schlegel List of Market Barriers	Barriers RPP is Designed to Address
Information search costs	✓
Performance uncertainties Asymmetric information	
Access to financing	
Split incentives	
Bounded rationality (frustrated by old rules of thumb) Organizational practices or custom	✓
Inability to separate product features—for pricing Inability to reverse an EE decision	
Hassle or transaction costs Hidden costs	
Externalities not visible	

---

9   Prahl and Keating 2014.

10   Cite MPER 2 or 4 that mentions retailer requests for ES TVs to manufacturers.

11   PG&E Gas 3668-G filing.

Eto, Prahl, and Schlegel List of Market Barriers	Barriers RPP is Designed to Address
Mispricing due to regulation	
Product or service unavailability	✓

Evaluators can use available ESRPP market characterization data and logic models, as described below, to test whether the program, as implemented, continues to address product barriers.[12] Market-characterization research might need to be refreshed after several years.[13]

3. Practice

The recommended approaches to evaluating RPP in this document are limited to the minimal evaluation activities required to assess attribution of RPP. Further, it only reports methods in this document that have been used in prior evaluations of RPP as it is currently designed.

3.1. Using Logic Models to Test Attribution

3.1.1. Logic Modeling

Programs are always designed with a purpose. All programs have end-goals, with activities identified to meet those end-goals. Program theory articulates how an intervention intends to result in the end-goal. The program theory not only establishes the program objective and activities to meet those objectives, but also details resources needed (including budget and staffing), anticipated or actual barriers, and activities to address the barriers.

A logic model visually illustrates the program's theory. Logic models link program activities to expected outputs (or direct evidence that the activity has occurred). Outputs then result in an outcome. These results could be near-term (e.g., training offered to retailers results in a retail staff that is more educated on energy efficiency), midterm (increasing retailer awareness of efficient options), or long-term (e.g., the television market in a specific geography is transformed and only sells program-qualifying units) outcomes. For further reading, the "Logic Model Development Guide" by the Kellogg Foundation provides a thorough overview of logic modeling.[14]

Program theory and logic models can be useful for all types of demand-side management programs; however, they are particularly important for market transformation programs like RPP. As noted earlier, leveraging a program to influence an entire supply chain tends to be more complex than using direct

---

[12] Ken Keating, *Guidance on Designing and Implementing Energy Efficiency Market Transformation Initiatives* (Sacramento, CA: California Public Utilities Commission, December 9, 2014).

[13] *See* Ken Keating, *Guidance on Designing and Implementing Energy Efficiency Market Transformation Initiatives* (Sacramento, CA: California Public Utilities Commission, December 9, 2014); and Ed Vine, R. Prahl, and S. Meyers. 2010. "An Approach for Evaluating the Market Effects of Energy Efficiency Programs." *Energy Efficiency* 3, 257–266.

[14] W.K. Kellogg Foundation, "Logic Model Development Guide" (Battle Creek, MI January 2004).

rebates to encourage a measure's adoption among consumers. Therefore, being thoughtful about *why* each activity is important and valuable for program success is crucial.

Fundamentally, there are four components to a program-theory logic model.

1. **Inputs:** What is needed for an activity to take place. Inputs include resources (funds and staffing), information, and tools.
2. **Activities:** Discrete activities are required for the program to function. For example, for RPP, activities could include "obtain and analyze sales data," and "pay incentives."
3. **Outputs:** Considered the immediate results of the activities, outputs are tangible and observable. These typically are considered the "bean-counting' elements of logic models. They often are structured as metrics—including specifics such as numbers and time frame—and do not depend on the market's response to an activity.
4. **Outcomes:** Outcomes are the goals or objectives of the activities and outputs. A logic model can denote outcomes by timeframe, such as short-term, midterm, and long-term outcomes. It is important to note that outcomes are not typically "yes/no," check-box items. Given that outcomes typically are longer-term end-goals, progress against outcome often is measured on a continuum and over time.

Developing program theory and logic models is important in both the design and evaluation stages of a program such as RPP. In the design stage, this ensures that program administrators and planners—such as EPA and partner utilities—are thinking about the intent of all activities and anticipated outputs that will affect and, hopefully, result in, the program reaching its end-goal through outcomes.

Program theory and logic models are equally important for program-evaluation purposes; they provide a roadmap to understanding the thinking behind the drivers of the program and for market-progress indicators that must be measured as part of the evaluation process). Perhaps as just importantly, however, evaluation uses the logic model to validate the components in the program theory and to assess whether the activities, as implemented, are likely to meet the program's long-term goals.

Evaluation considers both the outputs and outcomes established in the program theory and logic modeling process and assesses the program's progress against those outputs and outcomes using both qualitative and quantitative methods. The evaluation team gathers data—such as program records, sales data, and other forms of evidence—to ascertain what has happened and then use that information as an evaluation input. The evaluation team next assesses the outcomes established and—typically through primary data collection and other analysis—ascertains whether, or to what extent, those outcomes are being met. The team also explores how effective the program processes are and determines whether there are opportunities to make changes to those processes to more effectively achieve the desired outputs and outcomes of the logic model.

Assessing progress related to both outputs and outcomes, and understanding the intended link between these areas, are important for testing the validity of the program theory. If the team observes activities that are resulting in the expected outputs but not the intended outcomes of those outputs, then that analysis could indicate that the activity is not yet having the intended impact.

Further, it is through an understanding of the intended cause of and effects of a program model that an evaluation team can develop an appropriate methodology for measuring attribution. Attribution—or proportion of sales attributable to program intervention—should consider all the points of influence on target-market actors (e.g., retailer, customers, manufacturers). The program theory and logic models indicate the points of influence expected to determine whether the program's hypothesis is correct and the interventions are effective. Therefore, program theory and logic models are valuable for informing the methodology (sampling, data collection, and analysis) for assessing attribution.

3.1.2. Retail Product Platform's Basic Logic

Figure 2 provides a simple logic model for the RPP, derived from PG&E, Northwest Energy Efficiency Alliance (NEEA), and ENERGY STAR RPP logic models.[15] Note the components described above and the cause-and-effect links to each other. Primary activities include recruiting and engaging retailers, setting product specifications, providing incentives to retailers, participating in discussions on ENERGY STAR specifications and standards, and collecting detailed sales data. These activities should lead directly to an increase in the number of participating retailers, increased awareness of specifications by retailers, increased number of comment letters supporting changes to standards and ENERGY STAR specifications, and a comprehensive database of retailer sales data. If program logic follows, then these activities should result in retailers considering incentives in assortment and marketing decisions, reliable market share data to inform program design and product selection, an increase in the proportion of efficient models sold, increased retailer requests to manufacturers for efficient designs, increased ENERGY STAR qualifying criteria, and more-efficient federal standards. Also note, however, that each program sponsor should develop their own logic model. Although this example can serve as a guidepost, sponsors must define logic and create outcomes and outputs that align with their unique program or utility business model. Once defnied, sponsors then should create measurable metrics that are specific to their logic model.

The Retail Product Platform's influence spans three retailer segments: sustainability staff, the staff that interface with energy-efficiency programs; merchants, those who make decisions about product assortments; and marketing staff, staff members who make marketing decisions. The RPP also has the potential to influence manufacturing staff and the EPA staff members who work on ENERGY STAR specifications.

---

[15] Northwest Energy Efficiency Alliance (NEEA) (September 13, 2017), "Portfolio Products Market Test Assessment" (Portland OR), http://neea.org/docs/default-source/reports/rpp-market-test-assessment-report.pdf?sfvrsn=4; and ENERGY STAR https://www.energystar.gov/ESRPP.
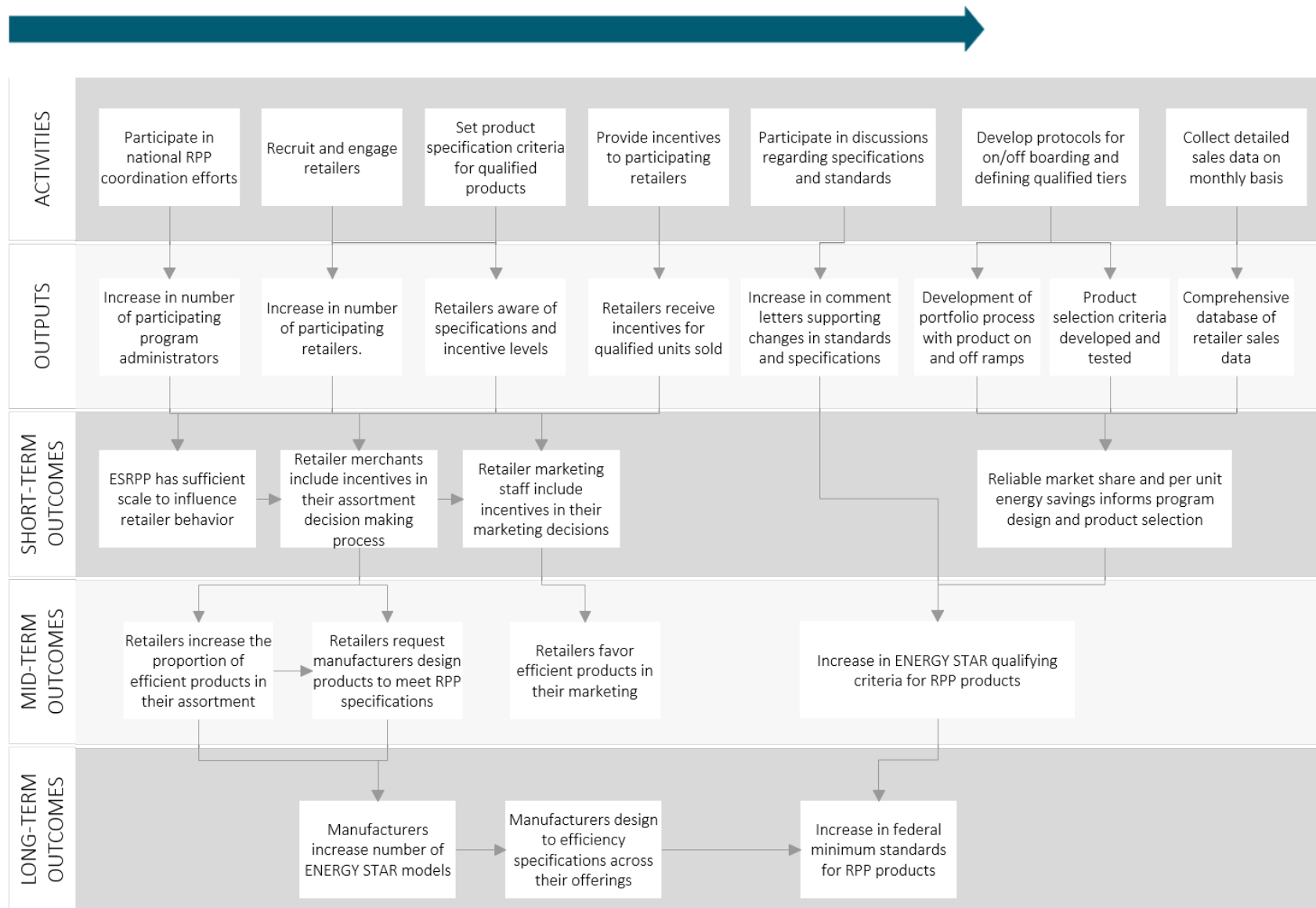
Figure 3. Generic RPP logic model

3.2. Current Methods to Assess RPP

3.2.1. Interviews: Why Interviews Are So Critical

There is considerable potential for uncertainty in estimates of market conditions; it is not possible for a quantitative counterfactual to account for all the conditions that might influence product sales, as discussed below. Thus, it is hoped that qualitative data from key market actors will play an important role in both assessing attribution and supporting the estimates of market lift that a quantitative analysis generates. Qualitative research supports this in three ways, as described below.

First, qualitative findings can verify that the conditions exist for program influence to occur. For example, for program incentives to influence retailer behaviors around assortment and promotion of efficient products, the retail staff members responsible for those decisions must be aware of the incentive at the time the decision is made, and the retailer must allocate the incentive funds in a way that benefits the staff members responsible for those decisions. If these conditions are not met, any observed increase in sales of qualified products is likely to reflect some market change independent of the program.

Second, qualitative data can provide a sense of the role that incentives play in retailers' decisions. Retailer decisions around product assortment and promotion are both complex and sensitive. As a result, retailers are likely to be unwilling or unable to provide detailed specifics about the way program incentives influenced specific decisions to quantify the impacts of those decisions. Past studies, however, have found that retailers can describe, in general terms, the way program incentives influence their decisions.[16] For example, retailers have described program incentives as being a tie-breaker in assortment decisions, leading them to select the efficient option when choosing between two otherwise similar models. Understanding that incentives factor into retailer decisions in this way provides support to any observed sales lift.

Finally, qualitative data can identify specific actions that retailers have taken to more precisely focus quantitative analyses. Although retailers might not provide detailed descriptions of all the ways program incentives have influenced their decisions, qualitative research could reveal certain, specific promotions or actions retailers have taken to increase sales of qualified products. Evaluators then might be able to focus more granularly on identifying the sales impacts of those specific actions. These individual actions will likely account for a small portion of the overall program effect, but their more precise measurement can provide context for larger-scale, but less concrete, estimates.

Fundamentally, incentives such as those RPP offers seek to influence merchant purchasing practices, with the expectation that the retailers will then act to shift end-user purchase decisions toward efficient models. Qualitative research with retailers and other key market actors allows for an assessment of the program's effects closer to this primary point of influence than does an analysis focused on sales data alone. This report provides some example interview topic areas for retailer sustainability staff (the

---

[16] http://neea.org/docs/default-source/reports/rpp-market-test-assessment-report.pdf?sfvrsn=4; http://neea.org/docs/default-source/reports/consumer-electronics-television-initiative-market-progress-evaluation-report-2.pdf?sfvrsn=8.

retailer staff interacting with RPP on a daily basis), merchants (retailer staff making purchasing decisions), and marketing staff (retailer staff making marketing decisions). Because sustainability staff often are RPP's interface with retailers, evaluators can ask sustainability staff all questions regarding marketing and purchasing. Conducting interviews with key manufacturers to assess RPP's influence on their business practices also is recommended.

Table 2. Interview Topics

Understand the buying cycles for each product incented in RPP.
What is the typical buying cycle for [product]? Is the buying cycle annual, bi-annual, ongoing, or cyclical with a longer-term refresh?

Understand the buying process for each product.
How do merchants make decisions on model selection for [product]? When do merchants make purchasing decisions for [product]? What information do merchants use to make purchasing decisions? How and why are certain products selected for in-store sale versus online only? Do incentives count toward the merchant group's profits? What information do sustainability reps pass on to merchants? What program information do merchants see – incentives, projected incentives, model numbers?

Understand how merchants balance purchases of new products versus product refreshes.
How often do merchants refresh models for [product] already in stores? How often do merchants consider new models not currently sold by [retailer]? What is the perceived risk for purchasing new models versus refreshing currently assorted models?

Understand the level of influence incentives have on merchant model selection and marketing efforts.
How do merchants consider incentives from manufacturers in their purchase decision? How do merchants consider utility incentives in their purchase decision? How do retailer marketing staff choose products to promote? What types of incentives to marketing staff consider when making marketing decisions? How do marketing staff consider utility incentives when making marketing decisions?

Note that structuring questions so that the answers can be compared over time is critical, This is not to suggest taking a Likert-scale approach, as staff turnover at big-box stores can make numbered scales

difficult to compare. Rather, ensuring questions are set up to thoughtfully probe how, over time, merchant purchasing behaviors are or are not changing as a result of RPP.

3.2.2. Implementation Plans and Shelf Surveys

The current RPP program model includes the collection of "implementation plans" from participating retailers. These plans are designed to document retailer marketing and promotion efforts in respect to RPP-targeted products and to confirm that marketing is occurring. Provided annually to the RPP sponsors via the EPA coordination, the plans serve to provide evidence that the retailer acknowledges some action on its part tied to the incentives it is receiving.

Implementation plans, although useful, are not sufficient to determine attribution and should not be used to do so. In the implementation plans, retailers could note what they have done to promote qualified models, but evaluators have no way of telling how much, if any, of the plan content retailers were planning to do outside of RPP. Further, the plans do not detail how retailers will change assortment—one of the program's biggest levers. Implementation plans are most useful as a qualitative data point, indicating that the retailers are taking steps to market qualified products and increase their sales.

Shelf surveys—a process in which a field team catalogs the products on the shelf in a given store and documents changes over time—also can provide an assessment of progress along the logic's continuum. Although not necessary to evaluate an RPP, if administrator/RPP sponsors include store-level interventions—such as point-of-purchase (POP) activities—then shelf surveys could assist evaluators in assessing the effectiveness of these elements. For program administrators that believe their regional assortment could differ from national assortments, shelf surveys can also enable evaluators to determine variability in their region's store-level assortment.

Neither implementation plans nor shelf-surveys are a means for assessing attribution. Rather, evaluators should use these as data points to confirm that the hypotheses of market outcomes in the program logic model are visible.

3.3. Baselining to Quantify Effects

3.3.1. Basic Logic of Baselines

In shifting from a resource-acquisition framework to a market-transformation framework, the unit of analysis for impacts also shifts from the individual utility program level to the broader market level. Tracking and analysis of market data therefore also must shift to a market level. Thus, the evaluation of market-transformation programs relies heavily on establishing a baseline against which the program impacts can be measured. Unlike resource-acquisition programs, market-transformation evaluations require more upfront coordination between the evaluation and implementation teams, data-collection needs must be clarified prior to launch, metrics established, short-term, midterm, and long-term market indicators defined. Without early and closer coordination, sponsors risk developing indicators that cannot be measured or collecting data that does not meet evaluation needs. Additionally, a comprehensive market study must be conducted to establish the market's baseline conditions. It

typically is more cost effective to gather this baseline data at the program's onset or early in the program implementation period. Because RPP is starting with six product categories and could increase from there in subsequent years, setting rigorous baselines could be difficult to do all at once, therefore discussions with regulators about acceptable levels of detail are necessary.

During the initial implementation of the program, a formal market assessment should be used to establish the baseline levels of the market indicators, such as the current market share of the incented products, retailer awareness of energy efficiency levels, and current stocking and promotional practices. In some jurisdictions, market progress evaluations carried out by independent contractors are conducted regularly to track market changes against the initial baseline. In other segments, progress is tracked through more informal means, such as analysis of program records, contacts with supply-side market actors, and reviews of market-share statistics prepared by vendor organizations or other government agencies at the local, regional, and national levels.

Baselines and progress commonly are tracked using market-transformation indicators. Indicators are chosen based on the expected outcomes from the program logic model; the information needs of the evaluation audiences; and the availability, cost, and timeliness of the data. The expected market changes occur at different stages of the intervention.

Evaluators historically have taken two approaches to determine the influence of midstream and upstream programs—the establishing of baseline sales of efficient products through either historical comparison or geographic comparison. Both approaches can be problematic for RPP.

Geographic comparisons examine the sales of efficient products under the program across geographic regions. For upstream buy-down programs, in which a given retailer provides lower prices for a product in region 1 but not in region 2, this approach is viable. For RPP, however, where research[17] has confirmed that retailer assortment decisions are made at a national level with little to no geographic variation in stores, this is not a viable approach to capture program impacts. Geographic comparisons for RPP only should be used to estimate impacts of in-store marketing efforts where in-store implementation varies across the geographies chosen.

Historical comparisons evaluate sales of efficient products under the program relative to a defined period before the program's launch to determine changes in sales. This approach has the potential to capture the effects of all the actions a retailer takes to increase sales of energy-efficient products, but it also runs the risk that changes in the market independent of the program will limit the relevance of the baseline period as a comparison. Although changes to widespread economic conditions or more-targeted supply-chain disruptions could affect the applicability of a historical baseline to any product, this risk is particularly acute for product categories that have short product-refresh cycles and those undergoing rapid technological change.

For example, televisions experience rapid product-refresh cycles, with manufacturers replacing their entire range of product offerings each year. Also, over the past 15 years, televisions have undergone a

---

[17] Cite ESRPP interviews from 2016

period of rapid technological change, with flat-screen technologies replacing cathode ray tubes, LCDs overtaking plasma, and light emitting diodes (LEDs) replacing compact fluorescent lights (CFLs) as a backlighting technology. Given this rapid change, sales of products meeting a particular efficiency level in one year would be of limited benefit in estimating a program's influence on sales of efficient products in subsequent years. In contrast, some models of pool pumps remain on the market relatively unchanged for 20 years, making a historical comparison of sales much more relevant (Research Into Action 2013). In any case, program administrators using a historical-comparison approach should closely track market and technology changes that might limit the relevance of the baseline period.

Program administrators and evaluators are building on the more-traditional historical baseline and comparison area approaches to assess the influence of RPP. One approach builds on the historical baseline approach to create mathematical models predicting the diffusion of efficient technologies over time. Other evaluations have successfully identified impacts from specific activities to promote efficient products by using targeted analyses comparing periods immediately before and after limited-time promotions, or comparing stores that received promotions with those that did not. In one case, these promotions implemented an experimental design to facilitate this analysis.[18]

The forecasted baseline then is compared to the actual post-implementation program-period sales data. The difference between the program-period data and the forecasted baseline is the net effect of the program. Figure 4 illustrates this approach. Given that this is proposed and managed as a mass-retail and multiple-product program, although sales measurements can be made at the measure level the program impact will be derived as the sum of the products at either the product level or the product subcategory level. In this way, variation in product shipments or unexpected results does not jeopardize the product portfolio.

---

[18] NEEA Consumer Electronics Television Initiative Market Progress Evaluation Report #1. http://neea.org/docs/default-source/reports/consumer-electronics-television-initiative-e11-230_combinedv2.pdf?sfvrsn=13.
NEEA Consumer Electronics Television Initiative Market Progress Evaluation Report #2. http://neea.org/docs/default-source/reports/consumer-electronics-television-initiative-market-progress-evaluation-report-2.pdf?sfvrsn=10.
NEEA Consumer Electronics Television Initiative Market Progress Evaluation Report #3. http://neea.org/docs/default-source/reports/consumer-electronics-television-initiative-market-progress-evaluation-report-3.pdf?sfvrsn=7.
NEEA Television Initiative MPER #4. http://neea.org/docs/default-source/reports/television-initiative-mper-4.pdf?sfvrsn=4.
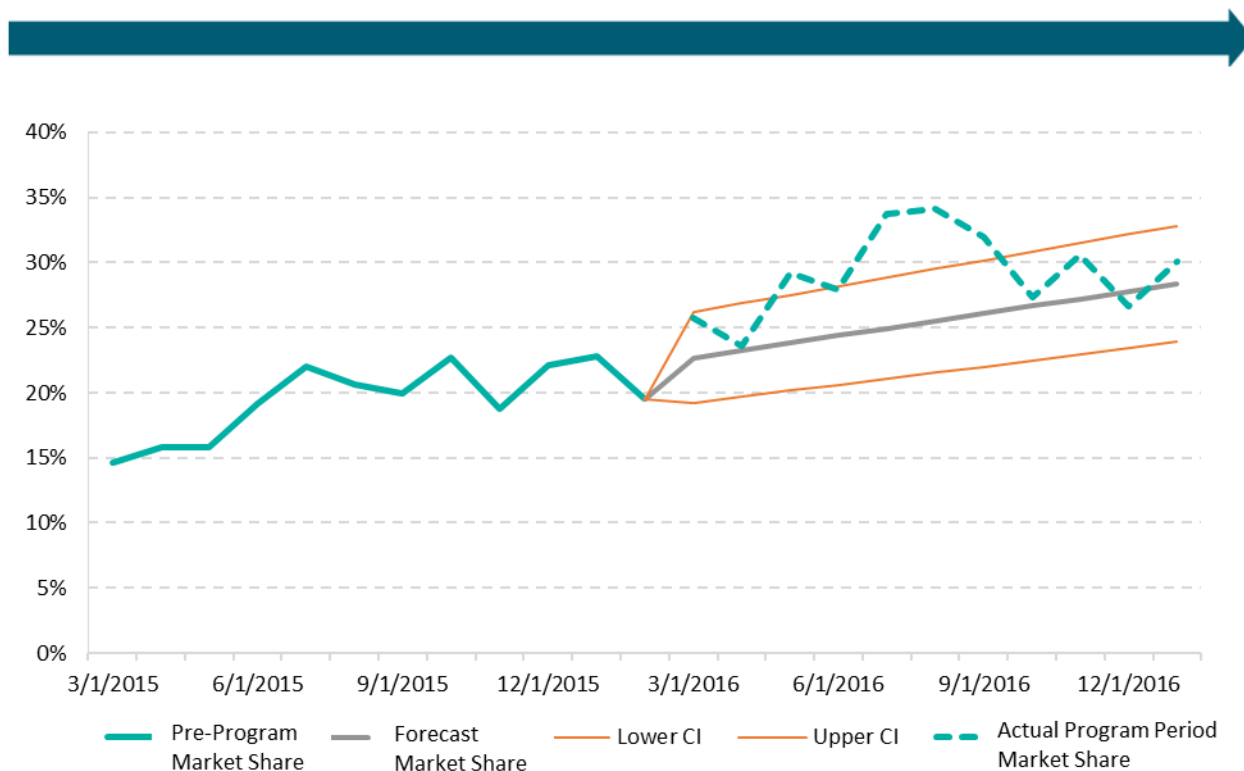
Figure 4. Example of forecasted baseline

3.3.2. Baseline Approach

To estimate ESRPP baselines in the short term and midterm, forecasting the market share of efficient products without program intervention is recommended. As part of its design, the RPP has a data portal that is designed to support this type of analysis, and prior evaluations[19] have shown that this approach is feasible as a starting point for RPP program administrators to estimate baseline sales. Note that no evaluator has yet tested a method to address change in baselines after the first several years of program implementation. Specifically, no evaluation has yet to address the change in the baseline that occurs once the "pre" sales data used to forecast sales are sullied by sales data from program implemented years. In other words, evaluators have not yet attempted to address how to create forecasted baselines using sales data that were influenced by the program incentives. Sales of products vary across years, therefore after several years the one year of pre-program data cannot continue to be used to estimate baselines. This issue has yet to be addressed by evaluators, but two regional organizations, NEEA and Bonneville Power Authority (BPA), have attempted to address these issues when designing and implementing programs. To further discussion in this area, the Appendix presents the approach used by BPA. Because no third-party evaluation of RPP has proceeded far enough to attempt long-term baselining approaches, these are included as alternative approaches and not as recommended approaches. These do enable a beginning discussion of potential long-term baselining approaches with

---

[19] NEEA Retail Products Program Market Test Assessment. http://neea.org/docs/default-source/reports/rpp-market-test-assessment-report.pdf%3Fsfvrsn%3D4&ved=0ahUKEwjp5vvJiYXXAhWIq1QKHU3IBgoQFggoMAE&usg=AOvVaw3ZRucGb28pXnJjZ46isEoy.

the full awareness that, as evaluations are completed and programs evolve, the guidance laid out in the document also likely will evolve.

The evolving baseline issues aside, given the limitations of forecasts based on historical data, it is important for baseline estimates to account for additional factors such as technological and market changes. The analysis should focus on forecasting the market share of qualified products, defined as the percentage of program-qualified sales relative to overall reported sales.[20] As a first step, evaluators can run simple comparisons of means (t-test) between the pre-program sales data and data from the program period for the products available for the analysis. Performing a t-test can help to determine whether the mean annual market share is statistically different between pre-program and program periods.

Next evaluators can run an ordinary least squares (OLS) regression to determine the influence of (1) time, and (2) the program, on the market share of qualified products. Including the time variable in the model is important, as this helps account for the general trend of market share in the period before program intervention. The functional form of the regression is shown in the equation below.

$$Y_t = \beta_0 + \beta_1 Time_t + \beta_2 Program_{t,i} + \varepsilon_t$$

Where:

Y_t = Market share in month t

$Time_t$ = Number of months from the beginning of the historical sales stream

$Program_{t,i}$ = Dummy (binary) variable representing pre-program (0) and program period (1)

β_0 = Model Intercept

β_1 = Secondary coefficient, represents the time-sensitive monthly change to the market share

β_2 = Primary coefficient of interest. Represents the change in market share due to program

ε_t = Model error term

To estimate the counterfactual—or what the market share of qualified products would have been absent the program—the evaluation team can develop two separate forecasts: a forecast based on exponential smoothing, and a forecast based on including only the time variable given in the equation above (removing the program-based variable). The advantage of an exponential smoothing forecast is that it places the greatest weight on the most recent month, with each previous month receiving exponentially less weight in the forecast. If there were clear indicators of seasonality (as in the case of seasonal products such as air conditioners) or if the evaluation teams have access to additional years of

---

[20] As defined in the ICF online retailer product database dashboard, "The penetration is calculated by taking the total qualified products and dividing by the total number of products sold per category. Unverified products are not included in the calculation." https://www.retailproductsplatform.com/.

sales data that show evidence of seasonal or monthly sales trends, then these factors also could be added to the model.

To estimate potential market lift associated with program activities, the output from the modeled sales forecast with the program intervention should be compared with the modeled forecast without program intervention. The difference between these two forecasts represents the absolute program-induced lift in market share of qualified products. As an example, if the post-intervention market share was 40% and the non-program forecasted market share was 20%, then the absolute market lift would be 20%. The relative program-induced sales lift is represented by the absolute percentage of market share, classified as lift divided by the actual program market share. Using the same example, the relative program sales lift would be 20% divided by 40%, or 50% (i.e., 50% of the program-period sales reflected program influence). As a cross-check, the evaluation teams can compare the modeled market share with program intervention to the exponentially smoothed market-share forecast for the same period without program intervention.

3.3.3. Potential Limitations of a Forecasted Baseline Approach

Relying on sales data as the basis to form the baseline condition, although the preferred method, still involves a substantial degree of uncertainty. Evaluators have identified some key factors that limited the comparability between the historical sales data and the data submitted during the evaluation period.

- Incomplete sales data

- Changing qualification specifications (ENERGY STAR)

- Cannibalization across two efficiency levels

The sections that follow provide additional detail on each of these limitations, as well as additional factors having the potential to complicate the process of developing a baseline from historical sales data.

3.3.3.1 Incomplete Sales Data

A prior evaluation of RPP for NEEA found incomplete data from one retailer.[21] For both freezers and dryers, the number of unique models included in the retailer's historical data is considerably less than the number included in data submissions during the program. In contrast are other retailers' reported sales of a much more consistent number of unique models before and during the program (Figure 5).

---

[21] NEEA Retail Products Program Market Test Assessment. http://neea.org/docs/default-source/reports/rpp-market-test-assessment-report.pdf%3Fsfvrsn%3D4&ved=0ahUKEwjp5vvJiYXXAhWIq1QKHU3IBgoQFggoMAE&usg=AOvVaw3ZRucGb28pXnJjZ46isEoy
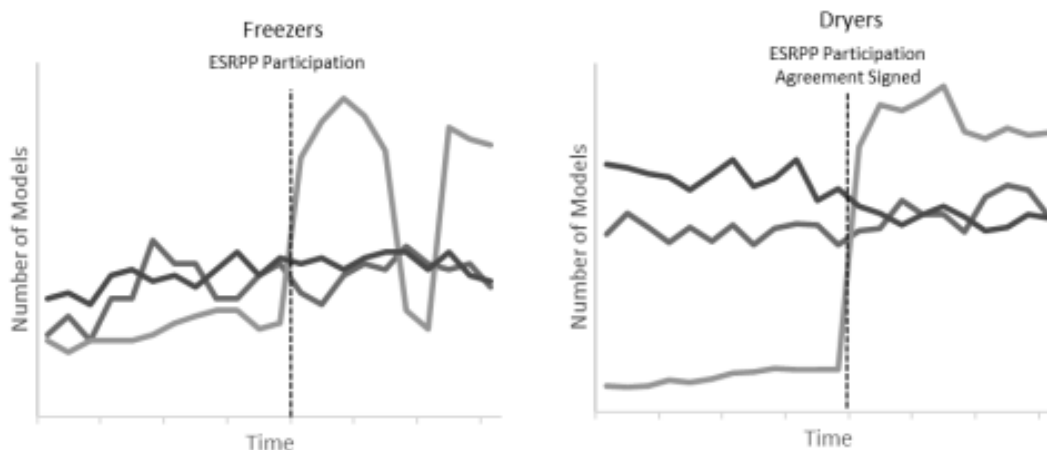
Figure 5. Demonstration of data summarization issues: Retailer pre-program versus program-period sales inconsistent reporting

This type of inconsistency in the pre-program data requires removal of that retailer from the analysis to estimate baseline market shares or efforts to obtain missing data. Further, potential for issues with an individual retailers' data submissions could compound evaluation issues. Note, however, that these issues should be resolved prior to evaluation. If the evaluation work identifies new gaps, those become process finding and should be flagged for correction. Because of the aggregated nature of much of the data in the RPP data portal,[22] having to remove one individual retailers' data also could limit the analyses that are possible. For example, the recent NEEA evaluation could not forecast baseline market shares by product tier because those data were not available at the retailer level for non-qualified products.

3.3.3.2 ENERGY STAR Specification Changes

To be relevant as a comparison against models sold during the program, a baseline estimated from historical data must estimate the market share of products meeting the same specification the program incentivized. The RPP data portal, however, currently lists models as qualified based on the ENERGY STAR specification in place at the time they were sold. As a result, if the ENERGY STAR specification changes during the year before a product's addition to the RPP portfolio, then products classified as basic or advanced tier in the data portal might not meet the same specification as products with the same classification sold during the program. For example, the room air conditioner specification changed in October 2015. As a result, an air conditioner sold in July 2015 that the data portal classified as basic tier would have qualified under the ENERGY STAR Version 3.0 specification, and an air conditioner sold in July 2016 and classified as basic tier would have qualified under ENERGY STAR Version 4.0. Thus, lags in product sell-through could lead to products being incorrectly categorized as

---

[22] The RPP data portal currently aggregates sales data so evaluators only can examine model-level sales data across retailers. Retailer-level data allow for granular analysis of qualified models only. These restrictions, although preserving retailer market intelligence, are a limiting factor in calculating retailer-level estimates.

specifications change. Utilities should work with the program-data implementer to identify a solution to this challenge.

3.3.3.3 Cannibalization Across Two Efficiency Levels

To drive the market to greater efficiency levels, RPP defines two qualification tiers for some products: a basic tier and an advanced tier. Ideally, baselines are estimated independently for each tier, which is what was done here. In some cases, however, retailers' actions could motivate consumers to select an advanced-tier model rather than a basic-tier model, or vice versa, rather than simply choosing a qualified model over a non-qualified model. Thus, one tier might cannibalize the market share of the other tier. Figure 6, for example, shows historical air cleaner sales across the pre-program and program periods. Although basic-tier sales decline from 2015 to 2016, advanced-tier sales increase. Overall, qualified sales (including both tiers) decline slightly between the pre-program (2015 sales, at 88%) and program (2016 sales, at 80%) periods (as denoted by the dashed lines below). This potential cannibalization complicates any analysis of the program's influence on sales of each tier of qualified products.
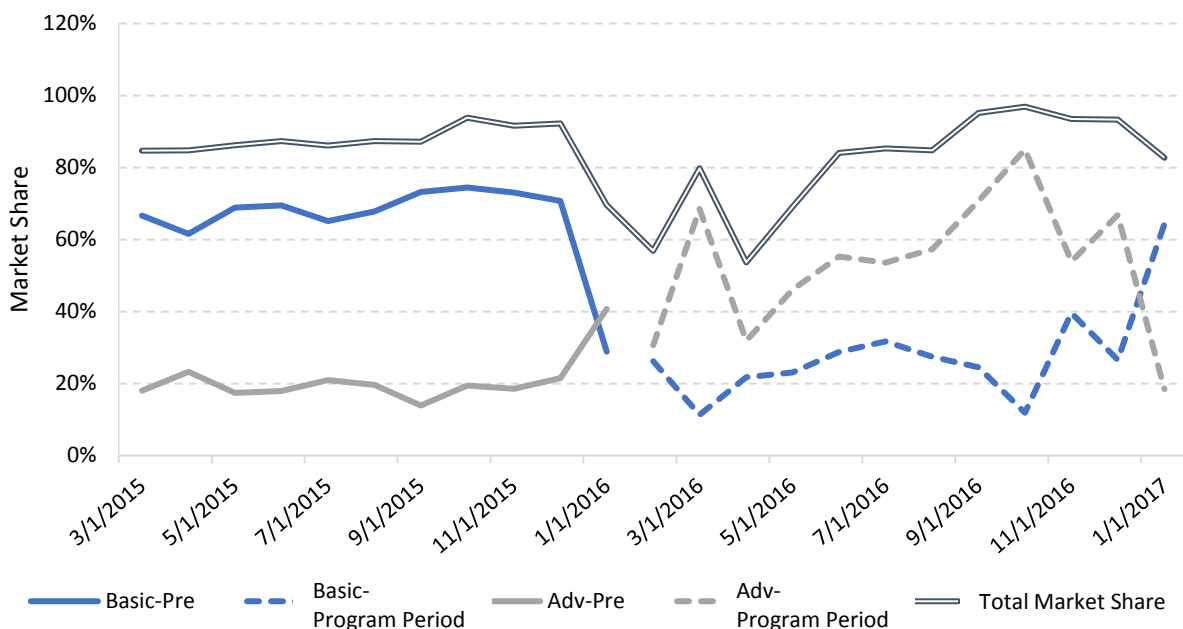

Figure 6. Demonstration of data summarization issues: historical market share of qualifying air cleaners

4. Additional Challenges

A variety of other factors exogenous to the program influence market-level sales, including—at a high-level—competitive retailer and manufacturer behaviors, the retail environment, and the general economic conditions at the time. In many cases, these factors affect sales of both efficient and inefficient products. Thus, although they might affect the volume of qualified products sold, they will have a more limited impact on market share of qualified products. Nonetheless, it is possible that

certain factors can act more strongly on some segments of the market than others. Some more-specific, non-program, market-level factors that affect sales include the following.

- Seasonality—this is particularly evident for products like air conditioners (where more than 97% of sales occur between April and September).

- Weather—this is particularly evident for products like air conditioners, where consumers are considerably less price and product sensitive during extreme heat wave events and tend to purchase whatever is available in the store coincident with a heat wave.

- Promotions—these could include retailer- or manufacturer-specific marketing activities independent of the program.

- Product lifecycle—the degree with which model assortments are turned over and the number of unique models.

- Halo effects—this represents the bias shown by customers toward certain products because of a favorable experience with other products made by the same manufacturer.

- Low sales volumes—some products experience lower sales volumes, and therefore sales data shows considerably greater variance across monthly market share of qualified products (a small shift in qualified sales could have a great impact on the market share of qualified products).

5. Cost Effectiveness

A new program design like RPP brings up some unique challenges regarding determining cost-effectiveness. Although providing specific guidance on cost-effectiveness is not a purpose of this document, a quick discussion on how the how the components of RPP is considered in a cost-effectiveness analysis is worthwhile. Regardless of the jurisdictional test of choice, most often the analysis of program costs to benefits is examined in terms of first-year savings to first-year costs, with inputs for costs of energy, incremental costs (compared to the inefficient choice) of included products, program implementation, marketing, and sometimes even evaluation tallied on the cost side. These sides of the equation come together to show how much benefit will be seen for the cost investment in the program. With RPP it is widely acknowledged that sales will be limited in the early years, that the first-year costs will, in fact, more likely influence sales (and thus savings) in years two or three than they might in year one. As such, any cost-effectiveness analysis must look at how to measure the cost investment in each year over what could be both a delayed influence (none during the year the money was spent) and a multiyear of influence (sales that occur in out years from spending in previous years). For example, in retailer discussions it has been verified[23] that for product categories such as washers and dryers the assortment could change every 18 months or so, and the order lead time can be as much as 6 months. This means that the program costs that influenced a change in assortment might not result in any savings for 12 to 18 months and those savings/sales could continue to accrue for 18 months beyond the program period.

---

[23] EPA retailer interview report by Cadmus.

A potential advantage for RPP in the context of cost-effectiveness is the coordinated nature of the program. By working collaboratively, sponsors are able to share the costs of services such as data aggregation, retailer coordination, and product criteria development. As these processes are established and mature, they in theory should lower per-unit administrative costs over time. As RPP moves forward, evaluators and program planners should consider how to best reflect the costs and benefits in their analysis to take a longer view and carefully consider the value of investments today that lead to long-term national shifts in market share. This task—which is unlikely to be easy—could serve as another point at which collaboration across sponsors and evaluation teams will be needed.

6. References

Barnes, H., L. M. Rockville, MD, W. Steigelmann. January 2009. "Market Transformation? Resource Acquisition? Not Really. It's Market Share!" Prepared for AESP.

Bonneville Power Administration. Market Research and Analysis 2010–2015. Retrieved from www.bpa.gov/goto/MomentumSavings.

Cadmus Group, Inc. October 2016. "ESRPP Pilot Interview Findings." Prepared for EPA & ENERGY STAR.

DeHoratius, N., Research into Action. March 2013. "Consumer Electrics Television Initiative Market Progress Evaluation Report #2." Prepared for Northwest Energy Efficiency Alliance (NEEA).

DeHoratius, N., Research into Action. July 2014. "Consumer Electronics Television Initiative Market Progress Evaluation Report #3." Prepared for NEEA.

Eckert, J.B. July 1995. "The Super-Efficient Refrigerator Program: Case Study of a Golden Carrot Program." Prepared as an account of work sponsored by an agency of the United States Government.

Energy Market Innovations, Inc. November 2011. "Consumer Electronics Television Initiative Market Progress Evaluation Report #1." Prepared for NEEA.

Eto, J., R. Prahl, and J. Schlegel, July 1996. "A Scoping Study on Energy-Efficiency Market Transformation by California Utility DSM Programs." Prepared for The California Demand-Side Measurement Advisory Committee.

Keating, K, November 2014. "Guidance on Designing and Implementing Energy Efficiency Market Transformation Initiatives"

Nadel, S., N. Elliott, and T. Langer. June 2015. "Energy Efficiency in the United States: 35 Years and Counting." Prepared for ACEEE.

Prahl, R., and K. Keating. October 2014. "Building a Framework to Support Energy Efficiency Market Transformation in California." Prepared for the California Public Utilities Commission.

Randolph, E. February 2016. "Advice Letter PG&E 3668-G/4765-E." Request for Authority for Retail Products Platform (RPP) Pilot within PG&E's Residential Energy Efficiency Plug-Load and Appliances Sub-Program.

Research Into Action. April 2013. "Consumer Electronics Television Initiative Market Progress Evaluation Report #2." Prepared for NEEA.

Research Into Action. 2016. Pacific Gas & Electric Company. 2014, 2015a, 2015b & 2015c. Retrieved from web.

Research Into Action, and Apex Analytics. July 2015. "Television Initiative MPER #4." Prepared for NEEA.

Research Into Action, APEX Analytics, LLC, and ILLUME Advising, LLC. September 2017. "Retail Products Portfolio Market Test Assessment." Prepared for NEEA. Web.

Rosenberg, Mitchell, and Lynn Hoefgen. March 2009. "Market Effects and Market Transformation: Their Role in Energy Efficiency Program and Evaluation." Prepared for CIEE Market Effects Program.

Taylor, B., D. Trombley, and J. Reinaud. December 2012. "Energy Efficiency Resource Acquisition Program Models in North America." Prepared for ACEEE.

Tomlinson, H. March 2016. "Creating a More Energy Efficient Future for Residential Customers." The ENERGY STAR Retail Products Platform.

U.S. Energy Information Administration. Retrieved from https://www.eia.gov/.

Vine, E., R. Prahl, S. Meyers, and I. Turiel. December 2009. "An Approach for Evaluating the Market Effects of Energy Efficiency Programs."

W.K. Kellogg Foundation. January 2004. "Logic Model Development Guide."

Appendix: Tracking Total Market Change at BPA

Disclaimer: These are the opinions of BPA staff and could differ from what is recommend in the main sections of this report.

Introduction

In the Northwest, Bonneville Power Administration (BPA) performs comprehensive end-use analysis to assess the change in energy consumption due to energy efficiency adoption. This appendix provides an overview of BPA's market analysis methods and ends with a discussion of lessons learned that could inform savings estimation for the Retail Products Portfolio.

Overview of BPA Market-Analysis Methods

The BPA quantifies energy-efficiency savings due to the change in electric consumption across entire markets and end-uses. This analysis uses the baseline scenario of the regional planning forecast developed by the Northwest Power and Conservation Council and compares this baseline to the change in efficiency mix across product categories over time. BPA's market analysis seeks to quantify savings from increasing adoption of energy-efficient technologies, regardless of attribution or whether the products were incentivized by program effort.

To do this, BPA analyzes both the efficient and inefficient products entering the marketplace. The data to support this analysis comes from multiple sources which characterize the building stock (the installed products consuming energy) and the product flow (the new products every year, which create change in the building stock). One critical data source is regular onsite stock assessments, which provide the physical characteristics of buildings and the technologies installed in homes. This is combined with information on the new equipment being sold annually (the product flow), generally via sales data. The combination of the stock- and product-flow data provides a bottom-up look at energy consumption and how that energy consumption changes over time.

This "whole market" approach provides a comprehensive estimate of energy savings due to changes in efficiency. What BPA has discovered is that tracking savings for only those units that are delivered via programs ends up missing a good portion of the energy efficiency impacting the grid. Based on research performed to date,[24] BPA has found that most units sold into the market are not rebated by utility programs. For instance, in the retail lighting market, about 50% of the efficient units installed into homes went through utility programs, despite this being one of the most comprehensive programs with no barrier to entry for end-users. The impact of poor information on efficiency in programs is that econometric load forecasts have been overpredicting regional load, missing the uptick in energy efficiency adoption, and planners have suboptimum information on the remaining energy-efficiency potential. It should be noted that as states use energy efficiency to meet greenhouse gas reductions,

---

[24] BPA's market research and analysis can be found at: www.bpa.gov/goto/MomentumSavings.

quantifying all carbon reduction will increase the accuracy of energy-efficiency's contributions to meeting these goals.

BPA's analysis avoids assessing attribution to program investments. One key finding from this work is that markets are interrelated and complex, and interventions from multiple actors create ripple effects that impact demand for energy-efficient technologies. Assigning attribution neither quantifies the benefit to the electric system of energy efficiency (and in fact detracts resources from quantifying the full benefit in favor of attribution research) nor improves program effectiveness.

What Market-Analysis Methods Suggest for Tracking Savings for the Retail Products Portfolio Program

There are a few high-level lessons from BPA's market-analysis methods that can help inform the quantification of energy savings from RPP.

1. **Include efficient and inefficient products**: Understanding the change in product mix over time, across efficient and inefficient products, provides the basis for a robust analysis of energy savings. Consumer electronic product sales tend to evolve in non-linear ways (such as desktop to laptop to tablet); because of this, BPA developed "competition groups" that enable products that are different yet meet the same need to be analyzed together. Developing these competition groups can be very complex (What exactly does compete with a sound bar?); the benefit to performing this analysis is that shifts in consumer behavior can be anticipated sooner—before they decimate a program's expected savings.

2. **Track saturation of products over time**: Consumer electronics products are well-known for their shifting saturation levels, as consumer preferences evolve. Regular assessment of the saturation of products is a foundational piece to accurately estimating the product category's energy consumption.

3. **Map the supply chain and estimate market shares by channel**: The actions retailers take in the program impact the choices their competitors make and pricing from manufacturers over the long term. BPA's market analysis includes market actor research to determine market position and market size to best compile disparate data sources; it is likely that the efficiency that occurs from interventions will occur market-wide. In other words, in an interconnected marketplace, finding a method to track product sales outside of the target retailers likely will result in showing additional energy savings.

4. **Focus effort on the big levers at the market level**: Market analysis isn't more expensive than other types of evaluation and planning research. Research costs could grow unwieldy however, if resources are allocated to data points with little impact on savings. For instance, saturation and efficiency mix knowledge across a market will have a significant impact on the savings; understanding leakage will not, as the effect of leakage is marginal because the markets are interconnected.