Itron
Knowledge to Shape Your Future

Electric | Gas | Water
information collection, analysis and application

# Verification of Reported Program Impacts from 2013 EmPOWER Maryland Energy Efficiency Programs with Recommendations to Improve Future Evaluation Research

Prepared for:

EmPOWER Maryland Utilities
and the Maryland Public Service Commission

Prepared by:

Itron, Inc.
11105 Inwood Avenue
Silver Spring, MD 20910
410-353-5491

August 1, 2014

# Table of Contents

## List of Tables

## List of Figures

# 1

# Overview

## 1.1 Introduction

As the EmPOWER Maryland Independent Evaluator, Itron, Inc. (Itron) supports the Public Service Commission (PSC)'s oversight of the statewide evaluation of EmPOWER energy efficiency programs being conducted by Navigant Consulting, Inc. (Navigant) and The Cadmus Group (Cadmus), collectively referred to as the "Evaluation Team." The objectives of the statewide evaluation are to assess compliance with legislatively mandated EmPOWER Maryland energy efficiency and demand response goals, inform energy efficiency program and portfolio cost-effectiveness, and enhance program design and implementation.

This report summarizes findings and recommendations from Itron's review and verification of the 2013 EmPOWER Maryland energy efficiency programs evaluation. In addition to providing verified savings estimates for the 2013 EmPOWER programs, we flag issues identified in our review of the 2013 Evaluation Reports[1] and recommend actions to improve evaluation efforts in future years. Verification of the Evaluation Team's estimates is intended as a companion report to the 2013 Evaluation Reports and to give increased confidence to stakeholders in Maryland that the evaluated savings from the EmPOWER programs are real and credible.

The 2013 Evaluation Reports are a collection of reports, which include a Calendar Year (CY) 2013 overview report, along with individual reports for each program. The program-level reports provide detailed discussion and findings pertaining to Evaluation Year 4 (EY4) program activity occurring from June 1, 2012 through May 31, 2013. The findings from the EY activity were used to adjust savings reported in the utility tracking systems from activity occurring January 1 through May 31, 2013. Activity occurring between June 1 and December 31, 2013 was not formally evaluated, but values were adjusted to reflect findings from the EY analysis and any errors in the tracking data. The CY2013 savings is equal to the sum of the estimated savings from those two periods.

[1] See Navigant and Cadmus, *EmPOWER Maryland Final Impact Evaluation Report Calendar Year 2013*, presented to Baltimore Gas & Electric (BGE), Potomac Electric Power Company (PEPCO), Delmarva Power & Light (DPL), Southern Maryland Electric Cooperative (SMECO), and Potomac Edison (PE), June 30, 2014. Individual program evaluation reports, including drafts and finals, were also reviewed by Itron. All references to the *2013 Evaluation Reports* are to the "final" versions of these various reports received circa June 30, 2014, unless indicated otherwise.

Itron's verification analysis confirmed all of the evaluated energy (MWh) and demand (MW) savings estimates for the program Year (PY) 2013 EmPOWER portfolio. There were significant differences between evaluated and verified savings for the Residential New Construction (RNC) programs and Residential HVAC programs, but the combined savings from these programs make up only a couple percent of the overall portfolio and the differences in savings were both positive and negative, thus partially offsetting each other.

The verification process involves extensive discussions between Itron and the Evaluation Team. Most issues were reconciled as part of the verification process, thus the evaluated and verified savings are generally closely aligned. The 2013 Evaluation Report is, on the whole, the culmination of a highly professional, thoughtful, and careful evaluation of the 2013 EmPOWER programs.

The focus of this Verification Report is on issues and uncertainties in savings methods or estimates that led to our recommendations for changes to be made in future evaluation cycles. In other words, the emphasis is on unresolved issues and changes that we would like to see made in future evaluation cycles.

Itron recommends a number of actions to further increase the accuracy and reliability of evaluations conducted in 2014 and beyond. Most of these recommendations involve minor adjustment to evaluation methods and/or assumptions and will be resolved directly with the Evaluation Team. Where more significant issues are flagged, Itron will work with the Evaluation Team, staff and stakeholders to ensure they are addressed in future evaluations.

## 1.2 Verification Process and Approach

### *1.2.1 Verification Process*

The verification process involves significant collaboration and interaction between Itron and the Evaluation Team. This process allows Itron and the Evaluation Team to independently work through many smaller technical issues and focus the PSC staff and stakeholders on higher-level data collection and evaluation policy issues.

The process Itron used for verifying evaluated savings from the 2013 programs was similar to the approach that was used for the 2009–2012 programs.

- We reviewed and approved the Evaluation Team's evaluation plans around mid-year of 2013. Throughout the year, we worked with the Evaluation Team to anticipate and resolve any issues that arose in the course of the evaluation. This helped reduce, but not eliminate, the number of issues that needed to be resolved as part of the 2013 verification analysis.

- We provided comments on various drafts of the 2013 Evaluation Report*;* most of these comments were addressed in the final version of the report.
- Based on review of early 2013 Evaluation Report drafts, we collected data from the Evaluation Team and engaged in iterative correspondence to fill any gaps in the 2013 Evaluation Report discussion or data and to ensure that our interpretation of the reports was accurate.
- We shared draft versions of the 2013 Verification Report with the Evaluation Team before submitting them to PSC staff, the EmPOWER utilities, and other stakeholders. This allowed the Evaluation Team to comment (verbally and in writing) on the substance of the 2013 Verification Report and the way in which various issues were characterized. Where Itron concurred with the Evaluation Team's comments, the 2013 Verification Report sections were revised.
- We submitted a draft of the 2013 Verification Report to PSC staff and other stakeholders and discussed key findings and recommendations in a meeting in Baltimore, Maryland on June 9, 2014.
- Comments and additional findings from these meetings, along with written comments provided outside these meetings, were incorporated into the final 2013 Verification Report.

### *1.2.2 Verification Analysis Approach*

To structure the PY2013 verification, we reviewed draft versions of the 2013 Evaluation Report and flagged issues requiring further examination for each program area. In this initial review, Itron focused on the following types of questions:

- **2012 Verification Report recommendations:** Did the Evaluation Team respond to our previous recommendations? Do we accept the proposed approach?
- **Tracking system review:** Does the 2013 Evaluation Report document the process for reviewing utility tracking data and pulling appropriate samples based on the confidence and precision targets? Are the evaluated measures clearly defined in the 2013 Evaluation Report?
- **Primary data collection methods:** Are potential sources of bias identified and mitigated? Are survey instruments asking the right questions in the right ways? Did the samples meet the 90-20 confidence precision requirements specified in the Strategic Evaluation Plan?
- **Algorithms and assumptions:** Does the analysis use the Mid-Atlantic Technical Reference Manual (TRM) recommended algorithms and assumptions? If not, are the alternative algorithms, sources of data, and assumptions adequately explained and

justified? Are baseline assumptions clear and do they accurately reflect current market conditions in Maryland?

- **Reporting:** Are findings and key assumptions clearly reported? Is the report narrative clear and concise? Is the report organized so that readers can find the methods, assumptions, and findings easily?

These questions were asked with respect to the gross and net energy and peak demand savings.

Where the answers to these questions were "no" or Itron reviewers were unclear, the issues were flagged for further review. Various lines of inquiry were pursued for each program. Each program area was assigned an Itron reviewer, who worked with the Evaluation Team to collect additional data and gain clarity on the various issues.

Most of the tables in this report compare CY2013 savings estimates from four sources, which are described below:

- **Utility Semi-Annual Reports:** Submitted by the EmPOWER utilities to the Maryland PSC in January 2014, these public reports provide high-level estimates of program savings based on utility tracking systems.
- **Utility Tracking System Data:** Provided to the Evaluation Team by the utilities and their implementation contractors, these detailed measure level data are the basis for all of the evaluation and verification activities. Tracking systems are updated each year to reflect the latest evaluation findings.
- **2013 Evaluation Reports:** The Evaluation Team relies on tracking data, supplemented by various primary data collection and the Mid-Atlantic Technical Reference Manual (TRM) to estimate program savings. The 2013 Evaluation Reports include a Calendar Year Report, which is informed by a collection of program-level Evaluation Year reports.
- **Itron Verification:** Itron reviews tracking data, evaluation calculations, primary data collection processes and findings, and the Evaluation Reports. Itron works with the Evaluation Team to ensure that the Evaluation Reports are accurate, clear, comprehensive, and are based on industry standard evaluation practices.

In a perfect world, the savings estimates from each of these sources would be equal, but, as discussed below, the differences in savings estimates from these four sources can be quite large. Numerous factors drive differences in the respective results, including: mathematical and clerical errors, methodologies, changes in market conditions, and assumptions. Large differences are symptomatic of a breakdown somewhere in the evaluation and reporting system and should be investigated. There is no expectation that the four estimates should be equal, but over time they should be more closely aligned.

Depending on how the estimates from each source are used, there could be a need to retrospectively try to align them. For example, utility tracking systems are prospectively updated each year to reflect evaluation and verification findings. Utility Semi-Annual Reports, probably the most public of the various estimates, should reflect evaluation and verification findings. Evaluation methods should take into account verification findings and recommendations from current or previous years.

The next subsection summarizes the verification findings for the various programs. The final subsection summarizes the various recommendations along with related action items.

## 1.3 Verification Findings

### *1.3.1 Overview*

Table 1-1 compares gross and net statewide portfolio energy and peak demand savings estimates from the Semi-Annual Report, Tracking Systems, Evaluation Reports, and Itron Verification at the premise level. Verified gross annual energy savings totaled 810 GWh and gross peak demand totaled 125 MW.[2] Taking free ridership and spillover into account, verified net annual energy savings totaled 562 GWh and gross peak demand totaled 86 MW. Itron's verified gross and net savings at the overall statewide level were almost identical (less than 0.2% difference) to the final evaluated savings estimates. Verified savings differed from evaluated savings for only two small programs: Residential HVAC and RNC.

**Table 1-1: Statewide Reported, Evaluated, and Verified Savings**

| Savings Type | Gross or Net | 2013 Full Year Semi-Annual Report | 2013 Full Year Utility Tracking System | 2013 Full Year Ex Post Evaluated | 2013 Full Year Ex Post Verified |
|---|---|---|---|---|---|
| Annual Energy Savings (MWh) | Gross | 877,571 | 874,608 | 810,383 | 809,975 |
| | Net | 608,970 | 610,883 | 562,058 | 561,732 |
| Utility Coincident Peak Demand Savings (kW) | Gross | 159,467 | 124,912 | 125,547 | 125,409 |
| | Net | 109,249 | 85,348 | 85,520 | 85,698 |

Table 1-2 compares the verified savings estimates to Semi-Annual Reports and tracking system estimates. As shown, verified peak demand savings were only 79% of savings reported in the utilities' Semi-Annual Reports, while verified and tracking system savings were almost equal. As shown below, PEPCO was responsible for most of the discrepancy between statewide peak demand savings reported in the Semi-Annual Report and tracking systems. This suggests there

[2] Gross savings includes savings from free riders and, to a lesser extent, spillover. Unless specifically stated, peak demand savings in this report are based on utility peak periods; PJM demand estimates are provided for programs where savings are bid into the PJM forward market.

was some disconnect between the tracking system and the Semi-Annual Report peak savings estimates, at least at the time the Semi-Annual Reports were compiled. For energy savings, the verified gross and net savings were 92% to 93% of both tracking system and Semi-Annual report savings.

**Table 1-2: Statewide Verified versus Semi-Annual and Tracking System Savings**

| | Gross | | Net | |
|---|---|---|---|---|
| **Savings Type** | **Verified % of Semi Annual** | **Verified % of Tracking** | **Verified % of Semi Annual** | **Verified % of Tracking** |
| Annual Energy Savings (MWh) | 92% | 93% | 92% | 92% |
| Utility Coincident Peak Demand Savings (kW) | 79% | 100% | 78% | 100% |

Table 1-3 and Table 1-4 compare gross and net MWh savings, respectively, at the utility level.

**Table 1-3: Reported, Evaluated, and Verified Gross Annual Energy Savings by Utility (MWh)**

| | Gross Savings | | | |
|---|---|---|---|---|
| **Utility** | **2013 Full Year Semi-Annual Report** | **2013 Full Year Utility Tracking System** | **2013 Full Year Ex Post Evaluated** | **2013 Full Year Ex Post Verified** |
| BGE | 421,942 | 430,134 | 379,302 | 379,072 |
| PEPCO | 277,001 | 268,669 | 258,843 | 258,820 |
| DPL | 56,723 | 57,913 | 57,103 | 57,092 |
| SMECO | 41,782 | 40,771 | 38,985 | 38,933 |
| PE | 80,123 | 77,120 | 76,151 | 76,057 |
| **Total** | **877,571** | **874,607** | **810,384** | **809,975** |

**Table 1-4: Reported, Evaluated, and Verified Net Annual Energy Savings by Utility (MWh)**

| | Net Savings | | | |
|---|---|---|---|---|
| **Utility** | **2013 Full Year Semi-Annual Report** | **2013 Full Year Utility Tracking System** | **2013 Full Year Ex Post Evaluated** | **2013 Full Year Ex Post Verified** |
| BGE | 294,690 | 301,916 | 265,017 | 264,824 |
| PEPCO | 190,435 | 186,443 | 177,876 | 177,858 |
| DPL | 40,730 | 41,777 | 41,024 | 41,015 |
| SMECO | 27,738 | 27,252 | 25,813 | 25,770 |
| PE | 55,377 | 53,496 | 52,329 | 52,265 |
| **Total** | **608,970** | **610,883** | **562,059** | **561,732** |

Table 1-5 compares the verified annual energy savings estimates to Semi-Annual reports and tracking system estimates. Semi-Annual Report and tracking system estimates were generally closely aligned, with only a few percentage point separating the two. Verified savings were generally more than 90% of Semi-Annual and tracking system savings. The exception was BGE, for which verified savings as a percent of Semi-Annual and tracking system savings were significantly lower than the other utilities. This was due to issues with the tracking data for the BGE C&I programs (discussed below), which comprised approximately one-fourth of the utility-reported statewide portfolio energy and demand savings.

### Table 1-5: Verified versus Semi-Annual and Tracking System Annual Energy Savings by Utility

| Utility | Gross | | Net | |
|---|---|---|---|---|
| | Verified % of Semi Annual | Verified % of Tracking | Verified % of Semi Annual | Verified % of Tracking |
| BGE | 90% | 88% | 90% | 88% |
| PEPCO | 93% | 96% | 93% | 95% |
| DPL | 101% | 99% | 101% | 98% |
| SMECO | 93% | 95% | 93% | 95% |
| PE | 95% | 99% | 94% | 98% |
| **Total** | 92% | 93% | 92% | 92% |

Table 1-6 and Table 1-7 compare gross and net kW savings, respectively, at the utility level.

### Table 1-6: Reported, Evaluated, and Verified Gross Utility Peak Demand Savings by Utility (kW)

| Utility | Gross Savings | | | |
|---|---|---|---|---|
| | 2013 Full Year Semi-Annual Report | 2013 Full Year Utility Tracking System | 2013 Full Year Ex Post Evaluated | 2013 Full Year Ex Post Verified |
| BGE | 67,923 | 61,201 | 56,312 | 56,275 |
| PEPCO | 65,900 | 39,075 | 43,509 | 43,408 |
| DPL | 8,378 | 8,096 | 9,370 | 9,398 |
| SMECO | 5,932 | 5,729 | 5,912 | 5,921 |
| PE | 11,334 | 10,811 | 10,445 | 10,407 |
| **Total** | **159,467** | **124,912** | **125,548** | **125,409** |

### Table 1-7: Reported, Evaluated, and Verified Net Utility Peak Demand Savings by Utility (kW)

| Utility | Net Savings | | | |
|---|---|---|---|---|
| | 2013 Full Year Semi-Annual Report | 2013 Full Year Utility Tracking System | 2013 Full Year Ex Post Evaluated | 2013 Full Year Ex Post Verified |
| BGE | 46,214 | 41,709 | 38,410 | 38,549 |
| PEPCO | 45,726 | 26,879 | 29,738 | 29,706 |
| DPL | 5,891 | 5,726 | 6,614 | 6,641 |
| SMECO | 3,886 | 3,792 | 3,844 | 3,872 |
| PE | 7,533 | 7,242 | 6,914 | 6,931 |
| **Total** | **109,249** | **85,348** | **85,520** | **85,698** |

Table 1-8 compares the verified peak demand savings estimates to Semi-Annual Reports and tracking system estimates. As shown, PEPCO's verified demand savings were just 66% of Semi-Annual Report savings, while its verified savings as a percentage of tracking system savings was 111%. This suggests a major disconnect between tracking system savings and Semi-Annual Report savings. BGE's verified demand savings were also relatively low compared to Semi-Annual and tracking system savings. The other three utilities were markedly better in comparison. DPL-verified savings exceeded Semi-Annual and tracking system savings across the board.

### Table 1-8: Verified versus Semi-Annual and Tracking System Peak Demand Savings by Utility

| Utility | Gross | | Net | |
|---|---|---|---|---|
| | Verified % of Semi Annual | Verified % of Tracking | Verified % of Semi Annual | Verified % of Tracking |
| BGE | 83% | 92% | 83% | 92% |
| PEPCO | 66% | 111% | 65% | 111% |
| DPL | 112% | 116% | 113% | 116% |
| SMECO | 100% | 103% | 100% | 102% |
| PE | 92% | 96% | 92% | 96% |
| **Total** | 79% | 100% | 78% | 100% |

### *1.3.2 Program-Level Results*

At the program level, verified savings were generally within 10 to 20% of utility-reported savings, with a few exceptions:

- Most notably, BGE's Custom program data were particularly problematic, with verified savings equaling only 61% of reported energy savings and 33% of demand savings.
- For the utilities' Appliance Recycling programs, verified demand savings always exceeded tracking system demand savings. BGE and DPL verified savings were roughly double the tracking system savings; PEPCO and SMECO verified savings were approximately 1.5 times the tracking system savings.
- PEPCO's Master Metered Multi-family program verified energy and demand savings were only 74% and 68%, respectively, of tracking system savings.
- The RNC programs evaluated savings as a percent of tracking system savings were uniformly 112% of energy and 75% of demand for all five utilities.

Verified savings were equal to evaluated savings for all but two programs: the Residential HVAC programs and the RNC programs. As shown in Table 1-9, for the Residential HVAC programs, verified demand savings at the statewide level were 89% of the evaluated savings. Verified energy savings equaled evaluated savings.

**Table 1-9: Residential HVAC Verified Gross Savings as Percent of Gross Evaluated Savings**

| | BGE | PEPCO | DPL | SMECO | PE | Statewide |
|---|---|---|---|---|---|---|
| Utility Demand (kW) | 88% | 91% | 97% | 91% | 87% | 89% |
| Energy (kWh) | 100% | 100% | 100% | 100% | 98% | 100% |

The differences between evaluated and verified savings are due mostly to the Evaluation Team's use of averages for equipment performance ratings. Itron calculated savings using individual unit data, which were recorded in the tracking systems. Some additional difference resulted from Itron rejection of the Evaluation Team's zeroing impacts because of missing data. Section 6 discusses these issues in detail.

For the RNC programs, verified demand savings at the statewide level were about one fourth higher than evaluated savings and verified energy savings were somewhat lower, as shown in Table 1-10.

### Table 1-10: Verified Gross Savings as Percent of Gross Evaluated Savings—Residential New Construction Programs

| | BGE | PEPCO | DPL | SMECO | PE | Statewide |
|---|---|---|---|---|---|---|
| Utility Demand (kW) | 129% | 107% | 145% | 114% | 118% | 123% |
| Energy (kWh) | 93% | 97% | 96% | 94% | 94% | 94% |

Verified results for RNC kWh savings differ from the evaluated results due to discrepancies found in the calculation of savings. The evaluated kW demand reduction differences are due to the manner in which the evaluated sample results were scaled to the participant population. These issues are discussed in detail in Section 8.

The tables below summarize program level results for each utility. As with the previous subsection, we compare gross and net energy and peak demand savings estimates from the Semi-Annual Report, Tracking Systems, Evaluation Reports and the Itron Verification.

### Table 1-11: BGE Reported, Evaluated, and Verified Annual Energy Savings (MWh) by Program—Gross

| | Gross Savings | | | |
|---|---|---|---|---|
| **Program** | **2013 Full Year Semi-Annual Report** | **2013 Full Year Utility Tracking System** | **2013 Full Year Ex Post Evaluated** | **2013 Full Year Ex Post Verified** |
| C&I Prescriptive | 96,779 | 102,112 | 81,095 | 81,095 |
| C&I Small Business | 35,770 | 37,419 | 30,689 | 30,689 |
| C&I Custom | 55,956 | 65,583 | 40,216 | 40,216 |
| Res. Lighting | 181,448 | 170,645 | 180,602 | 180,602 |
| Res. Appliances | 11,487 | 9,450 | 8,464 | 8,464 |
| Res. Appliance Recycling | 6,099 | 6,602 | 7,592 | 7,592 |
| Res. HVAC | 7,229 | 7,731 | 6,688 | 6,687 |
| Res. Retrofit – QHEC | 21,477 | 24,966 | 18,768 | 18,768 |
| Res. Retrofit - HPWES | 2,591 | 2,522 | 1,712 | 1,712 |
| Res NC | 3,105 | 3,104 | 3,477 | 3,248 |
| **Total** | **421,942** | **430,134** | **379,303** | **379,072** |

### Table 1-12: BGE Reported, Evaluated, and Verified Annual Energy Savings (MWh) by Program—Net

| Program | Net Savings | | | |
|---|---|---|---|---|
| | 2013 Full Year Semi-Annual Report | 2013 Full Year Utility Tracking System | 2013 Full Year Ex Post Evaluated | 2013 Full Year Ex Post Verified |
| C&I Prescriptive | 69,681 | 73,521 | 58,388 | 58,388 |
| C&I Small Business | 26,470 | 27,690 | 22,710 | 22,710 |
| C&I Custom | 38,609 | 45,252 | 27,749 | 27,749 |
| Res. Lighting | 125,199 | 117,745 | 124,615 | 124,615 |
| Res. Appliances | 4,020 | 3,308 | 2,962 | 2,962 |
| Res. Appliance Recycling | 3,599 | 3,895 | 4,479 | 4,479 |
| Res. HVAC | 2,819 | 3,015 | 2,609 | 2,608 |
| Res. Retrofit – QHEC | 19,974 | 23,218 | 17,454 | 17,454 |
| Res. Retrofit – HPWES | 1,710 | 1,665 | 1,130 | 1,130 |
| Res NC | 2,608 | 2,607 | 2,920 | 2,728 |
| **Total** | **294,690** | **301,916** | **265,016** | **264,824** |

### Table 1-13: BGE Reported, Evaluated, and Verified Utility Coincident Peak Demand Savings (kW) by Program Category—Gross

| Program | Gross Savings | | | |
|---|---|---|---|---|
| | 2013 Full Year Semi-Annual Report | 2013 Full Year Utility Tracking System | 2013 Full Year Ex Post Evaluated | 2013 Full Year Ex Post Verified |
| C&I Prescriptive | 13,840 | 12,995 | 14,282 | 14,281 |
| C&I Small Business | 11,537 | 8,820 | 5,994 | 5,994 |
| C&I Custom | 9,827 | 8,916 | 2,899 | 2,899 |
| Res. Lighting | 22,193 | 19,813 | 23,465 | 23,465 |
| Res. Appliances | 1,712 | 1,421 | 1,200 | 1,200 |
| Res. Appliance Recycling | 925 | 978 | 1,872 | 1,872 |
| Res. HVAC | 3,261 | 3,192 | 3,179 | 2,803 |
| Res. Retrofit – QHEC | 2,396 | 2,866 | 1,853 | 1,853 |
| Res. Retrofit – HPWES | 646 | 646 | 402 | 402 |
| Res NC | 1,587 | 1,554 | 1,165 | 1,506 |
| **Total** | **67,923** | **61,201** | **56,311** | **56,275** |

### Table 1-14: BGE Reported, Evaluated, and Verified Utility Coincident Peak Demand Savings (kW) by Program Category—Net

| Program | Net Savings | | | |
|---|---|---|---|---|
| | 2013 Full Year Semi-Annual Report | 2013 Full Year Utility Tracking System | 2013 Full Year Ex Post Evaluated | 2013 Full Year Ex Post Verified |
| C&I Prescriptive | 9,965 | 9,356 | 10,283 | 10,282 |
| C&I Small Business | 8,537 | 6,527 | 4,436 | 4,436 |
| C&I Custom | 5,994 | 5,439 | 1,768 | 1,768 |
| Res. Lighting | 15,313 | 13,671 | 16,191 | 16,191 |
| Res. Appliances | 599 | 497 | 420 | 420 |
| Res. Appliance Recycling | 546 | 577 | 1,104 | 1,104 |
| Res. HVAC | 1,272 | 1,245 | 1,240 | 1,093 |
| Res. Retrofit – QHEC | 2,228 | 2,665 | 1,724 | 1,723 |
| Res. Retrofit - HPWES | 426 | 426 | 265 | 265 |
| Res NC | 1,333 | 1,305 | 979 | 1,265 |
| **Total** | **46,214** | **41,709** | **38,410** | **38,549** |

### Table 1-15: PEPCO Reported, Evaluated, and Verified Annual Energy Savings (MWh) by Program Category—Gross

| Program | Gross Savings | | | |
|---|---|---|---|---|
| | 2013 Full Year Semi-Annual Report | 2013 Full Year Utility Tracking System | 2013 Full Year Ex Post Evaluated | 2013 Full Year Ex Post Verified |
| C&I Prescriptive | 38,958 | 41,984 | 38,272 | 38,272 |
| C&I Small Business | 50,526 | 52,973 | 49,823 | 49,823 |
| C&I Custom | 28,521 | 19,430 | 21,498 | 21,498 |
| MMMF | 4,110 | 4,617 | 3,423 | 3,423 |
| Res. Lighting | 120,013 | 112,038 | 114,112 | 114,112 |
| Res. Appliances | 3,745 | 3,185 | 3,052 | 3,052 |
| Res. Appliance Recycling | 2,922 | 2,690 | 3,116 | 3,116 |
| Res. HVAC | 1,981 | 2,164 | 1,914 | 1,912 |
| Res. Retrofit - QHEC | 23,640 | 26,906 | 21,385 | 21,385 |
| Res. Retrofit - HPWES | 2,061 | 2,159 | 1,660 | 1,660 |
| Res NC | 523 | 523 | 586 | 566 |
| **Total** | **277,001** | **268,669** | **258,841** | **258,823** |

### Table 1-16: PEPCO Reported, Evaluated, and Verified Annual Energy Savings (MWh) by Program Category—Net

| Program | Net Savings | | | |
|---|---|---|---|---|
| | 2013 Full Year Semi-Annual Report | 2013 Full Year Utility Tracking System | 2013 Full Year Ex Post Evaluated | 2013 Full Year Ex Post Verified |
| C&I Prescriptive | 28,050 | 30,228 | 27,556 | 27,556 |
| C&I Small Business | 37,389 | 39,200 | 36,869 | 36,869 |
| C&I Custom | 19,679 | 13,407 | 14,834 | 14,834 |
| MMMF | 3,740 | 4,201 | 3,115 | 3,115 |
| Res. Lighting | 74,408 | 69,464 | 70,750 | 70,750 |
| Res. Appliances | 1,311 | 1,115 | 1,068 | 1,068 |
| Res. Appliance Recycling | 1,753 | 1,614 | 1,870 | 1,870 |
| Res. HVAC | 793 | 866 | 766 | 765 |
| Res. Retrofit - QHEC | 21,513 | 24,484 | 19,461 | 19,461 |
| Res. Retrofit - HPWES | 1,360 | 1,425 | 1,095 | 1,095 |
| Res NC | 440 | 439 | 492 | 476 |
| **Total** | **190,435** | **186,443** | **177,876** | **177,858** |

### Table 1-17: PEPCO Reported, Evaluated, and Verified Coincident Peak Demand Savings (kW) by Program Category—Gross

| Program | Gross Savings | | | |
|---|---|---|---|---|
| | 2013 Full Year Semi-Annual Report | 2013 Full Year Utility Tracking System | 2013 Full Year Ex Post Evaluated | 2013 Full Year Ex Post Verified |
| C&I Prescriptive | 29,282 | 5,135 | 5,789 | 5,789 |
| C&I Small Business | 11,745 | 11,966 | 14,931 | 14,931 |
| C&I Custom | 3,645 | 2,232 | 2,261 | 2,261 |
| MMMF | 512 | 570 | 389 | 390 |
| Res. Lighting | 14,683 | 12,982 | 15,080 | 15,081 |
| Res. Appliances | 540 | 456 | 410 | 410 |
| Res. Appliance Recycling | 483 | 402 | 586 | 586 |
| Res. HVAC | 1,394 | 1,342 | 1,254 | 1,135 |
| Res. Retrofit - QHEC | 2,720 | 3,088 | 2,188 | 2,188 |
| Res. Retrofit - HPWES | 584 | 600 | 394 | 393 |
| Res NC | 311 | 303 | 228 | 244 |
| **Total** | **65,900** | **39,076** | **43,510** | **43,408** |

### Table 1-18: PEPCO Reported, Evaluated, and Verified Coincident Peak Demand Savings (kW) by Program Category—Net

| Program | Net Savings | | | |
|---|---|---|---|---|
| | 2013 Full Year Semi-Annual Report | 2013 Full Year Utility Tracking System | 2013 Full Year Ex Post Evaluated | 2013 Full Year Ex Post Verified |
| C&I Prescriptive | 21,083 | 3,697 | 4,168 | 4,168 |
| C&I Small Business | 8,692 | 8,855 | 11,049 | 11,049 |
| C&I Custom | 2,224 | 1,362 | 1,379 | 1,379 |
| MMMF | 466 | 519 | 354 | 355 |
| Res. Lighting | 9,103 | 8,049 | 9,350 | 9,350 |
| Res. Appliances | 189 | 160 | 143 | 144 |
| Res. Appliance Recycling | 290 | 241 | 352 | 352 |
| Res. HVAC | 558 | 537 | 502 | 454 |
| Res. Retrofit - QHEC | 2,475 | 2,810 | 1,991 | 1,991 |
| Res. Retrofit - HPWES | 386 | 396 | 260 | 259 |
| Res NC | 261 | 255 | 192 | 205 |
| **Total** | **45,726** | **26,879** | **29,740** | **29,706** |

### Table 1-19: DPL Reported, Evaluated, and Verified Annual Energy Savings (MWh) by Program Category—Gross

| Program | Gross Savings | | | |
|---|---|---|---|---|
| | 2013 Full Year Semi-Annual Report | 2013 Full Year Utility Tracking System | 2013 Full Year Ex Post Evaluated | 2013 Full Year Ex Post Verified |
| C&I Prescriptive | 8,201 | 9,539 | 8,804 | 8,804 |
| C&I Small Business | 13,891 | 14,852 | 13,979 | 13,979 |
| C&I Custom | 3,148 | 3,322 | 3,675 | 3,675 |
| Res. Lighting | 24,959 | 23,155 | 24,537 | 24,537 |
| Res. Appliances | 865 | 742 | 709 | 709 |
| Res. Appliance Recycling | 551 | 519 | 611 | 611 |
| Res. HVAC | 586 | 657 | 551 | 551 |
| Res. Retrofit - QHEC | 4,022 | 4,613 | 3,787 | 3,787 |
| Res. Retrofit - HPWES | 251 | 266 | 168 | 168 |
| Res NC | 249 | 250 | 280 | 270 |
| **Total** | **56,723** | **57,915** | **57,101** | **57,092** |

### Table 1-20: DPL Reported, Evaluated, and Verified Annual Energy Savings (MWh) by Program Category—Net

| Program | Net Savings | | | |
|---|---|---|---|---|
| | 2013 Full Year Semi-Annual Report | 2013 Full Year Utility Tracking System | 2013 Full Year Ex Post Evaluated | 2013 Full Year Ex Post Verified |
| C&I Prescriptive | 5,905 | 6,868 | 6,339 | 6,339 |
| C&I Small Business | 10,280 | 10,990 | 10,345 | 10,345 |
| C&I Custom | 2,172 | 2,292 | 2,536 | 2,536 |
| Res. Lighting | 17,471 | 16,209 | 17,176 | 17,176 |
| Res. Appliances | 303 | 260 | 248 | 248 |
| Res. Appliance Recycling | 330 | 311 | 366 | 366 |
| Res. HVAC | 234 | 263 | 221 | 221 |
| Res. Retrofit – QHEC | 3,660 | 4,198 | 3,447 | 3,447 |
| Res. Retrofit – HPWES | 166 | 176 | 111 | 111 |
| Res NC | 209 | 210 | 235 | 227 |
| **Total** | **40,730** | **41,777** | **41,024** | **41,015** |

### Table 1-21: DPL Reported, Evaluated, and Verified Coincident Peak Demand Savings (kW) by Program Category—Gross

| Program | Gross Savings | | | |
|---|---|---|---|---|
| | 2013 Full Year Semi-Annual Report | 2013 Full Year Utility Tracking System | 2013 Full Year Ex Post Evaluated | 2013 Full Year Ex Post Verified |
| C&I Prescriptive | 1,328 | 1,290 | 1,457 | 1,457 |
| C&I Small Business | 2,388 | 2,375 | 2,954 | 2,953 |
| C&I Custom | 654 | 629 | 637 | 637 |
| Res. Lighting | 2,930 | 2,683 | 3,344 | 3,344 |
| Res. Appliances | 129 | 114 | 99 | 99 |
| Res. Appliance Recycling | 81 | 77 | 164 | 164 |
| Res. HVAC | 286 | 261 | 231 | 225 |
| Res. Retrofit - QHEC | 425 | 506 | 362 | 362 |
| Res. Retrofit - HPWES | 52 | 56 | 42 | 41 |
| Res NC | 106 | 107 | 80 | 116 |
| **Total** | **8,378** | **8,098** | **9,370** | **9,398** |

### Table 1-22: DPL Reported, Evaluated, and Verified Coincident Peak Demand Savings (kW) by Program Category—Net

| Program | Net Savings | | | |
|---|---|---|---|---|
| | 2013 Full Year Semi-Annual Report | 2013 Full Year Utility Tracking System | 2013 Full Year Ex Post Evaluated | 2013 Full Year Ex Post Verified |
| C&I Prescriptive | 956 | 929 | 1,049 | 1,049 |
| C&I Small Business | 1,767 | 1,758 | 2,186 | 2,185 |
| C&I Custom | 399 | 384 | 389 | 389 |
| Res. Lighting | 2,051 | 1,878 | 2,341 | 2,341 |
| Res. Appliances | 45 | 40 | 35 | 35 |
| Res. Appliance Recycling | 49 | 46 | 98 | 98 |
| Res. HVAC | 114 | 104 | 92 | 90 |
| Res. Retrofit - QHEC | 387 | 460 | 329 | 329 |
| Res. Retrofit - HPWES | 35 | 37 | 27 | 27 |
| Res NC | 89 | 90 | 67 | 98 |
| **Total** | **5,891** | **5,726** | **6,613** | **6,641** |

### Table 1-23: SMECO Reported, Evaluated, and Verified Annual Energy Savings (MWh) by Program Category—Gross

| Program | Gross Savings | | | |
|---|---|---|---|---|
| | 2013 Full Year Semi-Annual Report | 2013 Full Year Utility Tracking System | 2013 Full Year Ex Post Evaluated | 2013 Full Year Ex Post Verified |
| C&I Prescriptive | 4,316 | 4,356 | 3,506 | 3,506 |
| C&I Small Business | 1,529 | 1,550 | 1,187 | 1,187 |
| C&I Custom | 712 | 964 | 800 | 800 |
| Res. Lighting | 25,009 | 23,446 | 23,934 | 23,934 |
| Res. Appliances | 1,746 | 1,399 | 1,430 | 1,430 |
| Res. Appliance Recycling | 1,535 | 1,656 | 1,705 | 1,705 |
| Res. HVAC | 1,176 | 1,337 | 1,135 | 1,135 |
| Res. Retrofit – QHEC | 4,741 | 5,023 | 4,213 | 4,213 |
| Res. Retrofit – HPWES | 222 | 244 | 183 | 184 |
| Res NC | 795 | 795 | 891 | 839 |
| **Total** | **41,782** | **40,770** | **38,984** | **38,933** |

### Table 1-24: SMECO Reported, Evaluated, and Verified Annual Energy Savings (MWh) by Program Category—Net

| Program | Net Savings | | | |
|---|---|---|---|---|
| | 2013 Full Year Semi-Annual Report | 2013 Full Year Utility Tracking System | 2013 Full Year Ex Post Evaluated | 2013 Full Year Ex Post Verified |
| C&I Prescriptive | 3,108 | 3,136 | 2,524 | 2,524 |
| C&I Small Business | 1,132 | 1,147 | 879 | 879 |
| C&I Custom | 492 | 665 | 552 | 552 |
| Res. Lighting | 15,756 | 14,771 | 15,078 | 15,078 |
| Res. Appliances | 646 | 518 | 529 | 529 |
| Res. Appliance Recycling | 906 | 977 | 1,006 | 1,006 |
| Res. HVAC | 435 | 495 | 420 | 420 |
| Res. Retrofit - QHEC | 4,457 | 4,722 | 3,961 | 3,961 |
| Res. Retrofit - HPWES | 140 | 154 | 116 | 116 |
| Res NC | 668 | 668 | 748 | 705 |
| **Total** | **27,738** | **27,252** | **25,813** | **25,770** |

### Table 1-25: SMECO Reported, Evaluated, and Verified Coincident Peak Demand Savings (kW) by Program Category—Gross

| Program | Gross Savings | | | |
|---|---|---|---|---|
| | 2013 Full Year Semi-Annual Report | 2013 Full Year Utility Tracking System | 2013 Full Year Ex Post Evaluated | 2013 Full Year Ex Post Verified |
| C&I Prescriptive | 564 | 614 | 742 | 742 |
| C&I Small Business | 343 | 332 | 240 | 240 |
| C&I Custom | (17) | (17) | (22) | (22) |
| Res. Lighting | 2,949 | 2,711 | 3,100 | 3,100 |
| Res. Appliances | 243 | 202 | 187 | 187 |
| Res. Appliance Recycling | 240 | 247 | 359 | 358 |
| Res. HVAC | 532 | 507 | 467 | 423 |
| Res. Retrofit - QHEC | 520 | 564 | 412 | 413 |
| Res. Retrofit - HPWES | 38 | 46 | 34 | 34 |
| Res NC | 522 | 523 | 392 | 446 |
| **Total** | **5,932** | **5,729** | **5,911** | **5,921** |

### Table 1-26: SMECO Reported, Evaluated, and Verified Coincident Peak Demand Savings (kW) by Program Category—Net

| Program | Net Savings | | | |
|---|---|---|---|---|
| | 2013 Full Year Semi-Annual Report | 2013 Full Year Utility Tracking System | 2013 Full Year Ex Post Evaluated | 2013 Full Year Ex Post Verified |
| C&I Prescriptive | 406 | 442 | 534 | 534 |
| C&I Small Business | 254 | 246 | 177 | 178 |
| C&I Custom | (10) | (10) | (13) | (13) |
| Res. Lighting | 1,858 | 1,708 | 1,953 | 1,953 |
| Res. Appliances | 90 | 75 | 69 | 69 |
| Res. Appliance Recycling | 142 | 146 | 212 | 211 |
| Res. HVAC | 197 | 188 | 173 | 157 |
| Res. Retrofit – QHEC | 488 | 530 | 388 | 388 |
| Res. Retrofit – HPWES | 24 | 29 | 21 | 21 |
| Res NC | 438 | 439 | 329 | 374 |
| **Total** | **3,886** | **3,792** | **3,843** | **3,872** |

### Table 1-27: PE Reported, Evaluated, and Verified Annual Energy Savings (MWh) by Program Category—Gross

| Program | Gross Savings | | | |
|---|---|---|---|---|
| | 2013 Full Year Semi-Annual Report | 2013 Full Year Utility Tracking System | 2013 Full Year Ex Post Evaluated | 2013 Full Year Ex Post Verified |
| C&I Prescriptive | 18,498 | 18,498 | 16,708 | 16,708 |
| C&I Small Business | 2,203 | 2,202 | 2,077 | 2,077 |
| C&I Custom | 8,978 | 8,993 | 8,925 | 8,925 |
| Res. Lighting | 31,829 | 28,569 | 32,018 | 32,018 |
| Res. Appliances | 2,329 | 2,186 | 1,798 | 1,798 |
| Res. Appliance Recycling | 3,053 | 2,716 | 3,096 | 3,096 |
| Res. HVAC | 1,352 | 1,387 | 1,752 | 1,722 |
| Res. Retrofit - QHEC | 10,145 | 10,762 | 8,019 | 8,019 |
| Res. Retrofit - HPWES | 822 | 893 | 734 | 734 |
| Res NC | 914 | 914 | 1,024 | 960 |
| **Total** | **80,698** | **77,120** | **76,151** | **76,057** |

### Table 1-28: PE Reported, Evaluated, and Verified Annual Energy Savings (MWh) by Program Category—Net

| Program | Net Savings | | | |
|---|---|---|---|---|
| | 2013 Full Year Semi-Annual Report | 2013 Full Year Utility Tracking System | 2013 Full Year Ex Post Evaluated | 2013 Full Year Ex Post Verified |
| C&I Prescriptive | 13,319 | 13,319 | 12,030 | 12,030 |
| C&I Small Business | 1,079 | 1,079 | 1,018 | 1,018 |
| C&I Custom | 6,195 | 6,205 | 6,158 | 6,158 |
| Res. Lighting | 21,325 | 19,141 | 21,452 | 21,452 |
| Res. Appliances | 815 | 765 | 629 | 629 |
| Res. Appliance Recycling | 2,198 | 1,956 | 2,229 | 2,229 |
| Res. HVAC | 487 | 499 | 631 | 620 |
| Res. Retrofit - QHEC | 8,623 | 9,148 | 6,817 | 6,817 |
| Res. Retrofit - HPWES | 567 | 616 | 507 | 507 |
| Res NC | 768 | 768 | 860 | 806 |
| **Total** | **55,377** | **53,496** | **52,331** | **52,265** |

### Table 1-29: PE Reported, Evaluated, and Verified Coincident Peak Demand Savings (kW) by Program Category—Gross

| Program | Gross Savings | | | |
|---|---|---|---|---|
| | 2013 Full Year Semi-Annual Report | 2013 Full Year Utility Tracking System | 2013 Full Year Ex Post Evaluated | 2013 Full Year Ex Post Verified |
| C&I Prescriptive | 2,443 | 2,443 | 1,901 | 1,901 |
| C&I Small Business | 360 | 360 | 283 | 284 |
| C&I Custom | 1,026 | 1,057 | 996 | 995 |
| Res. Lighting | 3,785 | 3,275 | 4,252 | 4,252 |
| Res. Appliances | 302 | 289 | 251 | 250 |
| Res. Appliance Recycling | 604 | 618 | 675 | 676 |
| Res. HVAC | 853 | 711 | 759 | 659 |
| Res. Retrofit - QHEC | 1,034 | 1,102 | 844 | 844 |
| Res. Retrofit - HPWES | 463 | 492 | 137 | 137 |
| Res NC | 462 | 463 | 347 | 409 |
| **Total** | **11,396** | **10,810** | **10,445** | **10,407** |

**Table 1-30: PE Reported, Evaluated, and Verified Coincident Peak Demand Savings (kW) by Program Category—Net**

| Program | Net Savings | | | |
|---|---|---|---|---|
| | 2013 Full Year Semi-Annual Report | 2013 Full Year Utility Tracking System | 2013 Full Year Ex Post Evaluated | 2013 Full Year Ex Post Verified |
| C&I Prescriptive | 1,759 | 1,759 | 1,369 | 1,369 |
| C&I Small Business | 177 | 176 | 139 | 139 |
| C&I Custom | 626 | 645 | 607 | 607 |
| Res. Lighting | 2,536 | 2,194 | 2,849 | 2,849 |
| Res. Appliances | 106 | 101 | 88 | 88 |
| Res. Appliance Recycling | 435 | 445 | 486 | 487 |
| Res. HVAC | 307 | 256 | 273 | 237 |
| Res. Retrofit - QHEC | 879 | 937 | 717 | 717 |
| Res. Retrofit - HPWES | 320 | 339 | 94 | 95 |
| Res NC | 388 | 389 | 291 | 344 |
| **Total** | **7,533** | **7,242** | **6,913** | **6,931** |

## 1.4 Key Issues and Recommendations

This subsection summarizes the various topics and recommendations discussed as part of the program-level verification. Detailed background discussion—including evaluation and verification methods, major corrections to early evaluation drafts, and issues that were not resolved in the verification process—is provided in subsequent sections.

### *1.4.1 All Programs*

- Itron and the Evaluation Team will hold preliminary verification meetings in October/November 2014 to walk through primary data collection and calculation methods, including the CY roll-up calculations. This should reduce the level of effort required after the Draft Evaluation Report from the Evaluation Team is provided in January 2015.
- The development of CY2013 savings estimates was more complicated this year than in past years, driven by the need to accommodate both the PJM and EmPOWER evaluation requirements. This complexity led to concerns about the transparency and reporting of calculations used to estimate CY savings. For the 2014 roll-up, Itron and the Evaluation Team will seek ways to reduce the complexity and how to better facilitate Itron review of the roll-up calculations.

- Previously, two realization rates (RRs) were reported: one RR was based on data quality problems in tracking systems and the other RR was based on evaluation findings related to various savings parameters or methods. Providing these two RRs separately was useful in helping utilities gauge performance of tracking systems. The Evaluation Team should consider reporting the tracking system and evaluation RRs separately. Itron will discuss the feasibility and costs of this proposal with the Evaluation Team as part of 2014 Evaluation Planning.
- The Evaluation Team should include questions about the timing of purchases in all future free rider survey instruments. To date, EmPOWER free ridership surveys have focused on whether or not a measure would have been purchased or installed within one year of the program measure being purchased or installed. While EmPOWER goals are based on gross, not net, savings, the net-to-gross (NTG) ratios from these surveys are then used to develop life cycle savings estimates for purposes of cost effectiveness calculations. Using first year NTG ratios for life cycle cost and benefit estimates overstates life cycle savings.[3] The Evaluation Team concurs with this recommendation.
- The utilities should investigate differences between savings reported in Semi-Annual reports and tracking system data submitted to the Evaluation Team.

### *1.4.2 Residential Lighting*

- The utilities and the implementers should share scanned copies of bulb invoices with the Evaluation Team to allow them to independently verify the accuracy of the program tracking data.
- The utilities should show all parameter values used in reported energy and demand savings calculations in the program tracking data.
- The Evaluation Team uses TRM values to calculate savings, but it does not review tracking system parameters to be able to say why there are differences between the evaluated and tracking system savings estimates. The Evaluation Team should re-create strict TRM-based savings calculations as a check on utility-reported savings and highlight any differences between utility-reported savings and strict TRM-based savings in the Evaluation Report.
- The Evaluation Report should include sensitivity analysis associated with analytical decisions that were made in developing the price response model for NTG estimation. For example, the Evaluation Team decided to exclude data from the model that showed counterintuitive average price elasticity for a given model in a given store (i.e., purchase

[3] Illinois (ComEd) and NYSERDA are two examples of jurisdictions that account for deferred free ridership. It is one of three scores in the NTG algorithm. The deferred free ridership is only counted for the first three years—four years out is considered non-FR for the third score adjustment.

rates rising when prices rise). Performing a sensitivity analysis on the impact of these analytical decisions would increase the transparency of the evaluation. The Evaluation Team has indicated that Price Response Modeling (PRM) is not planned for EY5 but that this recommendation will be followed in EY6 PRM.

- Information on HVAC system vintage and efficiency should be collected as part of onsite data collection in residential light metering studies. This information should be used to inform the Waste Heat Factors for energy (WHFe) and demand (WHFd) estimates in building energy simulation modeling on a forward-looking basis. The Evaluation Team has indicated that WHF analysis is not planned for EY5 but that this recommendation will be followed in EY6 WHF modeling.
- The Evaluation Team should provide a confidence/precision interval around the estimate of the proportion of residential bulbs being installed in non-residential sockets. Because there is sampling associated with the in-store intercepts in the form of selecting certain program participating stores and the sample of program bulb purchasers who agree to participate in the intercept survey, a sampling error term should be calculated for this parameter. This parameter should also be included in the Monte Carlo simulation to generate an overall estimate of the confidence/precision interval for energy and demand savings as a function of sampling error. Note that when sampling error associated with this term is incorporated, this may necessitate increased sample sizes to meet the overall program savings confidence/precision requirements.
- The Evaluation Team should include a discussion of possible sources of bias and steps taken to minimize bias in the savings estimation process. An excellent discussion was included on this topic in the PY2011 Evaluation Report.

### *1.4.3 Residential Appliance Rebates*

- Itron supports the Evaluation Team recommendation to discontinue incentives for electric water heaters.

### *1.4.4 Residential Appliance Recycling*

- NTG survey spillover batteries should include questions asking whether the respondent had received incentives from other programs for these measures.
- The Evaluation Team's appliance recycling regression model should be used in 2014 and program implementers should provide the data needed to run the model. The Evaluation Team should include age and other characteristics that are used as input parameters for the regression model in the Evaluation Report to allow readers to easily trace at least high level calculations.
- Partial use factors should be evaluated as part of the 2014 evaluation.

- The Mid-Atlantic TRM should include the Evaluation Team's regression model as an alternative method for estimating savings. It should be provided alongside the current method for now. The existing TRM kWh savings formula for appliance recycling should include a factor to account for reductions in appliance waste heat.

### *1.4.5 Residential HVAC*

- The utilities should collect *in situ* equipment age, nameplate efficiencies, and if possible, equipment condition. In past Verification Reports, Itron recommended that the utilities track in their program databases information about the efficiency and capacity of replaced equipment. This recommendation was not implemented for the 2013 programs. The lack of adequate evidence on the existing replaced equipment led to the ground source heat pump (GSHP) measure baseline being revised, which significantly decreased the estimated impacts for this measure.
- BGE should remove all peak demand reduction claims for furnace fan motor replacements with Electronically Commutated Motors (ECMs), clarify the conditions for assigning high and low energy savings based on the primary heating source and whether the replacement is paired with either a Central Air Conditioner (CAC) or an Air Source Heat Pump (ASHP) replacement, and ensure that the ECM energy savings are not double-counted. Although BGE improved their scrutiny of this measure, additional improvements are necessary. Both the evaluation and verification reviews agree that there is likely no peak demand reduction due to an ECM replacing an existing furnace fan motor. When ECMs are installed in conjunction with either CAC or ASHP replacements, it is very likely that the equipment performance ratings already capture the ECM impacts and assigning additional ECM savings entails double-counting the impacts.
- The Evaluation Team should subject the BGE natural gas furnace replacement measure with ECM to a more rigorous engineering review since the measure represents close to 7.5% of the EY4 electricity energy savings results for the utility program.
- The EmPOWER utilities should make the HVAC tune-up information data collection and requirements uniform statewide and use the nameplate ratings of the existing equipment to estimate the savings. All of the HVAC service offerings, including HVAC tune-ups and duct sealing, should be integrated into a single offering of HVAC diagnostic and repair services.
- The PE HVAC tune-up measures should report CAC and ASHP equipment separately.
- The NTG evaluation should look at free ridership and spillover over time to determine whether program design and marketing changes are impacting participant free ridership and spillover. Also, spillover determination should include only measures that do not receive financial incentives.

### *1.4.6 Residential Retrofit*

- The Evaluation Team should consider performing at least limited on-site metering for a sample of Home Performance with ENERGY STAR (HPwES) participants.
- Low verification rates for HVAC and Direct Hot Water (DHW) measures should be investigated in the upcoming evaluation cycle. If these low verification rates persist, savings claims may need to be adjusted.
- The Evaluation Team should ensure that the wording of future NTG survey batteries does not bias survey respondents by assuming facts or knowledge that may not be accessible to respondents or the likelihood of free riders. Future surveys should contain more open-ended questions to confirm assumptions before proceeding to quantitative questions used in NTG scoring.
- The Evaluation Team should conduct a CFL persistence study be carried out in the next evaluation because of the need to develop more accurate estimates of lifecycle savings for CFLs, which represent over 70% of the savings for the Residential Retrofit programs.
- The Evaluation Team should investigate the use of smart meter data to develop estimates of peak savings from the HPwES programs. The average load shape profiles currently being used are not likely to be representative of the participant program and are less accurate than using actual metered data.
- A more intensive evaluation effort should be conducted in EY5 to identify the reasons for declining savings per household and lower conversion rates. At the same time, Program Administrators should be encouraged to pilot test new sales strategies or program designs given the lack of success in improving the effectiveness of this program over the last three years.

### *1.4.7 Residential New Construction*

- On-site data collection efforts which focus on improving the quality of Home Energy Rating System (HERS) rater verification and certification efforts and prepares them for enforcement of the International Energy Conservation Code (IECC) 2012 codes and standards. Evaluation efforts should document when shortcomings are found in a sample of homes and track what happens to those homes as the issues are hopefully addressed and remedied by the builder.
- An alternate incentive structure should be considered that pays incentives based upon annual energy savings instead of HERS scores, which are poorly correlated with per-home energy savings.
- Program Administrators should periodically review and adjust, and/or update the reference home rule set to reflect current ENERGY STAR Reference Home Guidelines, applicable IECC requirements, findings from the ICF Baseline Study, and Federal Appliance Standards. Enforcement of the 2015 National Appliance Energy Conservation

Act (NAECA) water heater standards are on the horizon and will have a significant impact on the ability of builders to comply with program requirements.

- Program Administrators should ensure that program requirements and reference home rules do not inadvertently incentivize installation of electric resistance water heaters where natural gas is available.
- The ICF Baseline Study should be expanded to include a larger sample size in underrepresented jurisdictions. Findings should accurately reflect the typical construction characteristics found.
- Program Administrators should engage with the developer of REM/*Rate* to ensure that it includes all appropriate as-built and baseline specifications for the upcoming program cycles for lighting and appliances.
- The peak coincident demand factor should be based on the program savings shape, not on the overall residential load shape. The savings load shape for residential new construction was shown to be twice as large as the Evaluation Team's estimate using the residential load shape.
- Implementers should collect billing usage data for the participant population to either verify new home occupancy before incentives are paid or adjust savings with an occupancy adjustment factor.

### *1.4.8 Commercial & Industrial Prescriptive and Direct Install*

- The Program Implementers and the Evaluation Team should harmonize their descriptions of building types. Building types are a critical parameter for estimating savings for these programs.
- The Evaluation Team should develop or adopt a standardized approach for estimating savings from variable frequency drives. The Northeast Energy Efficiency Partnerships (NEEP) Variable Speed Drives Load Shape Study may be used when it becomes available in 2014, but the Evaluation Team should also consider conducting primary M&V.
- The Evaluation Team should review the approach and underlying self-reported lighting hours of operation that were used to develop the customer self-report ratio (CSRR) to determine if these values should be estimated separately for each utility. Each utility currently uses a slightly different method for estimating the customer self-reported lighting operating hours, but the evaluation treated these methods equally in calculating the overall CSRR.
- The Evaluation Team should consider revising and updating the NTG ratio values for these programs. The current values are several years old now.

- The Evaluation Team should advise the utilities on how they will evaluate T12 to T8 linear fluorescent measures in coming years. The Mid-Atlantic TRM provides an approach for the phase-out. The Evaluation Team should determine if the TRM approach is appropriate for Maryland and, if so, ensure that utility tracking system calculations are consistent with the TRM approach.

### *1.4.9 Custom Commercial & Industrial*

- In future years, the Evaluation Team proposes to give zero savings for Custom projects that do not have sufficient documentation. Clear threshold criteria are needed for when savings would be zeroed due to lack of documentation. Itron will discuss criteria for "sufficient" documentation with Evaluation Team as part of 2014 Evaluation Planning.
- The Evaluation Team should provide guidance to the Program Administrators about the baseline selection process used by the Evaluation Team to ensure baseline selection methods become more consistent with the selection processes and ultimate baseline condition choices of the Evaluation Team.
  - For early retirement (ER) claims, the Evaluation Team should provide guidance on the best practices used to determine remaining useful life (RUL) period based on the information collected through site visits and interviews.
  - The Evaluation Team should inform the implementers that the Evaluation Team will only accept ER claims for those projects with a RUL of two years or greater.
  - For new construction projects, the Evaluation Team should provide the process selecting the applicable code using the facility permit date.
- The Evaluation Team should ensure that the savings calculations for a given project are based on representative conditions found at the facility. When the "as found" site conditions are not able to be considered as valid representation of typical site conditions, the evaluation assumptions of typical site conditions should be based on well-grounded sources (e.g., customer agreements to confirm increase in occupancy or production rates, benchmarked industry averages for the measure/facility type, etc.).
- The Evaluation Team should develop estimates of NTG for the Custom program at the utility level next year. This will provide the Program Administrators with better guidance and the ability to adjust program design practices to reduce the incidence of free riders.

# 2

# Calendar Year Roll-Up

## 2.1 Summary

The bulk of the evaluation and verification effort was devoted to analysis of activity occurring from June 1, 2012 through May 31, 2013 (EY4). The findings from the EY4 analysis (i.e., the EY4 realization rates) were used to adjust savings reported in the utility tracking systems from activity occurring January 1 through May 31, 2013.

Activity occurring between June 1 and December 31, 2013 was not formally evaluated, but utility tracking data were adjusted to reflect findings from the Evaluation Year analysis and any errors in the tracking data. The CY2013 evaluated savings were calculated as the sum of the estimated savings from these two periods: January 1 through May 31 and June 1 through December 31.

Itron reviewed the CY2013 roll-up calculations and, after a few corrections were requested and made, can generally attest to their accuracy. Itron could not reproduce all of the calculations fully for a couple programs, but the differences were small and we concluded were likely due to the differences in program activity between the EY4 period (June 1, 2012 through May 31, 2013) and the period to which the EY4 realization rates were applied (January 1 through May 31, 2013).

The development of CY2013 savings estimates was more complicated this year than in past years. This complexity was driven first and foremost by the need to accommodate both the PJM and EmPOWER evaluation schedule requirements. The task was further complicated by the agreed need to apply realization rates at the measure level to account for changes in program measure shares; in previous years, realization rates were applied at the program level.

The complexity of the CY2013 analysis gives us some concerns about the transparency and reporting of calculations used to estimate CY savings. Itron and the Evaluation Team will seek ways to reduce the complexity of the 2014 roll-up and how to better facilitate Itron review of the roll-up calculations.

## 2.2 Roll Up Methods

Two approaches were used to obtain CY2013 ex-post savings estimates from the Evaluation Team's CY2013 ex ante savings estimates:

- Parameter Update Approach. For all program elements receiving deemed savings, the Evaluation Team applied updated parameters and installation rates from the EY4 report to the tracking data from the later period of the year. According to the Evaluation Report, these programs received a deemed savings update in June 2013, so all of these program elements would be expected to have a different gross realized savings ratio (GRSR) for the period after June than before June.
  - For programs with sufficient tracking data or consistent assumptions, parameter updates were applied to a census of projects. These programs/program elements included Residential Lighting, Appliances, Appliance Recycling, Retrofit–Quick Home Energy Checkup (QHEC), Retrofit–HPwES-Audit, and HVAC.
  - For programs with insufficient tracking data, a sample of project calculators was drawn and desk reviews were used to update the savings parameters and calculate new site-specific savings. These programs included C&I Prescriptive, C&I Small Business (SB), and Master-Metered Multi-family (MMMF).
- Apply EY4 GRSRs: For all program elements receiving site-specific or custom savings, the Evaluation Team applied the GRSRs from EY4 to the program-tracked savings. These programs do not use deemed savings, and would therefore be expected to have relatively constant GRSRs between the first part of the year and the second part of the year. These included Residential Retrofit–HPwES whole house measures, Residential New Construction (RNC), C&I Custom, and C&I Retrocommissioning (RCx).

Table 2-1 summarizes for each program the approach that was used and whether it was applied to a sample or census of program activity.

**Table 2-1: Summary of Evaluation Team EY-to-CY Approaches**

| Evaluation Team Program Group | GRSR Approach | Sample or Census? |
|---|---|---|
| Residential Lighting | Parameter update | Census |
| Residential Appliance Rebate | Parameter update | Census |
| Residential Appliance Recycling | Parameter update | Census |
| Residential HVAC | Parameter update | Census |
| Residential Retrofit-QHEC | Parameter update | Census |
| Residential Retrofit-HPwES-Audit | Parameter update | Census |
| Residential Retrofit-HPwES-WholeHouse | Used EY4 GRSR | NA (Program-level) |
| Residential New Construction | Used EY4 GRSR | NA (Program-level) |
| C&I Prescriptive | Parameter update | Sample + Desk reviews |
| C&I SB/DI | Parameter update | Sample + Desk reviews |
| Master-Metered Multi-family | Parameter update | Sample + Desk reviews |
| C&I Custom and RCx | Used EY4 GRSR | NA (Program-level) |

In the following sections, we summarize our review of the Evaluation Team's CY calculations.

### *2.2.1 Residential Lighting Program*

The calculation of Residential Lighting program CY savings was a relatively straightforward process. Because the EY and the CY each encompass a 12-month cycle, most of the parameter values used to calculate energy and demand savings were unaffected in the rollup. For example, Hours of Use (HOU), which is derived as a weighted average across a 12-month cycle based on a sinusoid regression to account for seasonal changes in daylight hours, was unaffected. Similarly the parameter values for Peak Load Coincidence Factor, In-Service Rate (ISR), cross-sector installation of residential program bulbs in non-residential sockets, lighting-HVAC interactive effects, and NTG were all unaffected by the shift from EY to CY.

The only energy and demand savings calculation parameters affected by the shift were the numbers of program bulbs and the Delta Watts Multipliers (DWM). The DWM is derived as a weighted average across program bulbs and is therefore affected by the specific mix of program bulbs sold by a given utility in a given time period. The Evaluation Team derived separate DWM by utility for the first and second halves of the CY and applied these to the specific bulb sales that took place in each half of the CY.

Itron traced the derivation of the utility-specific and time period-specific DWM values and confirmed their accuracy. We were then able to fully replicate evaluated CY savings estimates.

### *2.2.2 Residential Appliance Recycling Programs*

The calculation of CY savings for the Residential Appliance Recycling programs was straightforward. We were able to fully replicate the CY savings calculations using the tracking data received from the Evaluation Team.

### *2.2.3 Residential Appliance Rebate Programs*

The calculation of CY savings for the Residential Appliance Rebate programs was straightforward. We were able to fully replicate the CY savings calculations using the tracking data received from the Evaluation Team.

### *2.2.4 Residential HVAC Programs*

The Evaluation Team calculated calendar year results by applying the same savings calculation formulas and parameter assumptions from the EY4 impact evaluation to the calendar year reported measures. Itron followed the same basic methodology to verify the 2013 CY results, but used the key parameter values from the utility tracking data in lieu of the Evaluation Team's assumed average "kW Saved per Ton" parameter values for CAC and ASHP.

We identified only one small problem with the CY roll-up for the residential HVAC programs: it appears that the Evaluation Team did not zero out the BGE furnace replacement measure demand reduction claims for the January 1 through May 31 period in the CY roll-up report as was done in the final EY4 Evaluation Report. Otherwise, Itron was able to fully replicate the CY savings estimates.

### *2.2.5 Residential Retrofits*

CY 2013 savings for Residential Retrofit programs were calculated differently for HPwES and QHEC. For QHEC, the Evaluation Team multiplied quantities of each CY measure by the EY4 evaluated unit energy savings values. This implicitly assumes that installation rates, baselines, and NTG ratios were the same during both periods.

To calculate HPwES CY savings, the Evaluation Team applied realization rates from the EY4 analysis to all CY households that received the types of measures included in the EY4 billing analysis. For households that received only direct install measures, the team applied the EY4 measure level unit energy savings values.

We did not attempt to fully replicate these calculations. The assumptions and methods used for the Retrofit programs CY estimates appeared reasonable, however. We reviewed measure-level participation and savings data for the two separate reporting periods—January 1 through May 31 and June 1 through December 31— to search for any anomalies or large shifts in savings per measure installed or per participants in these two time periods. We found some significant

differences in reported activity (read measure installations) and savings levels between the first six and last six months by program and by utility but the evaluated savings estimates appeared to accurately reflect these differences.

### *2.2.6 Residential New Construction*

Itron review of the RNC CY calculations found some differences between the ex ante claimed savings from the EY4 Evaluation Report and the values in the CY2013 roll-up report. While we were unable to fully replicate the evaluation calculations, the differences were small and we concluded that the differences were likely due to the different time periods reported in the EY4 Evaluation Report as compared to the CY roll-up report.

### *2.2.7 C&I Prescriptive, SB Direct Install and PEPCO MMMF*

Using the tracking data provided by the Evaluation Team, Itron was able to fully replicate the CY savings calculations for all five of the utilities' Prescriptive and SB Direct Install programs, as well as PEPCO's MMMF program.

### *2.2.8 C&I Custom*

The methods used for the CY calculations are straightforward and reasonable, but Itron did not receive the Evaluation Team's spreadsheet calculations in time to allow full replication of the calculations.

# 3

# Residential Lighting Programs

## 3.1 Verification Summary

The goal of this analysis is to verify the evaluated energy savings and peak demand savings realization rates produced by the Evaluation Team for the EY4 EmPOWER Residential Lighting programs. Itron performed a review of the methods, results, and recommendations included in the Draft Evaluation Report.

The verified gross and net energy and demand values shown in Table 3-1 are based on having reviewed and re-created individual savings parameter values and overall savings calculations as described in this Verification Report. Based on review of the EY4 Final Evaluation Report, we accept the realization rates and savings estimates for the Residential Lighting programs in full and recommend no adjustments to the savings estimates. Table 3-1 presents a high-level comparison of gross evaluated savings estimates from the Evaluation Team and verified savings from Itron.

**Table 3-1: Summary of Evaluated & Verified Gross Savings—Residential Lighting Programs in EY4**

| | Gross Impacts | | | | Net Impacts | | | |
|---|---|---|---|---|---|---|---|---|
| | Evaluated | | Itron-Verified | | Evaluated | | Itron-Verified | |
| Utility | MWh | kW* | MWh | kW* | MWh | kW* | MWh | kW* |
| BGE | 209,587 | 27,226 | 209,587 | 27,226 | 144,615 | 18,786 | 144,615 | 18,786 |
| PEPCO | 107,099 | 14,153 | 107,099 | 14,153 | 66,401 | 8,775 | 66,401 | 8,775 |
| DPL | 22,777 | 3,104 | 22,777 | 3,104 | 15,944 | 2,173 | 15,944 | 2,173 |
| PE | 30,707 | 4,079 | 30,707 | 4,079 | 20,574 | 2,733 | 20,574 | 2,733 |
| SMECO | 24,153 | 3,128 | 24,153 | 3,128 | 15,216 | 1,971 | 15,216 | 1,971 |
| **Total** | **394,323** | **51,690** | **394,323** | **51,690** | **262,751** | **34,437** | **262,751** | **34,437** |

* Utility Coincident Peak Savings: Peak demand savings are in reference to utility-defined peak period rather than PJM-defined peak period.

## 3.2 Program Summary

All five EmPOWER Maryland utilities offer Residential Lighting programs in which incentives are paid upstream to lighting manufacturers. These incentives are passed down through retailers to customers in the form of lower prices on qualified energy-efficient lighting products. By order of the Maryland PSC, all five utilities offer similar Residential Lighting programs with incentives that can be negotiated up to a specific cap.

In EY4, approximately 7.5 million energy-efficient lamps were sold through EmPOWER Residential Lighting programs. Energy-efficient lamps were sold through home improvement channels, warehouse channels, and mass merchandising (or "big box") channels. The Residential Lighting programs contributed 46% of evaluated CY EmPOWER MWh savings and 39% of demand savings.

## 3.3 Evaluation Summary

The EY4 Evaluation Report represents the fourth year in which the Evaluation Team has evaluated the residential lighting programs across all EmPOWER utilities. The Evaluation Report endorses and is structured around the algorithms for gross lighting energy and demand savings as laid out in the Mid-Atlantic TRM version 3.0. The basic structure of the evaluation has been to ascertain the most appropriate value for each of the parameters in the TRM savings algorithms, apply these parameter values to incented bulbs as documented in the program tracking data, and compare the resulting savings calculations to utility reported savings in the form of an RR that shows evaluated savings relative to reported savings. The Evaluation Report approach and verification finding for each parameter are reported below, along with a review of overall gross and net savings calculations.

### *3.3.1 Program Bulb Sales*

The number of program bulbs sold is a fundamental parameter in driving program savings. The Evaluation Team received the program tracking data from the Program Implementer (ICF for all utilities except PE, in which case it is Honeywell). These tracking data include reported energy and demand savings values for each record, where each record represents program bulb sales for a particular bulb model through a specific retail store on a specific date. The tracking data included 145,482 records, representing 7.7 million statewide bulb sales. Tracking data from each utility included bulb description, stock-keeping unit (SKU) number, bulb-rated wattage, quantity, and date purchased. The Evaluation Team reviewed the tracking data from each EmPOWER utility for internal consistency and noted any typographical errors in pricing fields, incorrect bulb type designations, and SKU numbers that did not match between datasets.

### *3.3.2 Delta Watts*

The difference in wattage between program bulbs and the inefficient bulbs they are expected to replace was calculated in the evaluation as a DWM. This is the calculated difference in wattage between baseline and measure, expressed as a multiple of the efficient program bulb wattage.

The overarching approach to the DWM calculation was to assign baseline wattage based on bulb lumens, generate a delta watts figure for each record as the difference between baseline wattage and program bulb wattage, and calculate the weighted average DWM for each utility accordingly. To determine the appropriate baseline wattage for a given program bulb, the Evaluation Team used three main data sources.

The Evaluation Team used the Environmental Protection Agency (EPA) ENERGY STAR-Qualified Products List to find lumens for the large majority of program bulb models and merged these onto the program tracking data by model number.

For bulb models not listed in the EPA-qualified products list, the Evaluation Team used the relationship between CFL wattage and lumens based on a linear regression of known models to extrapolate lumens for unknown models from CFL wattage. This was also done separately for LED bulbs.

The Evaluation Team then used lumen bins to assign baseline wattage. For general service lamps covered under federal Energy Information and Security Act (EISA) legislation, the Evaluation Team used lumen bins provided by EISA to assign appropriate baseline wattage. For reflector lamps, the Evaluation Team created lumen bins and corresponding baseline wattages based on a table of minimum reflector lamp efficacy requirements (expressed as lumens per watt) provided by the U.S. DOE Office of Energy Efficiency & Renewable Energy (EERE) for reflector lamps in different nominal wattage ranges. The Evaluation Team noted that a variety of specific reflector bulb types are excluded from the EERE guidelines and made accommodations as such in the spreadsheet-based calculations of delta watts. The Evaluation Team further noted that EISA legislation came into effect partway through EY4 for 75W incandescent lamps. That is, starting January 1, 2013, the manufacture or sale of 75W traditional incandescent lamps was no longer legal in the U.S., and the Evaluation Team used the wattage of EISA-compliant halogen bulbs (53W) as the baseline for relevant program bulbs sold after January 1, 2013. Using this series of nested specifications, the Evaluation Team calculated the difference in wattage between the baseline and efficient case for each record, expressed this difference as a multiple of the efficient bulb wattage, and applied these values on a record-by-record basis to the tracking data when generating the overall savings estimates.

### *3.3.3 Hours of Use and Peak Load Coincidence Factor*

For bulbs installed in residential sockets, the Evaluation Team carried forward the average daily residential HOU value of 3.0 and peak load coincidence factor (CF) of 0.0919 that were derived in the EY3 evaluation. The EY3 HOU and CF estimates were based on data collected in the 2010 and 2011 light metering studies. In the EY3 evaluation, metering data were extrapolated to develop estimates of annual HOU based on a sinusoid regression approach. The Evaluation Team currently has another residential light metering study underway, and the data from that study will be available to inform updated HOU and CF values for EY5.

For the portion of program bulbs installed in non-residential sockets, the Evaluation Team used an average HOU value of 10.08 and CF of 0.72 for non-residential bulbs provided by the Evaluation Team. The Evaluation Report cites the EmPOWER Maryland 2013 Draft Evaluation Report for the C&I Prescriptive & SB programs as the source for this value and for discussion of the methods used in its derivation.

Based on an estimate of 5% of program bulbs going into non-residential sockets (see Cross-Sector Installation of Program Bulbs in Non-Residential Sockets subsection below), the Evaluation Team then generated a weighted average HOU value for all bulbs of 3.35 by applying the above HOU values to the proportions of program bulbs installed in residential and non-residential sockets. Similarly, the Evaluation Team generated a weighted average CF value for all bulbs of 0.1233 by applying the above CF values to the proportions of program bulbs installed in residential and non-residential sockets

### *3.3.4 In-Service Rate*

For the program bulb ISR, the Evaluation Team carried forward the value of 0.88 that was derived in the EY2 evaluation. The derivation of that value is based on data from 86 homes, 6,565 total lamps, 6,054 total lighting sockets, 1,077 total CFLs, and 584 lamps in storage. During the onsite visits conducted during the EY2 evaluation, bulb inventories included noting bulb types in sockets as well as in storage. The ISR was calculated as the number of CFLs that were in sockets divided by the sum of CFLs in sockets and in storage. Since there is no reliable way to know for certain whether a given CFL is a program bulb, this installation rate of CFLs generally is assumed to be representative of the installation rate of program CFLs.

In parallel with the ISR value of 0.88 for bulbs purchased and installed in EY4, the Evaluation Team also calculated savings that incorporate bulbs purchased and stored in EY2 and EY3 and installed in EY4. To derive these estimates, the team focused on the ISR of 0.88 that was applied in the EY2 and EY3 evaluations. Following guidance from the Uniform Methods Project protocol, the Evaluation Team noted that the lifetime installation rate of CFLs can be estimated at 97% (it can be presumed to be this high or higher for LEDs). They also noted that approximately half of stored bulbs are installed in each of the two years following the year of

purchase to reach this lifetime installation rate. Separately for bulbs purchased and stored in EY2 and in EY3, the Evaluation Team used the annual rate at which stored bulbs are installed—49.5% of stored bulbs per year—and multiplied it by the 12% of program bulbs not installed in the year of purchase. This yielded a value of 5.95% for the percentage of program bulbs from EY2 and from EY3 that were installed in EY4. The Evaluation Team applied the EY4 savings parameter values to calculate savings from these bulbs and added these to EY4 bulb purchases to generate overall savings estimates that include these delayed installations.

Itron accepts the counting of savings from bulbs purchased and stored in prior program years and installed in the current evaluation year.

### *3.3.5 Cross-Sector Installation of Program Bulbs in Non-Residential Sockets*

The Evaluation Team carried forward the value of 5% non-residential installations that was derived in the EY2 evaluation. This value was developed from in-store intercepts data, conducted with 445 customers in 2011 purchasing light bulbs at 20 participating stores. The EY2 Evaluation Report notes that of the 1,128 program bulbs represented by completed intercepts, 1,058 were destined for residential sockets, 55 for business sockets, and 15 for other applications. Data were not collected in that evaluation regarding the specific business type in the case of business installations.

### *3.3.6 Lighting-HVAC Interactive Effects: Waste Heat Factor*

As is described in the Evaluation Report, lighting waste heat contributes to home heating in the winter and adds to the cooling load in summer. Estimates of the size of this effect are characterized as Waste Heat Factors for energy (WHFe) and demand (WHFd) and are incorporated into the overall lighting savings algorithms. In EY4, the Evaluation Team retained the same overall method for estimating these factors and refined the approach relative to EY3. As in EY3, the overall approach was to build an energy simulation model using the BEopt™ front end of the DOE2 building energy simulation model from the U.S. DOE.[1] Whereas in EY3 the heating component of the simulation model was limited to heat pump heating systems, the

[1] Building Energy Optimization software (BEopt) was developed by the National Renewable Energy Laboratory to facilitate application of the U.S. DOE's DOE2 building energy simulation model. The evaluation team first used BEopt in EY3 to develop an hourly building energy simulation model built around a prototypical Maryland home with lighting schedule input from the PY2011 light metering study data, HVAC data from the PY2010 EmPOWER HVAC monitoring study, and iterative adjustment of model parameters such as window orientation and the amount of thermal mass in the building to calibrate the model to billing data from 50 residential non-gas customers in Maryland. See the EY3 Residential Lighting Evaluation Report and Verification Report for additional details on methods and findings from that approach.

EY4 model also incorporated electric resistance heating and used equipment saturation values from the 2011 KEMA Maryland Energy Baseline Study[2] to generate weighted average values.

The Evaluation Team reflected in their model the weighted averages of space cooling and space heating technologies by housing type within each utility service territory. The model also included weighted averages of home size, number of floors, home foundation type, window type, cooling system efficiency, vintage of central air conditioning system, and a correction for waste heat exiting through unconditioned spaces in the case of recessed lighting. These data came from a combination of the KEMA Baseline Study, the U.S. Census, the U.S. Energy Information Administration, and the Evaluation Team's own bench testing of recessed lighting in EY3. The modeling team generated separate building energy simulation models using data from each of five meteorological stations across the state using current meteorological data and generated weighted average WHFe and WHFd values that reflected the weather conditions unique to each service territory. Similar to EY3, the models were calibrated based on billing data from a subset of utility customers. The net effect of these refinements was to slightly lower the values for WHFe and WHFd relative to EY3.

### *3.3.7 Gross Savings Calculations*

The Evaluation Team's approach to calculating gross energy and demand savings was to derive an evaluation-based estimate of the most appropriate value for each savings parameter (as described in subsections above) and to apply these values to all program bulbs to yield overall savings. The Evaluation Team took the bulb counts by utility and by bulb type from the utility tracking data. Bulb-specific delta watts values were applied to each record. A single weighted average HOU, CF, and ISR were applied to all records for all utilities. Utility-specific WHFe and WHFd values were applied to all records. Applying these parameter values yielded energy and peak demand savings for each individual record in the tracking data, which were summed by utility to yield total evaluation-based ex post energy and demand savings.

### *3.3.8 Net Savings Estimation*

The Evaluation Team used a price elasticity approach to estimate program NTG ratios for each utility. The basic premise of the approach was to use observed patterns of sales rates and price changes (regardless of whether the price changes represented the program incentive or came from other sources) to generate estimates of price elasticity for each bulb type. These elasticity estimates were then applied to the program incentive amount to yield an estimate of program impact, or NTG.

[2] KEMA, Inc., *Maryland Energy Baseline Study*, Residential Sector, prepared for the Maryland Public Service Commission, February 23, 2011.

Data for this price response model were provided by the Program Implementer at a greater level of detail than had been available in prior program years, including sales data by month for each unique combination of retail location, package, bulb model number, and incentive level. These data came in the form of markdown reports that listed bulb sales at the retail location level, program sales trackers that listed prices at the retail chain level, and a list of promotional events and dates.

The Evaluation Team developed a regression model that effectively attributes different shares of a bulb's overall price to the different factors driving that price, including original retail price, incentive provided by utility, target retail price, number of bulbs per package, bulb type designation, and retailer. The Evaluation Team cleaned the data and used quality control criteria to identify and remove records associated with issues that could distort the model. An example of this is looking into cases of zero sales for a bulb and determining whether this is due to lack of demand, stocking abnormalities, or faulty data.

In developing the model, the Evaluation Team controlled for a number of factors. For example, they controlled for background seasonal patterns in bulb sales by normalizing the model to the seasonal patterns in bulb sales observed in the prior program year. Some variables, such as in-store promotion events, were tested in the model and dropped from the final model due to having no significant effect on bulb sales rates.

The regression model ultimately yielded a set of equations by bulb type and utility that predict program bulb sales rates as a function of a variety of other parameters, including bulb price. By entering the average incentive amount and price elasticity by retail channel and bulb type into the model, the Evaluation Team could predict bulb sales with and without the utility incentive program. The difference between these two rates relative to sales with the program represents the NTG value.

The final models produced from the regression took the following form:

$$\ln(Q_{it}) = \beta_1 + + \sum_{\theta,\delta}(\beta_{\theta,\delta}\left[ln(P_{it}) * (Retail\ Channel_{\theta,i}) * (Bulb\ Type_{\delta})\right]) + \sum_{\pi}(\beta_{\pi} ID_{\pi,i}) + \sum_{\delta}(\beta_{\delta} Time\ Effects_{\delta,t}) + \varepsilon_i + \gamma_t$$

*Where*:

| | | |
|---|---|---|
| ln | = | Natural log |
| Q | = | Quantity of bulb packs sold during the month |
| P | = | Retail price in that month |

| | | |
|---|---|---|
| ID | = | Dummy variable equaling 1 for each unique retailer and model number; 0 otherwise |
| Time Trend | = | Quantitative trend representing the impact of sales trends not related to the program[3] |
| $\varepsilon_i$ | = | Cross-sectional random-error term |
| $\gamma_t$ | = | Time-series random-error term |

### *3.3.9 Confidence, Precision, and Error*

The Evaluation Team calculated confidence intervals based on the values of HOU, CF, and ISR for bulbs installed in residential and commercial sockets. These were the only variables in the gross savings algorithms that involved sampling. Overall relative precision and standard error calculations were developed using Monte Carlo simulation.

The Evaluation Team calculated the standard deviation around per-bulb energy and demand savings, and calculated the lower limit confidence interval based on 90% one-tailed distribution. When applied to overall energy and demand savings by utility, these yielded an energy savings standard error of 0.06 and relative precision of 0.07. Utility coincident peak demand savings yielded a standard error of 0.13 to 0.15 across utilities, with relative precision of 0.13.

In the price response model, standard error by bulb type and retail channel ranged from 0.15 to 0.29. For overall NTG estimation by utility, the Evaluation Team calculated "bootstrap" standard errors to determine the NTG ratios' sensitivity. In this approach, the Evaluation Team drew 500 new samples (with replacements) from the original data, estimating coefficients with each sample and calculating a new NTG ratio.

## 3.4 Verification Findings

The basic structure of the verification process was to trace the development of the individual input parameters that were used in the energy and demand savings algorithms and to re-create savings values by utility based on the final parameter values. Itron used the Evaluation Report, supporting data and workbooks, and conference calls with the Evaluation Team to clarify methods and data sources. The verification findings are laid out by input parameter in the subsections that follow.

[3] The time trend for this analysis represents shifts in sales due to non-program related seasonality. It was calculated using normalized sales of program bulbs in the previous year that experienced no in-store promotions or price changes.

### *3.4.1 Program Bulbs*

Itron thoroughly reviewed the program tracking data that were provided by the Evaluation Team. While the tracking data appeared comprehensive in terms of bulb sales records, with no self-evident problems or gaps, Itron notes that there is currently no meaningful mechanism by which the Evaluation Team can triangulate the tracking data against an independent data source to verify program bulb sales.

In addition, Itron notes that there is no place in the Evaluation Report where program bulb sales are specifically listed, either at the aggregate level or broken out by utility, bulb type, and other groupings. This hampers the transparency of the evaluation and makes it difficult to re-create savings values. This is in contrast to EY3 where the Final Evaluation Report included program bulb sales corresponding to each DWM value by utility and bulb type.

Itron recommends that a table be added to future Evaluation Reports that shows program bulb sales by utility, along with the final, weighted average values of all savings parameters used in the evaluation. This would enable a reader to more clearly understand the elements going into the calculation of total evaluated savings. In response to feedback from Itron, tables were added to the Evaluation Report that show the final residential and non-residential savings parameters by utility. However, program bulb sales overall and by utility are not listed in the Evaluation Report.

Itron strongly recommends that the utilities and implementer be required to share scanned copies of bulb invoices with the evaluator and that the evaluator use these as a tool for independently verifying the contents of the program tracking data on an annual basis.

### *3.4.2 Delta Watts Multiplier*

The Evaluation Report notes that lumens were identified for 80.6% of the 7.5 million program bulbs that were analyzed in the development of DWM by utility.

Itron was provided with the workbook that the Evaluation Team developed to generate the DWM estimates. Itron reviewed the record-level delta watts calculations and their sources. This review highlighted the fact that the Evaluation Team used a different approach than what is in the TRM for assigning baseline wattage. For example, for reflector lamps, the Evaluation Team adapted minimum efficacy (lumens/watt) standards from the U.S. DOE EERE data book, which yields different lumen bins for baseline wattage mapping than those in the TRM that are developed from an ENERGY STAR equivalence table.

Itron agrees that the data sources used for calculating DWM in the evaluation are appropriate. However, the reasons for advocating a different baseline wattage calculation approach than what is in the TRM should be described in the Evaluation Report.

Itron continues to recommend that the approach to communicating about the difference in wattage between program and baseline bulbs be shifted from a DWM approach to a more intuitive and simple delta watts approach (i.e., communicating the average difference in wattage between program and baseline bulbs as a simple weighted average number of watts, rather than a multiplicative ratio relative to the efficient wattage). If the DWM is kept in place, the Evaluation Team should include in the Evaluation Report the weighted average program bulb wattage for each utility, as this is necessary for re-creating the calculations of energy and demand savings from the constituent parameter values.

### *3.4.3 Hours of Use and Peak Load Coincidence Factor*

Given that a residential light metering study is currently underway and its data are not yet available for applying to program evaluation, Itron agrees with carrying forward the 3.0 average daily HOU and 0.0919 CF values for bulbs installed in residential sockets from the EY3 evaluation in EY4. This decision is further corroborated by the fact that 3.0 is the value in the Mid-Atlantic TRM based on the data from prior years' metering studies. Itron commends the data collection and analytical approach that the Evaluation Team has used for estimating this factor in the past.

By using the average daily HOU of 10.08 provided in the Evaluation Report for bulbs installed in non-residential sockets, as well as the estimate of 5% of program bulbs going into non-residential sockets, Itron was able to recreate the overall weighted average 3.35 HOU used in the evaluation-based energy and demand savings calculations.

Itron looked in the utility program tracking data and noted HOU values corresponding to 1.01, 2.44, and 3.00 average daily use. Itron inquired with the Evaluation Team about this, but the Evaluation Team said they were not aware of these varying values or their source. This observation reinforces the point made elsewhere in this Verification Report that in Itron's opinion, all parameter values used by utilities to generate reported savings should be shown for each record in the utility tracking data, and the evaluation should include calculating strict TRM-based savings values and comparing these with utility-reported values as a check on the utility calculations.

For clarity in the Evaluation Report, Itron recommends that the Evaluation Team consider calculating and showing total energy and demand savings estimates separately for bulbs installed in residential and non-residential sockets, and add these values to show overall program energy and demand savings. This is more transparent and intuitive than generating weighted average values for HOU, WHFe, WHFd, and CF and applying them to all bulbs.

### *3.4.4 Cross-Sector Installation of Program Bulbs*

The Evaluation Team re-created the calculation of the proportion of program bulbs going into residential and non-residential sockets based on the EY2 intercepts data from which it was originally developed and derived a value of 4.94% of program bulbs going into non-residential sockets.

Itron agrees with the application of this estimate to the EY4 evaluation, as well as with the effort currently in development by the Evaluation Team as part of the EY6 evaluation to provide an updated and more rigorously derived estimate for non-residential installation of program bulbs. The proposed approach to deriving this estimate, for which the Evaluation Team is seeking support from participating program stakeholders, is to conduct an online survey with program bulb purchasers advertised via stickers on program bulb packages

Itron notes that the original data collected to estimate this factor in EY2 did not include business type. They also represent data collected on a program whose structure has changed somewhat in the intervening years through the introduction of LEDs, concentration of the program in a narrower range of retail channels, and increased emphasis on specialty CFLs. If the current effort to conduct an online survey is successful in deriving a rigorous estimate of non-residential installations, this updated value should be applied to EY6 savings. Otherwise, a new round of in-store intercept surveys should be planned to derive an updated estimate, or another appropriate and current data source should be found to provide this estimate.

### *3.4.5 In-Service Rate*

Itron reviewed the methodology associated with developing the ISR estimate of 88% in EY2. Itron concurs with the onsite data collection methods that were used to generate the estimate of first year installation of program bulbs. Based on secondary literature review, including data collection conducted by Itron for other programs, Itron also concurs with the estimate that 97% of program bulbs are eventually installed in sockets. Thus, Itron concurs with the overall value for the ISR as well as with the adjusted calculation associated with bulbs purchased and stored in prior program years and installed in the current program year. Itron recommends that the Evaluation Team clarify the description of the Late Installations section of the Evaluation Report, as the current language does not provide a complete description for how the savings calculations for late installations are carried out.

Itron notes that the same ISR of 0.88 is assumed for the non-residential installation of program bulbs. Itron is not aware of a source for an independent estimate of an ISR in commercial settings and agrees with applying the residentially derived ISR value to these bulbs, but notes that additional information would be useful on this front.

Itron also notes that collecting data about the ISR is a component of the current residential light metering study that is underway and that information from this study will form the basis of the EY5 ISR value.

### *3.4.6 Lighting-HVAC Interactive Effects: Waste Heat Factor*

Itron carefully reviewed the assumptions, data sources, and distributions of technologies, home characteristics, and weather data that served as inputs to the building energy simulation models. Itron notes that this model and its results have emerged from an iterative process of model refinement over the past three program years. Each year has represented an important and qualitative improvement over the prior year's effort to estimate these factors. Itron verifies that the modeling approach, the input data sources, and the model calibration all appear appropriate and reasonable. As such, Itron makes no recommendations for changes in the calculation of the WHFe and WHFd factors and has no recommendations for improving the approach.

### *3.4.7 Gross Savings Calculations*

To re-create the gross energy and demand savings calculations that underlie the Evaluation Report, Itron applied all parameter values by utility to the utility tracking data. Itron was able to re-create the savings calculations to within 0.1% for each utility. Our recommendations regarding gross savings calculations center on increasing transparency, both in the utility tracking data and in the Evaluation Report.

As mentioned elsewhere in this Verification Report, for the sake of increased transparency and increased evaluability, Itron strongly recommends that the utilities and implementer be required to provide in the tracking data all savings parameter values that go into the reported savings calculations for each record. This would enable the evaluator to observe how accurately and consistently the utilities are applying TRM values for each savings parameter to program bulb sales. Currently, the evaluator generates evaluation-based estimates of the appropriate values for each savings parameter and generates program energy and demand savings estimates based on these values. The only formal relationship that is reported between the reported and evaluated savings values is an RR that compares the two values. This RR blends together potential differences between the reported and evaluated calculations in terms of DWM, HOU, ISR, cross-sector installation of program bulbs, and WHF. As such, putting all of the focus on RR diminishes the practical value of the evaluation in terms of improving program design and in terms of improving the ways in which utilities track their own programs.

This lack of transparency hampers the potential conversation between the evaluator, the PSC, and the utilities regarding the appropriate calculation of reported savings and possible errors in the tracking data at the level of individual records. The utilities receive no verification, per se, on whether savings parameter values are being appropriately applied to program bulb sales. This

is particularly important given the increased complexity of baseline wattage calculation during EISA phase-in and the resulting potential confusion in how reported savings were derived.

As a specific example of this need for data transparency, PSC staff brought to Itron's attention a table from the BGE Q3-Q4 Semi-Annual Report that showed kWh savings per bulb for CFL and LED bulbs. Notably, the table showed distinctly lower savings per bulb for LEDs than for CFLs, when it would be expected that LEDs would generally have higher kWh savings per bulb than CFLs. The Evaluation Team was not aware of this issue, did not have the data in hand that would allow them or us to understand this apparent anomaly, and was unable to obtain the data from the implementation contractor.

### *3.4.8 Net Savings Calculations*

Itron was provided with the detailed description of the price response model methodology in the Evaluation Report and was also provided with a workbook showing the key inputs to, structure of, and outputs from the model. In addition, Itron asked the Evaluation Team a number of targeted questions during conferences calls. Itron reviewed the structure of the model inputs, as well as the predicted savings values by utility with and without the incentive program emerging from the model, and re-created and confirmed the NTG calculations by utility.

The price elasticity approach used by the Evaluation Team to estimate NTG is innovative and sensible. It will also benefit from a number of important refinements over time. At a fundamental level, using price elasticity to gauge program impact roots the evaluation in actual rates of purchase that coincide with different price levels. In contrast to self-report estimates of program impact that rely on customers to predict what they would have done in the absence of the program, this approach rests exclusively on actual, observed correlations between price and purchase rate.

Itron supports and agrees with the price elasticity approach to NTG and verifies that the NTG values by utility that were calculated from the available data appear correct. While this approach has important advantages over a customer self-report approach, there are a number of ways in which this approach can and should be refined and improved over time, as follows:

- As noted in the Evaluation Report, if at all possible, the Evaluation Team should collect and incorporate price data by SKU for program and non-program bulbs over time into the model.
- Itron notes that the Evaluation Team built the price response model in part around sales tracker data provided by the Program Implementer, and the Evaluation Team also conducted shelf surveys in participating stores. Price data from the shelf surveys should be directly compared against the prices in the sales tracker to verify consistency between these sources.

- The Evaluation Report should contain further elaboration of the methods used to develop the price response model. For example, the degree of observed price variation that underlies the model should be listed in the Evaluation Report. In a follow-up email to Itron, the Evaluation Team noted that 32% of SKUs had some price variation and that these made up 62% of total sales across the four utilities included in the analysis. They also acknowledged that the final model extrapolates rates of change in bulb purchasing beyond the range of price changes actually observed in the raw data. Displaying both the observed and extrapolated rates of change in sales as a function of bulb price in the Evaluation Report at greater detail would improve transparency in this area.
- It is not clear from the Evaluation Report or supporting workbook whether the effect of pack size on price was appropriately accounted for in the regression. While it is included as a variable in the regression model, and a separate sensitivity analysis is performed that isolates the model to one to three bulbs per pack, it may be warranted to run separate models altogether on different pack sizes to truly isolate this major effect on price.
- Itron recommends that the Evaluation Report include sensitivity analysis associated with analytical decisions that were made in developing the price response model. For example, the Evaluation Team decided to exclude data from the model that showed counterintuitive average price elasticity for a given model in a given store (i.e., purchase rates rising when prices rise). Performing a sensitivity analysis on the impact of these analytical decisions would increase the transparency of the evaluation. The Evaluation Team has indicated that this recommendation will be followed in EY6 when price response modeling is revisited.
- Consider including data from a multi-year period when normalizing the regression model for seasonal effects on bulb sales rates. The EY4 regression model uses seasonal sales patterns from the prior year for normalizing. Data taken from a longer period would provide a smoother and more reliable indicator of seasonal sales patterns.
- On an ongoing basis, the Evaluation Team needs to make a strong effort to minimize any potential systematic bias in the price response model, and be aware of and articulate any systematic bias that remains and cannot be eliminated from the model.

Itron offers additional questions that may be useful in guiding the structure of this approach to NTG looking forward:

- Is there any way to confirm that, from a price elasticity standpoint, a price drop at a selected store due to a program incentive has the same effect on purchaser behavior as the same size price drop across all stores?
- Looking forward, how are you going to take account of other market dynamics besides price changes in accounting for changes in sales rates?

- Is each retail channel represented by reasonably equivalent stores for participating and non-participating?

In the Evaluation Report, the Evaluation Team notes as a central finding that "the average reduction in price increased from one half to two thirds of the original retail price. This increase in incentive is the key driver leading to the decrease in free ridership from EY3."

When Itron inquired about the notable increase in incentive from an average of 50% of original price to an average of 67% of original price, the Evaluation Team sent Itron the data shown in Table 3-2 which demonstrates the relationship between the NTG ratio and the average percent discount from the original price by utility and retail channel. NTG and rebate percentage in this table are very highly correlated, with a Pearson Product-Moment Correlation Coefficient of 0.945.

**Table 3-2: Relationship between NTG Ratio and Average Percent Discount from Original Price by Utility and Retail Channel**

| Utility | Retail Channel | Net of Free Ridership | Percent Rebate of Original Price |
|---|---|---|---|
| BGE | Home Improvement | 44% | 49% |
| | Warehouse | 89% | 73% |
| | Other | 40% | 51% |
| PEPCO | Home Improvement | 45% | 50% |
| | Warehouse | 82% | 66% |
| | Other | 41% | 55% |
| DPL | Home Improvement | 49% | 51% |
| | Warehouse | 91% | 81% |
| | Other | 44% | 55% |
| PE | Home Improvement | 44% | 49% |
| | Warehouse | 87% | 70% |
| | Other | 42% | 53% |
| SMECO | Home Improvement | 42% | 46% |
| | Warehouse | 85% | 69% |
| | Other | 46% | 55% |

The Evaluation Team did not provide further comment on whether a significant increase in funds had been made available to the utilities to increase the average incentive amount per bulb or whether the retail price dropped relative to the incentive amount.

Itron also notes that by increasing the size of the incentive on program bulbs, this may increase the NTG ratio, but it also means that dollars spent per free rider will go up (since larger

incentives mean more money spent on a person who would have purchased the bulbs anyway). From a Total Resource Cost test perspective, the question is whether the reduction in the free ridership rate offsets the increase in program and participant costs.

### *3.4.9 Confidence, Precision, and Error*

In general, Itron agrees with the approach used to calculate confidence and precision associated with sampling error. However, Itron notes that the proportion of program bulbs going into non-residential sockets is subject to sampling error. This value was assumed to be known in the calculations of error and precision in the evaluation and should instead be incorporated in the Monte Carlo simulations to generate overall confidence/precision and standard error estimates.

In addition, Itron notes that the EY2 Evaluation Report included a particularly clear and comprehensive discussion of potential sources of bias in the evaluation. In contrast, the EY4 Evaluation Report made effectively zero reference to potential sources of bias. Itron recommends that a discussion of bias and how it is to be minimized be included in each year's Evaluation Report.

### *3.4.10 Adequacy of Sampling Plan and Final Sample*

The only evaluation activity in the EY4 evaluation that directly involved sampling was the selection of program and non-program participating retail stores for shelf surveys.

The Evaluation Team slected the six retail chains that sold the most program bulbs. According to the Evaluation Report, the team "*selected a sample of stores to span the EmPOWER Maryland utilities' service territories, visiting 19 stores in April 2013 and another 27 stores in September 2013. In the study's first round in April 2013, 12 of the 19 stores had participated in the EmPOWER Residential Lighting programs, while the remaining seven had not. In the second round in September 2013, the team visited 14 participating and 13 nonparticipating stores.*" (Page 30.) "*To obtain a diverse sample that controls for external factors that may influence light bulb availability and price, the team segmented the sample by utility, store location, and median income level of the area residents (characterized as low-, medium-, or high-income areas).*" (Page 35.) *"Low-income counties had a median household income below $55,000; medium-income counties had a median household income between $55,000 and $85,000; and high-income counties had a median household income above $85,000."* (Page 36.)

*"Because so many retailers are involved with the utilities' programs, the pool of non-participating retailers was small, and the Evaluation Team was not able to retain the same distribution channels"* (page 38) in the participating and non-participating store groups. *"No sales data were available to guide the selection of non-participating retailers that also sell efficient lighting products."* (Page 38.)

*"The team visited between one and five stores each from non-participating retail chains that sell efficient lighting products. Four non-participating retail chains were visited in each round of the study."* (Page 38.)

The selection of participating and non-participating retail stores affects the results of the process evaluation in EY4 but does not directly affect the impact evaluation parameter values. This is because the price response model used to derive NTG was based on data only from participating stores. Note, however, that the Evaluation Team did preliminary analysis on a price effect of program involvement that goes beyond the incentive amount on program bulbs. While the Evaluation Team did not incorporate this effect into the overall NTG calculation for EY4 due to some data collinearity issues, they note that they may do so in future evaluation years as the data collection process is refined. Itron underlines the importance of finding parallel participating and non-participating stores in the same retail channels if this effect is to be estimated.

### *3.4.11 Responsiveness to EY3 Recommendations*

In the EY3 Verification Report, Itron offered numerous recommendations pertaining to the Residential Lighting programs evaluation. Recommendations from the EY3 Verification Report are enumerated below with comments about the Evaluation Team's responses in the EY4 evaluation:

1. **Recommendation:** Consider making a formal request that the utilities provide an estimate of the cost to track lumen output as a standard variable in the program tracking data. Consider phasing in this reporting requirement over time or for selected utilities. We recommend the topic should be discussed internally with the Program Administrators and then the Evaluation Team should provide their assessment of the pros and cons of this change at the next evaluation meeting.

   *Evaluation Team Response*: The Evaluation Team recommended in the EY4 Evaluation Report that the utilities include a lumens field in their tracking data. Itron continues to recommend that this topic be addressed with Program Administrators in the context of the ongoing evaluation meetings. The Evaluation Team has indicated that they will follow this recommendation on a forward-looking basis.

2. **Recommendation:** For discussion of error associated with HOU and CF calculations, include full details on the sample selection process for homes and for bulbs within homes, as well as the data cleaning process, the number of loggers and/or data points excluded from analysis as a function of data cleaning, and the justification for exclusion.

   *Evaluation Team Response*: No new logger data were collected in EY4 for these values. Errors associated with sampling conducted in 2010 and 2011 were included in overall standard error calculations in the EY4 report.

3. **Recommendation:** Use three house prototypes rather than one in the BEopt energy simulation model used for lighting-HVAC interactive effects estimation. Using three prototypes would allow for a more accurate representation of the diversity of single-family, multi-family, and mobile homes in Maryland. Weighted averages for the specific features of these representative homes could be derived based on available data and could inform the relative weight of each of the three house types in the overall factor estimates.

   *Evaluation Team Response*: The Evaluation Team used weighted averages for a variety of house characteristics to inform the building energy simulation modeling and drew upon multiple secondary data sources for this purpose. As such, the diversity of homes in each utility service territory was appropriately represented in the model.

4. **Recommendation:** Add the next level of rigor to the estimate of HVAC system efficiency in the lighting-HVAC interactive effects model by collecting primary data specifically on this factor. Because the estimation of WHFd is highly sensitive to system efficiency, which is, in turn, highly sensitive to the high outdoor air temperatures and humidity levels likely seen at time of peak summer load, primary data collection should be undertaken to specifically identify Coefficient of Performance during the PJM peak and utility peak hours as a function of temperature and humidity.

   *Evaluation Team Response*: This is an expensive and difficult recommendation to carry out. The Evaluation Team reflected vintage of HVAC systems in generating their estimates of system efficiency in the model but did not carry out primary data collection for this purpose. Itron recommends that information on HVAC system vintage and efficiency be collected on a low cost basis as part of the residential light metering study that is currently underway and in all future data collection efforts that involve residential onsite visits. The Evaluation Team has indicated that these data will be collected and incorporated into EY6 WHF modeling.

5. **Recommendation:** Perform a sensitivity analysis on the major inputs to lighting-HVAC interactive effects estimation as part of the evaluation. Lighting-HVAC interactive effects are a significant component of the overall savings calculation on the demand side. Sensitivity analysis should be undertaken to frame the range of impacts associated with reasonable variability across a number of key model inputs.

   *Evaluation Team Response*: Variations in input values in the current model structure provide a kind of de facto sensitivity analysis in the model output workbook. Itron recommends that this sensitivity be further explored, displayed, and discussed in the Evaluation Report itself.

6. **Recommendation:** Provide a confidence/precision interval around the estimate of the proportion of residential bulbs being installed in non-residential sockets. Because there is

sampling associated with the in-store intercepts in the form of selecting certain program participating stores, a sampling error term should be calculated for this parameter. This parameter should also be included in the Monte Carlo simulation to generate an overall estimate of the confidence/precision interval for energy and demand savings as a function of sampling error.

*Evaluation Team Response*: This was not done. The Evaluation team's de facto response was to plan a sticker-based survey to increase the rigor of this estimate looking forward, but to treat the proportion of bulbs going into non-residential sockets as an assumption rather than a sampling-based value in the EY4 evaluation. Itron supports this approach so long as secondary studies are cited that support this assumption alongside the intercepts data that was collected in EY2.

7. **Recommendation:** Repeated from 2011: "[Determine whether for future program years] savings from installation of Residential Lighting program bulbs in commercial sockets should be attributed to the Residential Lighting programs. If so, make a point of collecting information on the type of business in which the bulbs are expected to be installed so that HOU and CF differentiated by business type can be applied to the resulting data."

   *Evaluation Team Response*: This was not done. While cross-sector installation of program bulbs in non-residential sockets became a component of the formal evaluation-based savings calculation starting in EY3, Itron is not aware of a formal discussion having taken place with Program Administrators about inclusion of this factor. Thus, the standing recommendation made here is repeated in the current Verification Report for the third year in a row.

8. **Recommendation:** Include a discussion of possible sources of bias in the savings estimation process. Itron notes that an excellent discussion was included on this topic in the PY2011 Evaluation Report. That discussion provides essential context setting the relevance and scope of the standard error estimation and should not be overlooked. Itron recommends that a concise summary of possible sources of bias be included in each year's Evaluation Report.

   *Evaluation Team Response*: This was not done. Bias is effectively not mentioned in the Evaluation Report.

## 3.5 Recommendations

The recommendations below are targeted at further improving certain aspects of the evaluation next year and in future evaluation years.

- Total program bulb sales are not listed in the Evaluation Report, either statewide or by utility. Itron recommends that a table be added to the Evaluation Report that shows the final, weighted average values of all savings parameters used in the evaluation. This should include program bulb sales by utility, the savings parameter values applied to program bulbs for each utility, and the resulting gross and net energy and demand savings values, such that a reader can observe and recalculate total evaluated savings.
- Utilities and implementers should share scanned copies of bulb invoices with the evaluator and that these be used by the Evaluation Team to independently verify the accuracy of the program tracking data.
- If the Evaluation Team pursues a means of calculating DWM other than what is in the TRM, the basis for this difference be included in the Evaluation Report, as well as the lumen bins developed by the Evaluation Team for each bulb type.
- The approach to communicating about the difference in wattage between program and baseline bulbs should be shifted from a DWM approach to a more intuitive and simple delta watts approach (i.e., communicating the average difference in wattage between program and baseline bulbs as a simple weighted average number of watts, rather than a multiplicative ratio relative to the efficient wattage). If the DWM is kept in place, the Evaluation Team should include the weighted average program bulb wattage for each utility in the Evaluation Report, as this is necessary for re-creating the calculations of energy and demand savings from the constituent parameter values.
- The utilities should show all parameter values used in reported energy and demand savings calculations in the program tracking data.
- The Evaluation Team should re-create strict TRM-based savings calculations as a check on utility-reported savings and highlight any differences between utility-reported savings and strict TRM-based savings in the Evaluation Report. The Evaluation Team has indicated that this recommendation will be followed in EY6.
- The Evaluation Team should consider including a savings summary table in the Evaluation Report that shows separate parameter values and savings subtotals for bulbs installed in residential and non-residential sockets, and add these values to show overall program energy and demand savings. This is more transparent and intuitive than generating weighted average values for HOU, WHFe, WHFd, and CF across residential and non-residential installations and applying these to all bulbs.
- Looking forward, if the Evaluation Team is not able to generate an updated estimate of the proportion of program bulbs going into non-residential sockets using a web-based survey, intercept surveys should be conducted to provide an updated value for this parameter or that another reliable and current data source be found. The value derived in EY2 needs to be updated.

- The Evaluation Report should include a sensitivity analysis associated with analytical decisions that were made in developing the price response model. For example, the Evaluation Team decided to exclude data from the model that showed counterintuitive average price elasticity for a given model in a given store (i.e., purchase rates rising when prices rise). Performing a sensitivity analysis on the impact of these analytical decisions would increase the transparency of the evaluation. The Evaluation Team has indicated that this recommendation will be followed in EY6 when Price response modeling will be revisited.
- The Evaluation Team should consider including data from a multi-year period when normalizing the price response regression model for seasonal effects on bulb sales rates. The EY4 regression model uses seasonal sales patterns from EY3 for normalizing. Data taken from a longer period would provide a smoother and more reliable indicator of seasonal sales patterns.
- On an ongoing basis, the Evaluation Team should make a strong effort to minimize any potential systematic bias in the price response model, and be aware of and articulate any systematic bias that remains and cannot be eliminated from the model. This would include making sure program stores and non-program stores whose data are included in the analysis are in the same retail channels to the greatest extent possible. The Evaluation Team should seek sales volume by retail channel and by product for both program and non-program stores.
- In the price response model, the Evaluation Team should make a greater effort to be explicit about the assumptions being made regarding causality between price changes and sales rate changes. Differences in sales rates cannot necessarily be fully attributed to changes in price.
- While variations in input values in the current lighting-HVAC interactive effects model structure provide a kind of de facto sensitivity analysis in the model output workbook, this sensitivity should be displayed and discussed in the Evaluation Report itself.
- The Evaluation Team should support the assumption that 5% of program bulbs are installed in non-residential sockets with additional studies from the literature.
- [Repeated from 2011 and 2012 Verification Reports] While cross-sector installation of program bulbs in non-residential sockets became a component of the formal evaluation-based savings calculation starting in EY3, Itron is not aware of a formal discussion having taken place with Program Administrators about inclusion of this factor. Pursue an administrative decision for future program years as to whether savings from installation of Residential Lighting program bulbs in commercial sockets should be attributed to the Residential Lighting programs.
- [Repeated from 2012 Verification Report]: The Evaluation Team should include discussion of possible sources of bias and steps taken to minimize bias in the savings

estimation process. Itron notes that an excellent discussion was included on this topic in the PY2011 Evaluation Report. That discussion provides essential context setting the relevance and scope of the standard error estimation and should not be overlooked.

# 4

# Residential Appliance Rebate Programs

## 4.1 Verification Summary

Based on review of the Appliance Rebate Program Evaluation Report for EY4 (June 1, 2012 through May 31, 2013),[1] Itron accepts the RRs and EY savings estimates for the Residential Appliance Rebate programs in full and recommends no adjustments to the savings estimates.

Table 4-1 reports the verified EY4 gross and net savings and RR estimates for the Residential Appliance Rebate programs.

### Table 4-1: Summary of Appliance Rebate Programs Verified Gross and Net Savings for EY4

| | Verified Gross Impacts | | | | Verified Net Impacts | | | |
|---|---|---|---|---|---|---|---|---|
| | EY Savings | | EY RR | | EY Savings | | EY NTG | |
| Utility | MWh | kW* | MWh | kW* | MWh | kW* | MWh | kW* |
| BGE | 8,131 | 1,174 | 0.96 | 0.95 | 2,846 | 411 | 0.35 | 0.35 |
| PEPCO | 3,053 | 416 | 1.01 | 1.00 | 1,062 | 146 | 0.35 | 0.35 |
| DPL | 691 | 101 | 1.01 | 0.95 | 242 | 35 | 0.35 | 0.35 |
| PE | 1,212 | 171 | 0.89 | 0.93 | 424 | 60 | 0.35 | 0.35 |
| SMECO | 1,363 | 183 | 1.11 | 1.05 | 504 | 68 | 0.37 | 0.37 |
| **Total** | **14,432** | **2,045** | **0.98** | **0.97** | **5,078** | **720** | **0.35** | **0.35** |

* Utility Coincident Peak Savings

The approach and methods Itron used to reach these verified savings numbers are provided in the balance of this section along with recommendations to improve the accuracy and usefulness of savings estimates in the next evaluation for these programs.

The key recommendations are summarized below.

- 2014 free rider batteries should include questions about the timing of purchases to be used in life cycle savings estimates; this is an issue for all of the programs.

[1] Navigant and Cadmus, *EmPOWER Maryland Final Evaluation Report Evaluation Year 4 (June 1, 2012 – May 31, 2013): Appliance Rebate Program*, March 21, 2014—Revised June 23, 2014.

- Itron supports the Evaluation Team recommendation to discontinue incentives for electric water heaters.

## 4.2 Program Summary

The Appliance Rebate programs offer incentives to customers for purchase of high efficiency (ENERGY STAR- and Consortium for Energy Efficiency-designated) appliances. The objective is to induce customers to purchase the high efficiency versions when purchasing appliances. All five utilities offered incentives for refrigerators, freezers, clothes washers, dehumidifiers, room air conditioners, and heat pump water heaters. All except BGE offered incentives for electric water heaters. The Residential Appliance Rebate programs contributed 2% of evaluated CY EmPOWER MWh savings and 2% of demand savings.

## 4.3 Evaluation Summary

The Appliance Rebate programs were subjected to relatively low evaluation rigor in 2013. Verification and NTG surveys were performed in 2012 so these were not justified in 2013, especially given the relatively small size of the programs. Additionally, evaluation and verification of the 2012 programs revealed few surprises.

The evaluation for 2013 was limited to engineering review calculations of savings using values from ENERGY STAR and the TRM. The 2013 evaluation did not include any primary data collection and no phone verification of purchases was conducted as part of the PY2012 evaluation.

NTG values were derived from self-report surveys conducted as in the 2012 evaluation. NTG values are based only on free ridership; non-participant spillover was outside the scope of the 2012 and 2013 evaluations. All program units were categorized as Tier 1 or Tier 2 and the corresponding NTG ratios from the 2012 evaluation were applied. This gives a different overall program NTG for 2013 than 2012.

## 4.4 Verification Approach

The verification of the EY4 evaluation of the Appliance Rebate programs included review of the report methods discussion, data, and results. Data showing the savings calculations were provided and spot checked for errors; a few minor discrepancies were identified and resolved with the Evaluation Team. Consistency with TRM prescribed values and algorithms was also confirmed.

### *4.4.1 Overall Findings*

Based on review of the 2013 Appliance Rebate Evaluation Report, Itron accepts the RRs and savings estimates for the Residential Appliance Rebate programs in full and recommends no adjustments to the savings estimates.

The report was complete, clear and well-organized. It was generally self-contained and included the necessary documentation to allow Itron staff to trace through the various calculations and understand the assumptions that were used. Most of the assumptions, algorithms, and relevant program data (e.g., participants, units, etc.) that were used to calculate overall program savings were provided for each measure in response to a data request. We were thus able to avoid iterative requests for data to confirm the evaluation results. Only a few minor issues were identified and these were discussed and resolved with the Evaluation Team.

In the 2012 Verification Report, Itron provided several recommendations to the Evaluation Team. These recommendations were discussed or addressed in the 2013 Evaluation Report. Only a few issues warranted further consideration and are discussed below.

Some Itron recommendations from 2012 will still need to be addressed as part of more rigorous evaluations proposed for 2014, but the Evaluation Team was mostly responsive to recommendations in the 2012 Verification Report insofar as they were applicable given the abbreviated 2013 evaluation. Specifically, the Evaluation Team reported ex post evaluated net savings alongside reported net savings. This made it much easier to see the combined effect of the evaluated NTG ratios and gross RR adjustments to the overall savings results.

## 4.5 Issues and Recommendations

The Evaluation Team addressed or clarified a number of issues identified by Itron in a draft version of the Evaluation Report provided in January 2013. Only four sets of issues (discussed below) merit further attention by the utilities and their Program Administrators. None of these issues warranted an adjustment in the gross or net MW or MWh savings values.

### *4.5.1 Binning Appliances into Tiers*

Categorizing of appliances into Tiers was done differently for purposes of estimating gross and net savings. For gross savings, the Evaluation Team estimated gross RRs based on the program tracking system tier classifications. As stated on page 5 of the 2013 Evaluation Report:

> *In assigning gross savings to different measures and tiers (e.g. the results reported in Section 2), the team used the utilities' tier assignments. The team chose this approach to*

> *preserve like comparisons for gross realized savings ratios in the tables of Section 2.[2] To preserve the interpretability of the gross realized savings ratios, the team retained the tiers assigned by the utilities for each rebated unit, and applied ex post savings to these units. By contrast, the measure-level tables in Appendix E (starting on page 31) illustrate the difference in tier assignments between the utilities and the evaluation team.*

The tracking systems for all utilities except PE classified appliances incorrectly; the tracking systems classified more than 80% as Tier 2, while the Evaluation Team classified more than 80% as Tier 3. The Evaluation Team used the TRM v.3.0 supplemented by the ENERGY STAR database to reclassify appliances into the correct tiers. Itron discussed the reason for the binning error with the Implementation Contractor, who reported that when the tracking system was set up there were only Tier 1 and 2 measures so it did not collect data on Tier 3 energy savings. The problem has been corrected and now includes fields for Tier 3 measures.

For purposes of the NTG analysis, the Evaluation Team classified all measures as either Tier 1 or Tier 2. Any Tier 2 or 3 measures, or measures that were not included in the 2012 NTG survey were binned into Tier 2. The NTG ratios for Tier 1 and Tier 2 for the 2012 analysis were then applied to the corresponding measures in the 2013 analysis. Itron inquired specifically about the classification of electric water heaters—a measure that the Evaluation Team itself recommends be removed from the program in 2014—as Tier 2 in the 2013 evaluation, thus applying the higher NTG score to this measure. The Evaluation Team response was:

> *At the time there were non-program eligible water heater units being sold in the stores, so the participant would have had a choice to purchase a less efficient unit that wouldn't be eligible for a rebate through the program.*

Itron was not entirely satisfied with this response, but having no alternative to suggest and given the small number of electric water heater savings, did not ask for an adjustment to the savings estimates.

### *4.5.2 Net-to-Gross Analysis*

As discussed in the PY2012 Verification Report, previous free rider surveys have not asked whether the high efficiency units would have been purchased in future years without the rebates. This type of question is asked in the California Evaluation Framework and is needed for life cycle savings calculations. The 2014 free rider surveys should include questions about the timing of purchases to be used in life cycle savings estimates.

---

[2] For example, had the team chosen to split the utilities' ex ante Tier 2 refrigerators into ex post Tier 2 and Tier 3, the realized saving ratios for each tier would have been very large (e.g., Tier 3 divided by 0 because no PE savings) or very small (Tier 2 numerator small because some savings sent to Tier 3).

### *4.5.3 Refrigerator Metering*

In the 2012 evaluation, the Evaluation Team metered refrigerators rebated during PY2011 by BGE. The meters were installed at each site and left in place for one month from mid-August to mid-September to measure energy consumption and other usage and environmental factors that impact energy consumption; in addition to metering electric current, the meters recorded internal refrigerator and freezer cabinet temperatures, ambient temperatures, and the frequency and duration of door openings.

Itron expressed concern that, since the meters were only in place for August, the savings could be biased. That month could have a high number of door openings, greater door ice usage, and higher ambient temperatures relative to other months in the year. The Evaluation Team acknowledged that the evaluation did not control for door openings and through-the-door feature usage (ice) and said the Team would consider winter metering activity in the PY2013 evaluation.

This metering was not included in the 2013 evaluation plan and Itron agreed that, given other evaluation priorities, this was not a priority issue for the 2013 evaluation.

### *4.5.4 Reporting*

In the 2012 Verification Report, Itron recommended that in future evaluation years, ex post evaluated net savings be reported alongside reported net savings so that the reader can easily observe the combined effect of the evaluated NTG ratios and gross RR adjustments to the overall savings results. We also asked the Evaluation Team to report utility gross and net savings results at the measure level. The Evaluation Team did so in the 2013 Evaluation Report.

## 4.6 Recommendations

In sum, the issues discussed above do not warrant revisions to the evaluated kWh or kW savings estimates from the Evaluation Team. The key recommendations are summarized below.

- Itron continues to underscore the need to adjust the baselines used to estimate gross savings when NTG results are conclusively low, since the EmPOWER Maryland goals are based on gross, not net, savings. Failure to do so will overestimate the level of savings for these programs and complicate the process of assessing progress toward EmPOWER Maryland goals.
- 2014 free rider batteries should include questions about the timing of purchases to be used in life cycle savings estimates; this is an issue for all of the programs.
- Itron supports the Evaluation Team recommendation to discontinue incentives for electric water heaters.

# 5

# Residential Appliance Recycling Programs

## 5.1 Verification Summary

Itron reviewed the evaluation approach, data, calculations, and assumptions used by the Evaluation Team to produce estimates of peak and annual energy savings for the Appliance Recycling programs. Based on these savings estimates, the Evaluation Team developed measure- and program-level RRs, equal to the evaluated savings divided by reported savings. These RRs were used by the Evaluation Team to estimate savings for PY2013 (January 1, 2013 thru December 31, 2013).

Based on the review of the EmPOWER Maryland Final Evaluation Report EY4 (June 1, 2012 – May 31, 2013) Appliance Recycling program, received July 8, 2014, along with previous review drafts, Itron accepts the savings and RRs estimates for the Residential Appliance Recycling programs in full and recommends no adjustments to the savings estimates.

Table 5-1 reports the verified gross and net savings and RR estimates for the Residential Appliance Recycling programs.

### Table 5-1: Summary of Verified Gross Savings—Appliance Recycling Programs EY4

| | Verified Gross Impacts | | | | Verified Net Impacts | | | |
|---|---|---|---|---|---|---|---|---|
| | EY Savings | | EY Realization Rate | | EY Savings | | EY Net-to-Gross | |
| Utility | MWh | kW* | MWh | kW* | MWh | kW* | MWh | kW* |
| BGE | 8,134 | 1,567 | 1.34 | 1.73 | 4,799 | 925 | 0.59 | 0.59 |
| PEPCO | 3,055 | 574 | 1.27 | 1.60 | 1,833 | 344 | 0.60 | 0.60 |
| DPL | 670 | 163 | 1.30 | 2.14 | 402 | 98 | 0.60 | 0.60 |
| PE | 3,198 | 684 | 1.39 | 1.25 | 2,303 | 492 | 0.72 | 0.72 |
| SMECO | 1,839 | 351 | 1.12 | 1.43 | 1,085 | 207 | 0.59 | 0.59 |
| **Total** | **16,896** | **3,338** | **1.31** | **1.56** | **10,422** | **2,066** | **0.62** | **0.62** |

* Utility Coincident Peak Savings

Source: Navigant and Cadmus, *EmPOWER Maryland Final Evaluation Report Evaluation Year 4 (June 1, 2012–May 31, 2013): Appliance Recycling Program*, March 21, 2014—Revised July 8, 2014.

## 5.2 Program Summary

Appliance Recycling programs offer incentives to customers for retirement and disposal of old refrigerators, freezers, and room air conditioners. The objective is to induce customers to discard old and inefficient secondary appliances and to ensure the units are not sold in secondary markets or given to neighbors, friends, or family. The Residential Appliance Recycling programs comprised 2% of the CY 2013 EmPOWER Maryland statewide portfolio MWh and 3% of the kW savings.

All five utilities offered incentives for refrigerator and freezer recycling in PY2013; incentives were also offered for room air conditioners, if recycled in conjunction with a refrigerator or freezer.

## 5.3 Evaluation Summary

The Appliance Recycling programs were subjected to relatively low evaluation rigor in 2013 in order to free resources for other priority evaluation concerns. The evaluation for 2013 was initially limited to engineering review of the inputs and savings assumptions in the TRM and multiplication of reported recycled appliances by the TRM values. In response to a request from Itron, the final version of the Evaluation Report is now based on unit energy savings estimates from the 2012 evaluation.

Verification and NTG surveys were performed in 2012 so these were not justified in 2013, especially given the relatively small size of the programs. Also, evaluation and verification of the 2012 programs revealed few surprises. The 2013 evaluation found one major error running across four of the utilities: reported demand savings for room air conditioners were exactly one hundredth of the TRM-specified value. This increased the statewide demand impacts by nearly one third.

The TRM does not include NTG values, so NTG values derived from 151 participant self-report surveys conducted as part of the 2012 evaluation were used. The 2013 evaluation did not include any primary data collection or verification with participants that appliances had been removed.

In response to recommendations by Itron in the 2012 Verification Report, the Evaluation Team also examined the processes and methods used by the Program Implementers to assess and verify the age of recycled units; this is an important parameter used to estimate unit savings. Based on this analysis, the Evaluation Team concluded that the age differences between utilities were real and did not result from differing data collection methodologies.

## 5.4 Verification Approach

The verification of the EY4 Appliance Recycling programs included review of the report methods discussion, data, and results. Itron requested and received supplemental data to support measure-level calculations and assumptions in the Evaluation Report. These were checked for errors and that they were consistent with TRM prescribed values. Calculations used to estimate free ridership were examined as part of the 2012 evaluation.

In the 2012 Verification Report we provided comments and suggestions to the Evaluation Team. Most of these recommendations have been discussed and addressed. Only a few issues warranted further consideration and are discussed below.

## 5.5 Overall Findings

Based on review of the 2013 Evaluation Report, Itron accepts the RRs and savings estimates for the Residential Appliance Recycling programs in full and recommends no adjustments to the savings estimates.

The 2013 Evaluation Report is generally complete, clear and well-organized. However, measure-level calculations were not included in the report and had to be requested. Review of the data received revealed no issues with the calculations. The TRM methods and assumptions were used correctly, but are overly simplistic. As discussed below, future evaluations should use the regression model that the Evaluation Team developed and has used in past years. Also, the cost-effectiveness analysis should include a sensitivity case comparing the benefit/cost results using TRM prescribed unit energy savings with benefit/cost results using average unit energy savings from the 2012 evaluation.

Some Itron recommendations from 2012 will still need to be addressed as part of more rigorous evaluations proposed for 2014, but the Evaluation Team was mostly responsive to recommendations in the PY2012 Verification Report insofar as they were applicable given the abbreviated 2013 evaluation. Specifically, the Evaluation Team reported ex post evaluated net savings alongside reported net savings. This made it much easier to see the combined effect of the evaluated NTG ratios and gross RR adjustments to the overall savings results. The Evaluation Team also worked with utilities and implementers to develop and institute a regime for validating age and other data reported by the appliance recyclers and investigated possible discrepancies between utilities pertaining to the age of recycled units. However, despite previous requests, the Evaluation Report did not contain measure-level information that would enable the reader to easily trace high level calculations.

## 5.6 Issues and Recommendations

This subsection highlights issues and related Itron recommendations.

### *5.6.1 Net to Gross Analysis*

Since NTG analysis will be conducted for 2014, and was not this year, Itron reaffirms recommendations from the 2012 Verification Report pertaining to NTG. First, spillover savings estimates were not included in evaluated net savings because of concerns about double-counting savings from other programs (most of the measures cited by respondents were measures that were eligible for incentives from other programs). To confirm this supposition, future spillover batteries should include questions asking whether the respondent had received incentives from other programs for these measures.

Second, consistent with the emphasis of the PY2012 evaluation on first-year savings, the free rider survey questions did not ask whether the program accelerated the disposal of the refrigerator or freezer. This type of question is asked in the California Evaluation Framework and is needed for life cycle savings calculations. Free rider batteries should include questions about the timing of purchases to be used in life cycle savings estimates; this is an issue for all of the programs.

### *5.6.2 Evaluation of Partial Use Factors*

Itron fully concurs with the proposed reevaluation of partial use factors as part of the 2014 evaluation to account for changes made since the initial program year due to changes in measure mix or program saturation.

### *5.6.3 Adoption of Uniform Method Protocols*

Itron supports adoption by the TRM of methods recommended in the U.S. Department of Energy (DOE)'s Uniform Method Protocols (UMP). The Evaluation Team is encouraged to adopt the UMP guidance for future Appliance Recycling program evaluations. The Maryland EmPOWER evaluations are not beholden to the TRM. While the TRM has primacy as the default source for assumptions and methods, the Evaluation Team has always been encouraged to seek better data sources and methods. Itron recommends that the UMP methods be used for future evaluations, regardless of whether they are required by the TRM.

### *5.6.4 Use of Multivariate Regression Model*

The Evaluation Team initially did not use their previously developed multivariate regression model to estimate the gross unit energy consumption for recycled refrigerators and freezers. This model, based on more than 400 appliances that were metered as part of evaluations in California and Michigan, predicts savings based on equipment characteristics such as age, size,

and configuration. The Evaluation Team has used the model in all of the previous EmPOWER evaluations.

The Evaluation Team initially used TRM prescribed values for refrigerators to estimate savings for refrigerators and freezers. As shown in Table 5-2, the average per-unit savings values for refrigerators and freezers from the 2012 evaluation were significantly higher than the TRM deemed values used by the Evaluation Team for the 2013 evaluation (the evaluations in both years used TRM values for room air conditioners).

### Table 5-2: Unit Savings and Cost-Effectiveness Using TRM versus Regression Model

| Utility | Appliance Type | Per Unit Annual kWh Savings | | 2012 Cost-Effectiveness | |
|---|---|---|---|---|---|
| | | 2013 TRM | 2012 PY Average | 2013 TRM | 2012 PY Average |
| BGE | Refrigerator | 760 | 1126 | 1.09 | 1.33 |
| | Freezer | 760 | 604 | | |
| | Room AC | -- | -- | | |
| PEPCO | Refrigerator | 760 | 947 | 0.78 | 0.88 |
| | Freezer | 760 | 761 | | |
| | Room AC | 145 | 145 | | |
| DPL | Refrigerator | 760 | 947 | 1.86 | 2.1 |
| | Freezer | 760 | 761 | | |
| | Room AC | 145 | 145 | | |
| SMECO | Refrigerator | 760 | 922 | 0.55 | 0.61 |
| | Freezer | 760 | 679 | | |
| | Room AC | -- | -- | | |
| PE | Refrigerator | 760 | 1,111 | 0.24 | 0.32 |
| | Freezer | 760 | 904 | | |
| | Room AC | 145 | 145 | | |

Itron posited that use of the TRM values could cause the total resource cost benefit cost ratios to be about one-quarter lower compared to using the average unit saving values from the 2012 evaluation. The Evaluation Team confirmed this and subsequently re-evaluated the program using PY 2012 average unit energy savings.

The Evaluation Team has agreed to use the regression model parameters again in 2014. Program implementers should prepare to provide the data needed to estimate savings using the model. In 2015, the Evaluation Team should re-estimate the model parameters.

### *5.6.5 Use Regression Model in TRM*

The 2013 Evaluation Report recommends that the TRM be revised to use a regression model; currently, the TRM and the utility tracking systems use an alternative approach that does not take into account the actual operational energy performance of the recycled units. While Itron fully supports use of the model in the 2014 evaluation and its incorporation into the TRM, it will likely be necessary to retain default values in the TRM as well, at least for the coming year. This type of model can only be used and useful if the implementers are providing the requisite input data. In previous years, the lack of data has thwarted use of the model for at least a few of the utilities' Appliance Recycling programs. Cost concerns initially led the Evaluation Team to not use the model in the 2013 evaluation. While the model was ultimately used, this cost concern further underscores the need to retain options.

### *5.6.6 Waste Heat Factors*

Finally, the Mid-Atlantic TRM includes kW reduction multipliers for refrigerators and freezers: a Temperature Adjustment Factor (TAF) set at 1.23 originating from a metering study, and a Load Shape Adjustment Factor set to 1.15. Taken together, these factors increase the estimated kW demand reduction by 41.45% (1.23 X 1.15). This is consistent with the California Database for Energy Efficient Resources (DEER).

However, for the kWh savings, the TRM makes no adjustments for indoor or outdoor conditions. This is inconsistent with the treatment of kW demand described above. If demand is increased by the TAF, the kWh savings should be reduced to reflect the TAF. The California DEER indicates that this adjustment would decrease kWh savings by 11% to 17% for the Southern California climate zones.

Since the TRM is accounting for the benefit of boosting the kW reduction due to the indoor environment, it should also account for negative impacts on kWh savings and include a Waste Heat Factor (WHF). Itron recommends that future versions of the TRM and evaluation should apply temperature adjustment factors to kWh savings for refrigerators and freezers to reflect that those savings will be reduced in winter due to reductions in waste heat.

### *5.6.7 Reporting*

Itron continues to request that measure-level details that will allow readers to easily trace high level calculations be included in Evaluation Reports. Statewide and utility tables like Table 5-3 below were included in the final report per request from Itron. They should be included as a matter of course in future Evaluation Reports.

**Table 5-3: Statewide 2012-2013 Installation Year Ex Ante Reported and Ex Post Evaluated Gross Annual Savings by Measure**

| Metric | Measure Category | Quantity | Ex Post Evaluated Gross Savings per Unit | Ex Post Evaluated Gross Savings | Evaluated Net Savings |
|---|---|---|---|---|---|
| Utility Coincident Peak Demand Savings (kW) | Freezer | 3,271 | 0.114 | 373 | 232 |
| | Refrigerator | 13,566 | 0.114 | 1,547 | 949 |
| | Room AC | 915 | 0.900 | 824 | 515 |
| Annual Energy Savings (MWh) | Freezer | 3,271 | 0.760 | 2,486 | 1,548 |
| | Refrigerator | 13,566 | 0.760 | 10,310 | 6,325 |
| | Room AC | 915 | 0.145 | 133 | 83 |

In addition, the Evaluation Team should provide summary statistics, including age and other characteristics that are used as input parameters for the regression model.

## 5.7 Recommendations

In sum, the issues discussed above do not warrant revisions to the evaluated kWh or kW savings estimates from the Evaluation Team. Itron does, however, offer the following recommendations:

- Survey spillover batteries should include questions asking whether the respondent had received incentives from other programs for these measures.
- Partial use factors should be evaluated as part of the 2014 evaluation.
- The Evaluation Team's appliance recycling regression model should be used in 2014 and program implementers should provide the data needed to run the model.
- The Mid-Atlantic TRM should include the Evaluation Team's regression model as an alternative method for estimating savings. It should be provided alongside the current method for now.
- The TRM kWh savings formula for appliance recycling should include a factor to account for reductions in appliance waste heat.
- The Evaluation Team should include age and other characteristics that are used as input parameters for the regression model in the Evaluation Report to allow readers to easily trace at least high level calculations.

# 6

# Residential HVAC Programs

## 6.1 Verification Summary

Itron reviewed the utility program tracking data, calculation methods and assumptions used to determine the evaluated ex post impact estimates for the Residential HVAC programs for EY4. The evaluated program activity for EY4 spans from June 1, 2012 through May 31, 2013. Itron reviewed the EY4 Draft Evaluation Report provided in January as well as revisions provided in February and March with responses to our earlier comments. The Evaluation Team provided supplemental data and information in mid-February and in early and late March. The verification review identified needed corrections, changes to assumptions, and recommended secondary verification of outlier results. The changes were discussed with the Evaluation Team, and they addressed most of the issues we identified in the April 4 draft version of the Evaluation Report. The evaluated and verified impacts are summarized in Table 6-1.

**Table 6-1: Summary of the EY4 Impacts for the Residential HVAC Programs**

| Utility | Gross Ex Post Impacts | | | | | | Net-to-Gross Ratio | Net Ex Post Impacts | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | Evaluated | | | Verified | | | | Evaluated | | Verified | |
| | MWh | Utility kW* | PJM kW# | MWh | Utility kW* | PJM kW# | | MWh | Utility kW* | MWh | Utility kW* |
| BGE | 6,373 | 2,932 | 2,936 | 6,365 | 2,617 | 2,544 | 0.39 | 2,486 | 1,157 | 2,482 | 1,021 |
| PEPCO | 1,614 | 1,092 | 1,132 | 1,613 | 976 | 988 | 0.40 | 646 | 437 | 646 | 391 |
| DPL | 424 | 179 | 149 | 423 | 169 | 130 | 0.40 | 170 | 72 | 169 | 68 |
| PE | 1,609 | 721 | 371 | 1,575 | 617 | 319 | 0.36 | 579 | 260 | 567 | 222 |
| SMECO | 1,022 | 422 | 403 | 1,022 | 376 | 338 | 0.37 | 378 | 156 | 380 | 140 |
| **TOTALS** | **11,041** | **5,347** | **4,991** | **10,998** | **4,755** | **4,318** | **0.39** | **4,258** | **2,081** | **4,244** | **1,842** |

* Utility Coincident Peak Demand Reduction

\# PJM Coincident Peak Demand Reduction

Source: Navigant and Cadmus, *EmPOWER Maryland 2013 Draft Evaluation Report, Evaluation Year 4 (June 1, 2012 – May 31, 2013), Residential HVAC Program*, April 4, 2014, Tables 3, 4, and 5.

Note: The table columns may not add up precisely due to rounding of individual values.

Overall, the verification effort of the EY4 ex post results found that the verified annual energy savings are within a fraction of a percent of the evaluation ex post results. However, the verified

Utility and the PJM peak demand reduction ex post results were less than the evaluation results by 11.6% and 13.5%, respectively. Table 6-2 summarizes and compares the Evaluation and Verification Teams' GRSRs for EY4.

**Table 6-2: Residential HVAC Programs Realized Savings Ratios for EY4**

| Utility | Gross Realized Savings Ratios | | | | | |
|---|---|---|---|---|---|---|
| | Evaluated | | | Verified | | |
| | MWh | Utility kW | PJM kW | MWh | Utility kW | PJM kW |
| BGE | 0.78 | 0.94 | 1.15 | 0.78 | 0.83 | 0.99 |
| PEPCO | 0.80 | 0.84 | 1.15 | 0.80 | 0.75 | 1.00 |
| DPL | 0.77 | 0.87 | 1.24 | 0.77 | 0.83 | 1.08 |
| PE | 1.47 | 1.80 | 1.11 | 1.44 | 1.54 | 0.96 |
| SMECO | 0.78 | 1.02 | 1.43 | 0.79 | 0.91 | 1.20 |
| **Statewide** | **0.84** | **0.98** | **1.17** | **0.84** | **0.87** | **1.01** |

The Evaluation Team used similar analytical and impact calculation methods for the 2013 evaluation as it did for the 2012 evaluation. The Verification Team followed the same basic calculation methods to arrive at the verified ex post results. However, for several key parameters, the Evaluation Team used averages while the Verification Team used the reported parameters of Seasonal Energy Efficiency Ratio (SEER), Energy Efficiency Ratio (EER), Heating Seasonal Performance Factor (HSPF), etc., for individual units from the utility tracking data. For example, the Evaluation Team used average values for the "kW Saved per Ton" parameter in their calculations to determine the ex post peak demand reduction for the ASHP and CAC measures.[1]

Those averages represent the values recommended in the EY3 Evaluation Report and do not represent the performance ratings of the actual equipment installed during the EY4 evaluation period.[2] Itron used the recorded EER value for each claimed unit to determine the peak demand reduction. Also, some of the differences between the evaluation and the verification results are due to Itron finding, in the utility tracking data, missing parameter values for several records for which the Evaluation Team had zeroed out the impacts due to the missing data. The Verification Findings subsection provides more specific discussions as to the differences for each measure category between the evaluation and verification results.

---

1 *EmPOWER Maryland Draft Eval Report EY4: Residential HVAC Program*, page 18, Table 20.

2 The parameter values summarized in Table 20 of the Draft Evaluation Report include the peak coincident demand factors (CDFs) for either CAC or ASHP equipment. Thus, the BGE Utility Coincident Demand for ASHP equipment listed as 0.119 would be 0.153 without a CDF applied. It is important to recognize this potential pitfall so that CDF values are not applied twice when calculating the peak demand reduction estimates.

## 6.2 Program Summary

The Residential HVAC programs constitute 1% of the evaluated calendar year 2013 EmPOWER Maryland statewide portfolio gross MWh savings and 5% of the statewide peak gross MW reduction. The programs' objective is to induce customers to install more energy-efficient HVAC systems and components than they may have considered without the programs' influence. The EmPOWER utilities offer financial incentives to customers that install high efficiency CAC and ASHP, "tune-up" their existing CAC and heat pump (HP) systems, install ductless mini-split air conditioners and GSHPs, and repair leaking air distribution ducts. In addition, both BGE and PE include the replacement of less efficient furnace fan motors with ECMs as part of specific measures, i.e., BGE with natural gas furnaces replacements and PE with HVAC tune-ups. Among the EmPOWER utilities, only BGE offers incentives for installing high efficiency natural gas furnaces.

The Residential HVAC programs encompass eight measure categories as summarized in Table 6-3. As in previous years, the EmPOWER Residential HVAC programs' dominant measures were the HP and CAC replacements, together accounting for nearly 79% of the MWh savings and 82% of the peak demand reduction. The summary results by utility and measure may be found in the Verification Findings subsection.

**Table 6-3: Percent Measure Distribution of the Verified Ex Post EY4 Results**

| Measure | BGE | | PEPCO | | DPL | | PE | | SMECO | | Statewide | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | MWh | kW | MWh | kW | MWh | kW | MWh | kW | MWh | kW | MWh | kW |
| CAC | 21.2% | 41.7% | 38.4% | 60.0% | 6.5% | 14.8% | 7.0% | 15.3% | 3.2% | 7.4% | 19.4% | 38.3% |
| ASHP | 60.1% | 45.6% | 53.9% | 30.2% | 70.0% | 55.8% | 45.9% | 33.8% | 82.2% | 73.9% | 59.6% | 43.5% |
| Ductless AC | 0.1% | 0.1% | 0.0% | 0.0% | 0.0% | 0.0% | 0.1% | 0.1% | 0.0% | 0.0% | 0.1% | 0.1% |
| Ductless HP | 1.4% | 0.9% | 0.9% | 0.4% | 1.2% | 0.9% | 0.0% | 0.0% | 2.1% | 1.6% | 1.2% | 0.8% |
| GSHP | 9.6% | 11.4% | 3.8% | 3.3% | 18.2% | 23.6% | 3.8% | 4.5% | 11.5% | 16.0% | 8.4% | 9.6% |
| Tune-ups | 0.0% | 0.1% | 1.9% | 4.1% | 4.0% | 4.9% | 43.3% | 46.3% | 1.0% | 1.1% | 6.8% | 7.2% |
| Duct Sealing | 0.1% | 0.3% | 1.1% | 2.0% | 0.0% | 0.0% | 0.0% | 0.0% | 0.0% | 0.0% | 0.2% | 0.6% |
| Furnaces | 7.5% | 0.0% | 0.0% | 0.0% | 0.0% | 0.0% | 0.0% | 0.0% | 0.0% | 0.0% | 4.3% | 0.0% |

## 6.3 Evaluation Summary

As with the prior evaluation efforts, the Evaluation Team conducted engineering reviews of most of the measure categories, and undertook phone verification and NTG surveys with customers across the five EmPOWER utility service areas. The Evaluation Team continued and completed the ASHP metering efforts for the 2012–2013 heating season. For the EY4 analysis, the new measurements were combined with prior metering results. All the measurement results were normalized to Typical Meteorological Year weather data. Using the weather-normalized

measurements, the Evaluation Team determined new heating season Equivalent Full Load Hours (EFLH) values for the ASHP equipment for each utility service area. The new ASHP heating EFLH values for EY4 are about 5% lower than the prior evaluation period.[3] The remaining recommended EFLH values remained unchanged from the previous evaluation.

The Evaluation Team completed a Standard Market Practice Study during the period. The Evaluation Team concluded that the Residential HVAC programs accounted for close to 82% of the high efficiency HVAC equipment sales within the five EmPOWER utility service areas. However, the Evaluation Team indicated that, due to the low percentage of the market covered by the distributor data collected for the study and the uncertainty surrounding ECMs installed with new CAC and ASHP systems, the results from the study were not incorporated into the NTG analysis. In addition, the Evaluation Team indicated that the distributor data may contain some bias since the distributors that participated in the study each partner with at least one of the utility's EmPOWER Residential HVAC programs.

The Evaluation Team conducted telephone surveys of program participants to assess their HVAC purchasing decisions, estimate free-ridership, spillover, and determine the NTG ratios for each utility program. Over a two-year period, the Evaluation Team interviewed 173 participants across the EmPOWER utilities. The Evaluation Team estimated spillover using the results from 101 surveys from the EY4 effort only. The free-ridership and spillover estimates were combined to arrive at each utility's program-level NTG ratio. The resulting NTG ratios are low due to the large percentage of participants who were considered either a partial or a full free-rider based on their survey responses.

The Evaluation Team conducted engineering reviews of the utilities' calculation methodologies and tracking data collected for HVAC tune-ups, CAC and ASHP replacements, ductless mini-split systems, and GSHP. For the HVAC tune-up measures, the Evaluation Team reviewed samples of contractor worksheets and invoices for BGE, PEPCO, and DPL, and tune-up files from PE and SMECO. The Evaluation Team only checked the duct sealing measure results for overall reasonableness.

For the BGE and PE measures that incorporate furnace fan motor replacements with an ECM, the Evaluation Team examined the reasonableness of the claimed energy savings and what additional potential evaluation research may be pursued in the future. The Evaluation Team accepted the annual energy savings but recommended no peak demand reduction credit for these ECM installations. For ECMs that are part of either CAC or ASHP replacements, the Evaluation Team indicates that their impact should already be accounted for in the equipment performance

[3] *EmPOWER Maryland Draft Eval Report EY4: Residential HVAC Program*, page 49, Table 30.

ratings, i.e., SEER and HSPF, and that estimating further impacts may lead to double-counting savings. Itron agrees with the Evaluation Team's ECM assessment.

## 6.4 Verification Approach

The verification of the EY4 Residential HVAC programs entailed a detailed review of the calculation methods, utility tracking data, and evaluation results. There were only a few separate follow-ups required with the Evaluation Team. The Evaluation Team's collaboration and prompt follow-up actions to address the identified issues was greatly appreciated.

Itron replicated the evaluation's initial results using the supplied utility tracking data summary workbook and the supplemental documentation on the SQL database savings calculation procedures and functions. To accomplish this step, Itron incorporated the evaluation's savings calculation methodologies directly into the tracking data summary workbook. This verification step found discrepancies, which the Evaluation Team traced to missing EFLH parameter updates in the calculations performed for the EY4 savings reported in the 2013 Draft Evaluation Report in early January 2014. Once the evaluation results were replicated, Itron used the revised utility tracking data workbook to adjust key parameters and assumptions to determine the final verified savings results.

As part of the review effort, Itron checked the detailed calculations for the revised ASHP heating EFLH, reviewed sample contractor forms for the HVAC tune-up measures, and examined the detailed savings calculations for the HVAC tune-up measures. As part of their follow-up to Itron's comments on the initial Draft Evaluation Report, the Evaluation Team provided a workbook with a random sample of 90 PE HVAC tune-up sites that they used to revise the recommended deemed value for the PE HVAC tune-up measure. Itron closely examined the calculations and reasonableness of the estimated savings for each site. In addition, Itron examined the utility tracking data for instances where furnaces with ECMs were installed at the same time as either a CAC or an ASHP replacement. Lastly, Itron examined the NTG and spillover analysis.

## 6.5 Verification Findings

Itron commented on the 2013 Draft Evaluation Report provided in early January 2014 by the Evaluation Team. The teams discussed the most significant findings through email and during a telephone conversation in early March 2014. The Evaluation Team addressed the most salient issues and incorporated revisions into the Draft 2013 Final Evaluation Report distributed in early April 2014. As is evident in Table 6-1, the overall magnitudes of the evaluation and verification MWh energy savings results are in close agreement. However, there are significant differences between the evaluated and verified peak kW demand reduction results.

The peak demand reduction differences are due primarily to the key parameters used to determine the impacts. The Evaluation Team used average parameter values recommended in EY3, by utility and a single value across all efficiency Tiers, while Itron used the as-reported performance ratings contained in the utility tracking data workbook. The distribution of performance ratings among the CAC and ASHP equipment found for each measure Tier in the utility tracking data yielded significantly different "Peak kW Reduced/Ton" averages from those recommended in the EY3 evaluation that the Evaluation Team used. Both Table 6-4 and Table 6-5 compare the "Peak kW Reduced per Ton" parameter values that the Evaluation Team assumed to determine the peak demand reductions and the values that Itron determined using the utility's recorded CAC and ASHP performance ratings in the tracking data.

**Table 6-4: Comparison of Peak kW Reduced per Ton Parameters for CAC Units**

| Utility* | Central AC Tier 1: ≥14.5 SEER, ≥12 EER | | Central AC Tier 2: ≥15 SEER, ≥12.5 EER | | Central AC Tier 3: ≥16 SEER, ≥13 EER | |
|---|---|---|---|---|---|---|
| | Evaluation** | Verification | Evaluation** | Verification | Evaluation** | Verification |
| BGE | 0.098 | 0.047 | 0.098 | 0.071 | 0.098 | 0.089 |
| PEPCO | 0.100 | 0.052 | 0.100 | 0.075 | 0.100 | 0.095 |
| DPL | 0.095 | 0.047 | 0.095 | 0.072 | 0.095 | 0.099 |
| PE*** | 0.087 | 0.046 | 0.087 | 0.068 | 0.087 | 0.080 |
| SMECO | 0.100 | 0.051 | 0.100 | 0.075 | 0.100 | 0.097 |

* The "kW Reduced per Ton" values summarized in the table include each utility's peak CDF multiplier.

** Source: *EmPOWER Maryland Draft Eval Report EY4: Residential HVAC Program*, page 18, Table 20.

*** PE set the Tier 2 threshold at "≥15 SEER, ≥12 EER" and Tier 3 at "≥16 SEER, ≥12 EER."

The Evaluation Team estimated the peak demand reduction using the assumed average "kW Reduced per Ton" values summarized in Table 6-4 and Table 6-5 multiplied by the utility recorded system rated capacities found in the tracking data:

$$Peak\ Demand\ Reduction = \frac{\Delta kW}{Ton} \times Tons$$

In contrast, Itron estimated the peak demand reduction for each claimed unit found in the utility tracking system data using the recorded rated system capacity and EER performance ratings:

$$Peak\ Demand\ Reduction\ = Tons \times \left(\frac{12}{EER_{Baseline}} - \frac{12}{EER_{Installed}}\right) \times CDF_i$$

where, $EER_{Baseline}$ is 11.18, representing a code baseline 13 SEER unit,
$EER_{Installed}$ is the installed EER in the tracking system data, and

$CDF_i$ is the peak coincident demand factor for either a CAC or ASHP unit for each utility.

Hence, the Evaluation Team's approach used an anticipated average EER difference from the code baseline for each utility and equipment type, and Itron estimated the EER difference from code baseline for each installed unit. The distribution of EER values recorded in the utility tracking data, representing the actual equipment installations for EY4, differs from the assumed distribution of EER values per Tier on which the Evaluation Team based the EY3 recommended "kW Reduction per Ton" averages. This is an expected variance between forecast averages and actual results. To be clear, the values in Table 6-4 and Table 6-5 are inputs to the equation above, which lead to the differences in kW reduction reported in Table 6-1.

### Table 6-5: Comparison of Peak kW Reduced per Ton for ASHP Units

| Utility* | Air-Source HP Tier 1: ≥14.5 SEER, ≥12 EER, ≥8.2 HSPF | | Air-Source HP Tier 2: ≥15 SEER, ≥12.5 EER, ≥8.5 HSPF | | Air-Source HP Tier 3: ≥16 SEER, ≥13 EER, ≥9 HSPF | |
|---|---|---|---|---|---|---|
| | Evaluation** | Verification | Evaluation** | Verification | Evaluation** | Verification |
| BGE | 0.119 | 0.069 | 0.119 | 0.097 | 0.119 | 0.130 |
| PEPCO | 0.122 | 0.071 | 0.122 | 0.102 | 0.122 | 0.136 |
| DPL*** | 0.117 | 0.063 | 0.117 | 0.097 | 0.117 | 0.141 |
| PE**** | 0.108 | 0.057 | 0.108 | 0.092 | 0.108 | 0.105 |
| SMECO | 0.122 | 0.069 | 0.122 | 0.100 | 0.122 | 0.133 |

* The "kW Reduced per Ton" values summarized in the table include each utility's peak CDF multiplier.

** Source: *EmPOWER Maryland Draft Eval Report EY4: Residential HVAC Program*, page 18, Table 20.

*** The Verification Team found that the evaluation's actual parameter values are 0.116 for Tier 1 and 2, and 0.115 for Tier 3.

**** PE set the Tier 2 threshold at "≥15 SEER, ≥12 EER" and Tier 3 at "≥16 SEER, ≥12 EER." The Verification Team found that the evaluation's actual Tier 3 parameter value is 0.105 and not 0.108.

Other measure categories that differ somewhat in peak demand reduction are summarized in Table 6-6. They include GSHP installations and HVAC tune-ups. The natural gas furnace replacement measure did not undergo a detailed engineering review by either the Evaluation Team or Itron. The measure's electric energy savings were only checked for reasonableness. It appears that BGE better scrutinized the measure in EY4 since only 304 furnace replacements out of 1,755 claimed a peak demand reduction along with higher ex ante energy savings of 419 kWh per furnace. The balance of furnace replacements claimed only 241 kWh per furnace with no peak kW demand reduction. However, additional improvements are needed. The assignment of either the high or low annual energy savings do not appear to follow an obvious pattern that considers whether the furnace replacement was paired with either a CAC or an ASHP replacement. Itron found many instances where the higher energy savings and peak demand

reduction claim was paired with either a CAC or an ASHP replacement. Also, the tracking data did not clearly identify what was the primary space heating source for each premise.

**Table 6-6: Summary of Statewide EY4 Ex Post Electric Impacts by Measure**

| Statewide | Evaluated | | | Verified | | |
|---|---|---|---|---|---|---|
| **Measure Categories** | **MWh** | **Utility kW** | **PJM kW** | **MWh** | **Utility kW** | **PJM kW** |
| CAC | 2,137 | 2,110 | 2,370 | 2,137 | 1,824 | 2,042 |
| ASHP | 6,575 | 2,379 | 2,619 | 6,554 | 2,070 | 2,276 |
| Ductless Mini-Split AC | 8 | 3 | - | 8 | 3 | - |
| Ductless Mini-Split HP | 167 | 32 | - | 130 | 35 | - |
| GSHP | 926 | 397 | - | 927 | 456 | - |
| Tune-ups | 729 | 396 | - | 741 | 340 | - |
| Duct Sealing | 24 | 27 | - | 24 | 27 | - |
| Furnaces | 477 | - | - | 477 | - | - |
| **TOTALS** | **11,043** | **5,347** | **4,989** | **10,998** | **4,755** | **4,318** |

Note: The table columns may not add up precisely due to rounding of individual values.

In addition, the natural gas furnace replacement measure description indicates that the furnace fan motor may be replaced with an ECM. When the evaluation and verification review teams discussed the potential impacts of ECMs, both engineering reviewers agreed that there was no peak demand reduction. Hence, the Evaluation Team and Itron zeroed out the claimed peak demand reduction.

According to the utility tracking data, over 82% of all furnace replacements included either a CAC or an ASHP replacement for the same customer. Since the performance ratings of either CAC or ASHP replacement equipment should already account for the ECM improvement, crediting additional savings to the furnace fan ECM risks double-counting the impacts. For example, Itron found 64 instances of furnace replacements that included an ASHP replacement. These 64 records likely double-count the ECM energy savings. However, the verification effort did not adjust the energy savings for these 64 records due to the lack of information on the primary space heating source and the small overall impact they represent. Itron recommends that this measure be thoroughly reviewed in the next evaluation.

The difference in peak demand reduction for the GSHP measure is due to the same reason that the CAC and ASHP measures differ: the evaluation used assumed average values for the "Peak kW Reduced per Ton" parameter and the verification used the performance ratings indicated in the utility tracking data to determine each installation's peak demand reduction. In addition, both the Evaluation Team and Itron agreed that the baseline for the GSHP measure should not be a code minimum ASHP unit in the absence of adequate evidence that existing ASHP equipment was replaced at the customer site. Therefore, the Evaluation Team and Itron agreed that the

baseline for the GSHP measure should reflect the minimum requirements listed in the ENERGY STAR specification for GSHP equipment: 14.1 EER.

The HVAC tune-up measures underwent further scrutiny this year and Itron concentrated specifically on the evaluated deemed savings assigned to the PE HVAC tune-up measures with and without furnace fan motor replacements with ECMs. The evaluation used a random sample of 90 PE HVAC tune-ups to establish an average deemed value for the measure. The utility-provided sample did not include either the nameplate performance ratings of the existing equipment or the estimated age of the equipment. Therefore, the Evaluation Team assumed that the sampled CAC and ASHP equipment performance ratings were 10 SEER and 6.8 HSPF. Itron found that the calculated impacts for seven sites significantly skewed the averages, producing standard deviations that were twice the magnitude of the evaluation's proposed average deemed values. Itron requested that either the estimated results for those seven sites be confirmed by secondary means, i.e., billing analysis, or they should be removed from the sample. Itron removed the seven outliers from the sample to derive an acceptable average deemed value for the PE HVAC tune-up measure.

The Evaluation Team provided two workbooks to Itron as a result of initial comments made on the 2013 Draft Evaluation Report. One workbook contained the revised ASHP heating EFLH estimates for EY4 incorporating the Evaluation Team's completed ASHP metering efforts. Itron found a data range discrepancy in the ASHP heating EFLH workbook. Once the data range discrepancy was corrected, the revised ASHP heating EFLH decreased by 5.3% compared to the original value that appeared in the initial Draft Evaluation Report. This decrease in ASHP heating EFLH contributed to the overall decrease in both the evaluated and verified heat pump energy savings.

The second workbook the Evaluation Team provided contained some example contractor forms along with sample savings estimates used to determine the deemed values for each utility HVAC tune-up measure. Itron reviewed the sample forms and the information they provided to inform the individual site savings estimates. Itron observed that neither the equipment's recorded nameplate performance ratings nor the indicated equipment age was taken into account when nameplate data was not available. The Evaluation Team uniformly assumed a 10 SEER and 6.8 HSPF for all equipment. Also, one technician form clearly indicated that the unit was an air conditioner, yet the savings estimate treated it as an HP. Based on these observations, the Evaluation Team undertook a closer review of the data translation from the contractor forms and made corrections, in particular for the PE sample sites.

Table 6-7 through Table 6-11 summarize the evaluated and verified Residential HVAC measure results for each utility.

### Table 6-7: Summary of BGE Ex Post Electric Impacts by Measure

| BGE | Evaluated | | | Verified | | |
|---|---|---|---|---|---|---|
| **Measure Categories** | **MWh** | **Utility kW** | **PJM kW** | **MWh** | **Utility kW** | **PJM kW** |
| CAC | 1,347 | 1,273 | 1,428 | 1,347 | 1,091 | 1,223 |
| ASHP | 3,832 | 1,370 | 1,508 | 3,825 | 1,193 | 1,321 |
| Ductless Mini-Split AC | 7 | 3 | - | 7 | 3 | - |
| Ductless Mini-Split HP | 92 | 18 | - | 90 | 24 | - |
| GSHP | 611 | 259 | - | 611 | 297 | - |
| Tune-ups | 2 | 2 | - | 2 | 2 | - |
| Duct Sealing | 6 | 7 | - | 6 | 7 | - |
| Furnaces | 477 | - | - | 477 | - | - |
| **TOTALS** | **6,374** | **2,932** | **2,936** | **6,365** | **2,617** | **2,544** |

Note: The table columns may not add up precisely due to rounding of individual values.

### Table 6-8: Summary of PEPCO Ex Post Impacts by Measure

| PEPCO | Evaluated | | | Verified | | |
|---|---|---|---|---|---|---|
| **Measure Categories** | **MWh** | **Utility kW** | **PJM kW** | **MWh** | **Utility kW** | **PJM kW** |
| CAC | 619 | 665 | 758 | 619 | 585 | 663 |
| ASHP | 872 | 337 | 373 | 870 | 295 | 325 |
| Ductless Mini-Split AC | - | - | - | - | - | - |
| Ductless Mini-Split HP | 14 | 3 | - | 14 | 4 | - |
| GSHP | 60 | 26 | - | 61 | 32 | - |
| Tune-ups | 31 | 40 | - | 31 | 40 | - |
| Duct Sealing | 18 | 20 | - | 18 | 20 | - |
| Furnaces | - | - | - | - | - | - |
| **TOTALS** | **1,614** | **1,091** | **1,131** | **1,613** | **976** | **988** |

Note: The table columns may not add up precisely due to rounding of individual values.

### Table 6-9: Summary of DPL Ex Post Impacts by Measure

| DPL | Evaluated | | | Verified | | |
|---|---|---|---|---|---|---|
| **Measure Categories** | **MWh** | **Utility kW** | **PJM kW** | **MWh** | **Utility kW** | **PJM kW** |
| CAC | 28 | 28 | 31 | 28 | 25 | 28 |
| ASHP | 297 | 107 | 118 | 296 | 95 | 102 |
| Ductless Mini-Split AC | - | - | - | - | - | - |
| Ductless Mini-Split HP | 5 | 1 | - | 5 | 1 | - |
| GSHP | 77 | 35 | - | 77 | 40 | - |
| Tune-ups | 17 | 8 | - | 17 | 8 | - |
| Duct Sealing | - | - | - | - | - | - |
| Furnaces | - | - | - | - | - | - |
| **TOTALS** | **424** | **179** | **149** | **423** | **169** | **130** |

Note: The table columns may not add up precisely due to rounding of individual values.

### Table 6-10: Summary of PE Ex Post Impacts by Measure Category

| PE | Evaluated | | | Verified | | |
|---|---|---|---|---|---|---|
| **Measure Categories** | **MWh** | **Utility kW** | **PJM kW** | **MWh** | **Utility kW** | **PJM kW** |
| CAC | 110 | 109 | 114 | 110 | 95 | 97 |
| ASHP | 734 | 237 | 257 | 723 | 209 | 222 |
| Ductless Mini-Split AC | 1 | - | - | 1 | - | - |
| Ductless Mini-Split HP | 35 | 6 | - | - | - | - |
| GSHP | 60 | 26 | - | 60 | 27 | - |
| Tune-ups | 669 | 342 | - | 681 | 286 | - |
| Duct Sealing | - | - | - | - | - | - |
| Furnaces | - | - | - | - | - | - |
| **TOTALS** | **1,609** | **720** | **371** | **1,575** | **617** | **319** |

Note: The table columns may not add up precisely due to rounding of individual values.

**Table 6-11: Summary of SMECO Ex Post Impacts by Measure Category**

| SMECO | Evaluated | | | Verified | | |
|---|---|---|---|---|---|---|
| **Measure Categories** | **MWh** | **Utility kW** | **PJM kW** | **MWh** | **Utility kW** | **PJM kW** |
| CAC | 33 | 35 | 39 | 33 | 28 | 32 |
| ASHP | 840 | 328 | 363 | 840 | 278 | 306 |
| Ductless Mini-Split AC | - | - | - | - | - | - |
| Ductless Mini-Split HP | 21 | 4 | - | 21 | 6 | - |
| GSHP | 118 | 51 | - | 118 | 60 | - |
| Tune-ups | 10 | 4 | - | 10 | 4 | - |
| Duct Sealing | - | - | - | - | - | - |
| Furnaces | - | - | - | - | - | - |
| **TOTALS** | **1,022** | **422** | **402** | **1,022** | **376** | **338** |

Note: The table columns may not add up precisely due to rounding of individual values.

The evaluation's NTG ratio determination combined the responses from participants over a two-year period. Thus, the NTG evaluation encompasses program policies, marketing, and messaging that span both EY3 and EY4. This approach improved the overall statistical significance due to the larger pool of surveyed participants, but it does not inform as to whether any program design changes moved the NTG ratio for any of the EmPOWER utilities. In addition, Itron finds that the determination of spillover should include only non-rebated measures that are closely related to space conditioning, e.g., attic and wall insulation, windows, and programmable thermostats; and exclude any measure that received separate financial incentives such as screw-in CFLs that are discounted through upstream incentives. Itron concurs with the Evaluation Team's assessment and decision not to use the results from the Evaluation Team's market practice survey in the NTG analysis due to the distributor's potential bias, the small representation of market sales data, and the uncertainty surrounding the ECM adjustments.

### *6.5.1 Responsiveness to EY3 Issues and Recommendations*

Itron identified several issues in the EY3 Verification Report and offered several recommendations. The issues and recommendations found in the EY3 Verification Report are enumerated below along with the Evaluation Team's follow-up in the EY4 evaluation:

1. **Issue:** In the 2012 Verification Report, Itron concurred with the Evaluation Team that further HVAC Tune-up information be collected to better ascertain the measure impacts.

   *Evaluation Team Response:* Based on the information provided as part of the 2013 evaluation, it appears that progress was made in this area.

2. **Issue:** In both the 2011 and 2012 Verification Reports, Itron recommended that further information about the replaced equipment be captured in the utility tracking data: efficiency, capacity, and age.

*Evaluation Team Response:* This recommendation does not appear to have been implemented in the utility's 2013 program. The EY4 evaluation did not address the need to capture information on replaced equipment.

3. **Issue:** In the 2012 Verification Report, Itron recommended that the free rider survey ask about the timing of purchases and whether the program accelerated the purchase of the incentivized HVAC units.

   *Evaluation Team Response:* These recommendations were incorporated into the telephone survey script.

4. **Issue:** In the 2012 Verification Report, Itron recommended that the ex post net savings should be summarized along with the gross savings in the Evaluation Report.

   *Evaluation Team Response:* This recommendation was implemented.

5. **Issue:** In the 2012 Verification Report, Itron recommended further scrutiny of ECM savings to ensure there savings are not double-counted.

   *Evaluation Team Response:* BGE improved its scrutiny of ECM savings but still claimed peak demand reductions that both the Evaluation Team and Itron zeroed out.

6. **Recommendation:** Utilities should provide information necessary to evaluate tune-up measures, as proposed by the Evaluation Team.

   *Evaluation Team Response:* This recommendation was implemented but is not uniform across the EmPOWER utilities.

7. **Recommendation:** In PY2013, utilities should begin collection of in-situ equipment age and efficiencies to ensure that the program is fully credited for savings from early replacement units and that the varied baseline assumptions are accounted for in cost effectiveness estimates.

   *Evaluation Team Response:* The recommendation was not implemented in the utility's 2013 program.

8. **Recommendation:** The authors of the next version of the Mid-Atlantic TRM should update EFLH and CF values for both HP and CAC, per the findings in the PY2012 evaluation report.

   *Evaluation Team Response:* The recommendation is not addressed in the EY4 Evaluation Report and does not appear to have been implemented in the latest updates to the TRM.

9. **Recommendation:** The PY2013 evaluation should include at least some spot verification to ensure ECM savings are not being double-counted as part of both CAC/HP savings and ECM savings.

   *Evaluation Team Response:* There was no field verification of ECM savings in 2013. The Evaluation Team provided an extensive discussion of evaluation research needs for ECMs. It appears that BGE did scrutinize the measure better in EY4 but additional improvements and examination are needed to ensure there is no double-counting of energy savings.

10. **Recommendation:** If savings from the duct sealing measure exceed 5% of the PY2013 program reported savings, more rigorous evaluation of the measure should be conducted.

    *Evaluation Team Response:* The duct sealing measure remained far below the 5% threshold and therefore a closer examination of the measure was not undertaken. The duct sealing measure may be combined with other measures outside of the Residential HVAC program.

11. **Recommendation:** Future evaluation years should report ex post evaluated net savings alongside reported net savings so that the reader can easily observe the combined effect of the evaluated NTG ratios and gross realization rate adjustments to the overall savings results.

    *Evaluation Team Response:* This recommendation was implemented.

## 6.6 Recommendations

Overall, Itron finds the impact evaluation adequate. The MWh savings are in close agreement once the Evaluation Team incorporated the verification's recommended corrections to the initial draft results presented in January 2014. The peak demand reduction differences are due to the different approach in key parameter values between the evaluation and the verification: the use of assumed averaged values versus the recorded as-is equipment performance ratings in the utility tracking data. Based on the current verification findings and past recommendations that are yet to be implemented, Itron recommends the following changes and improvements:

- The utilities should collect *in situ* equipment age, nameplate efficiencies, and if possible, equipment condition. In past Verification Reports, Itron recommended that the utilities track in their program databases information about the efficiency and capacity of replaced equipment. This recommendation was not implemented for the 2013 program. The lack of adequate evidence on the existing replaced equipment is the most important reason why the GSHP measure baseline was revised for EY4, significantly decreasing the impact results for this measure.

- BGE should remove all peak demand reduction claims for furnace fan motor replacements with ECMs, clarify the conditions for assigning high and low energy savings based on the primary heating source and whether the replacement is paired with either a CAC or an ASHP replacement, and ensure that the ECM energy savings are not double-counted. Although BGE improved their scrutiny of this measure, additional improvements are necessary. Both the evaluation and verification reviews agree that there is likely no peak demand reduction due to an ECM replacing an existing furnace fan motor. When ECMs are installed in conjunction with either CAC or ASHP replacements, it is very likely that the equipment performance ratings already capture the ECM impacts and assigning additional ECM savings entails double-counting the impacts.
- The Evaluation Team should subject the BGE natural gas furnace replacement measure with ECM to a more rigorous engineering review since the measure represents close to 7.5% of the EY4 electricity energy savings results for the utility's Residential HVAC program.
- The EmPOWER utilities should make the HVAC tune-up information data collection and requirements uniform statewide. The two key aspects of this effort are: Use the nameplate ratings of the existing equipment to estimate the savings, and, in lieu of removing the duct sealing measure from the portfolio, integrate all the HVAC service offerings, HVAC tune-ups and duct sealing offerings into a single HVAC diagnostic and repair service measure, as proposed in the California HVAC Maintenance Study.[4] The EmPOWER utilities should follow and adopt the recommendations emerging from the ongoing evaluation research work in California on HVAC Quality Maintenance and Quality Installation practices.
- The PE HVAC tune-up measures should report CAC and ASHP equipment separately.
- The NTG evaluation should use a longitudinal approach to determine whether program design and marketing changes are impacting participant free-ridership and spillover. Additionally, spillover determination should include only measures that do not receive financial incentives.
- Since the evaluated and verified peak demand reduction results exhibited significant differences in EY4 due to the actual distribution of EER values of CAC and ASHP installed systems, the Evaluation Team should use the actual EER values of installed equipment to determine the program's peak demand reduction results in lieu of average "kW Reduced per Ton" values.

[4] Davis Energy Group, Inc. and Western Cooling Efficiency Center, *HVAC Energy Efficiency Maintenance Study*, for Southern California Edison, December 29, 2010, CALMAC Study ID SCE0293.01, pages 60, 63, and 69.

# 7

# Residential Retrofit Programs (QHEC and HPwES)

## 7.1 Verification Summary

Based on review of the EY4 Residential Retrofits Program Evaluation Report,[1] Itron accepts the evaluated RRs and EY savings for the QHEC and HPwES programs produced by the statewide Evaluation Team. Table 7-1 and 7-2 provides a high-level summary of the evaluated and Itron-verified gross and net savings for the QHEC and HPwES programs, respectively, for EY4 (June 1, 2012 through May 31, 2013). Itron was able to replicate most, if not all, of the calculations used by the Evaluation Team to estimate EY4 gross and net savings and, where we identified issues, the Evaluation Team made the necessary adjustments.

### Table 7-1: Summary of Evaluated & Verified Gross Savings–QHEC Programs in EY4

| | Gross Impacts | | | | Net Impacts | | | |
|---|---|---|---|---|---|---|---|---|
| | Evaluated | | Itron-Verified | | Evaluated | | Itron-Verified | |
| Utility | MWh | kW* | MWh | kW* | MWh | kW* | MWh | kW* |
| BGE | 28,559 | 2,818 | 28,559 | 2,818 | 26,560 | 2,621 | 26,560 | 2,621 |
| PEPCO | 22,008 | 2,247 | 22,008 | 2,247 | 20,027 | 2,045 | 20,027 | 2,045 |
| DPL | 6,359 | 615 | 6,359 | 615 | 5,786 | 560 | 5,786 | 560 |
| PE | 5,894 | 612 | 5,894 | 612 | 5,010 | 520 | 5,010 | 520 |
| SMECO | 3,507 | 327 | 3,507 | 327 | 3,296 | 307 | 3,296 | 307 |
| **Total** | **66,326** | **6,620** | **66,236** | **6,619** | **60,680** | **6,053** | **60,680** | **6,053** |

Gross and net evaluated and verified savings are reported for the QHEC programs in Table 7-2.

[1] Navigant and Cadmus, *EmPOWER Maryland 2013 Final Evaluation Report Evaluation Year 4 (June 1, 2012–May 31, 2013): Residential Retrofit Programs*, April 4, 2014—Revised June 23, 2014.

**Table 7-2: Summary of Evaluated & Verified Gross Savings–HPwES Programs in EY4**

| | Gross Impacts | | | | Net Impacts | | | |
|---|---|---|---|---|---|---|---|---|
| | Evaluated | | Itron-Verified | | Evaluated | | Itron-Verified | |
| Utility | MWh | kW* | MWh | kW | MWh | kW | MWh | kW |
| BGE | 1,533 | 357 | 1,533 | 357 | 1,012 | 235 | 1,012 | 235 |
| PEPCO | 1,668 | 413 | 1,668 | 413 | 1,101 | 272 | 1,101 | 272 |
| DPL | 165 | 44 | 165 | 44 | 109 | 29 | 109 | 29 |
| PE | 443 | 82 | 443 | 82 | 305 | 57 | 305 | 57 |
| SMECO | 175 | 36 | 175 | 36 | 110 | 23 | 110 | 23 |
| **Total** | **3,983** | **931** | **3,983** | **931** | **2,637** | **616** | **2,637** | **616** |

Itron accepts the evaluated savings estimates as the best available point estimates and we recommend that the full range of uncertainty found during our review of net program savings be covered by running high and low sensitivity cases during the cost-effectiveness analysis for this program. Nevertheless, we identified a number of issues and offer a number of recommendations for improving the accuracy and value of future evaluations.

## 7.2 Progress on Past Recommendations to Improve Evaluation Methods

Itron provided five recommendations in 2012 to improve the accuracy and effectiveness of future load impact estimates and the effectiveness or performance of future Residential Retrofit program efforts. Below, we describe the progress by the Evaluation Team made toward these recommendations over the last year.

- **Recommendation:** Focus on increasing the accuracy of the gross savings estimates for the HPwES programs operated by PEPCO, DPL, SMECO, and PE by conducting additional billing analysis with a minimum sample size of at least 100 homes or using a census of all participating homes for utilities with fewer than 100 participants in PY2013.

  *Evaluation Team Response:* The Evaluation Team performed enhanced billing analysis for all five of the utility service areas and used a census of all participants for utility programs with fewer than 100 participants.

- **Recommendation:** Conduct an NTG survey analysis for 2013 HPwES programs rather than relying on use of deemed NTG ratios taken from similar programs in other states, as was the case for the first two years of net savings estimates for these programs.

  *Evaluation Team Response:* The Evaluation Team designed and implemented NTG surveys for both single-family and multi-family owners enrolled in this program and

estimated NTG ratios ranging from 0.71 to 0.84 for whole house measures. These estimates represent a significant improvement in accuracy relative to the use of deemed NTG values from other jurisdictions of 0.90 that were used in the 2012 and 2011 evaluations.

- **Recommendation:** Verify customer usage of various control settings on smart strips and which set of appliances are plugged into the smart strip. The goal should be to validate the ex ante savings estimates for smart strips.

  *Evaluation Team Response:* The Evaluation Team performed a more extensive literature review to find alternate sources of savings estimates and compare them to the deemed estimates available in TRM version 3.0. More detailed information on the relative penetrations of home entertainment centers and home offices were used by the Evaluation Team to derive revised deemed savings per smart strip measure. This resulted in significantly lower savings estimates than were used in previous years (savings of 47 kWh per smart strip compared with 102 kWh per smart strip found in versions 1 and 2 of the Mid-Atlantic TRM). The Evaluation Team proposes to update the deemed savings estimates for smart strips in TRM version 4.0 and we support this recommendation.

- **Recommendation:** Focus on process evaluation efforts to estimate conversion or measure installation rates for different types of audit delivery methods, and provide explanations for why different types of customer market segments appear to show higher conversion rates than others.

  *Evaluation Team Response:* The Evaluation Team completed two process evaluations of the HPwES audit programs and produced useful program improvement recommendations. Our review of the Process Report recommendations is included as part of the overall process evaluation review provided under separate cover.

- **Recommendation:** Work closely with PE on the design and implementation of the Small Commercial Energy Savings Kit program, given the experience to date in evaluating savings from the energy conservation kits targeted to single and multi-family homes. The team should explore delivering real-time evaluations every three to six months on how the program is being received in the market place and preliminary estimates or indication of free ridership for key measures such as CFLs.

  *Evaluation Team Response:* The Evaluation Team provided this assistance and provided estimates of savings for the Small Commercial Direct Install program, which can be found in the Commercial Prescriptive and Small Commercial Evaluation Report. To our knowledge, the Evaluation Team did not explore providing real-time evaluation of the either of the Residential Retrofit programs.

Overall, the Evaluation Team has done a commendable job in taking actions to improve the quality and breadth of their evaluation efforts for the Residential Retrofit programs.

## 7.3 Program Summary

The HPwES audit programs are designed to provide customers with a detailed assessment of the available energy efficiency options to reduce home energy use from a whole house perspective, while the QHEC programs provide customers with a more cursory audit assessment and then have the auditor assist the customer to install low cost measures at no cost. The shorter and less comprehensive audit is the QHEC program, which ensures some measurable level of savings is achieved as a result of each home audit and is delivered by leaving behind a few direct install measures at no cost. In 2013, PE took what it learned from running its Residential Energy Conservation program for residential customers and developed a new method of delivering energy conservation kits to small commercial building owners.

The QHEC programs comprised 7% of EmPOWER MWh savings and 5% of kW savings in CY 2013. The HPwES programs comprised 1% of EmPOWER MWh savings and 1% of kW savings in CY 2013.

## 7.4 Evaluation Summary

The EY4 evaluation conducted by the Evaluation Team included review and clean-up of program tracking data; engineering reviews of specific savings estimates for measures distributed in the QHEC programs and the direct install measures distributed as part of the HPwES programs; and telephone surveys to verify installation of measures for all five utilities and to gather information on the customer experience with the programs. In addition, a billing analysis was conducted to estimate savings from the HPwES programs for all five utilities and, for the first time, site verification activities were performed on a random sample of homes participating in the QHEC programs. Estimates of gross savings at the premise level were developed and compared to the savings reported by the Program Administrators.

Estimates of net savings were derived by fielding attribution surveys to four different market segments.

1. QHEC Participant Customers (Single Family) Direct-Install Measures NTG
2. QHEC Multi-family Property Manager Direct-Install Measures NTG
3. HPwES Participant Customer Direct-Install Measures NTG
4. HPwES Participant Customer Whole-House Measures NTG

These surveys were qualitatively different than in previous years because they attempt to estimate spillover savings in addition to adjusting gross savings estimates downward to account for free riders. Addition of the spillover savings resulted in increases in the average weighted NTG ratio for each utility's QHEC program, ranging from 0.84 for PE to 0.93 for SMECO, with a statewide average NTG point estimate of 0.91. This is a significant increase from the NTG

ratio estimated for the same program in 2011 at 0.74, which is probably due to the addition of spillover savings equivalent to 17.5% of direct program savings.

The addition of spillover savings was not as significant for the HPwES programs. The NTG ratios for the HPwES programs ranged from 0.80 to 0.90 for the 2010 to 2013 program years. The Evaluation Team proposes to replace these deemed NTG values with a single point NTG estimate of 0.66 for all HPwES programs operated in EY4 and beyond. We explore the possible reasons for this increase in the NTG ratio for the QHEC programs and drop-in NTG ratio for the HPwES programs in Subsections 7.6.8 and 7.6.9: Net Savings from QHEC and HPwES Programs.

## 7.5 Verification Approach

Itron's verification of the evaluated energy savings and peak demand savings estimates for the EY4 EmPOWER Residential Retrofit programs consisted of the following activities:

- Evaluation Plan Review: We reviewed and made changes to the evaluation plans for these programs, which were prepared in May of 2013. The activities in the final plan were then compared to the activities implemented and described in the Evaluation Report review.
- Evaluation Report Review: We reviewed the Draft Evaluation Report and provided comments to the Evaluation Team. We discussed these comments with the Evaluation Team and suggested changes that should be made to the Final Evaluation Report to either increase accuracy or the clarity of the final product. The Evaluation Team produced a final report that incorporated most of the recommended changes.
- Sample Design: Itron reviewed the sampling plan for each of the sub-programs and confirmed that the sample design and selected customer samples met the confidence and precision levels prescribed in the Evaluation Plan. Itron also explored whether the sample used to conduct phone surveys to determine installation rates was based on a random selection process. No issues were uncovered during this review.
- Engineering Review: Itron reviewed the methods used for estimating energy savings for the range of measures installed by the QHEC and HPwES programs. Our review included assessment of baseline sampling methods, comparison of input values to the recommended values or algorithms in the Mid-Atlantic TRM, and review of the quality of the documentation provided. We only encountered one issue related to how to adjust either gross or net savings estimates when there is evidence collected that confirms baseline usage has changed for a given set of buildings or customers. For example, there is clear evidence that some customers are now replacing existing CFLs with program CFLs, a clear change in the baseline usage for lighting. In our view, this should result in

lower gross savings rather than simply increasing the estimate of free ridership to adjust net savings.

- Review of Billing Analysis: Itron reviewed the process and regressions used to estimate savings per household and confirm they represent best practices for the industry. We found no issues with the methods or estimates.
- Program Savings Comparison: Itron compared the reported ex ante savings in the utility annual reports with those in an early draft of the Evaluation Team's report. This review uncovered some data entry errors and classification uncertainties, which were corrected in the final versions of the Evaluation Team's report.
- Program Savings and GRSR: Itron verified the ratio estimation method used to extrapolate the sample GRSR from the sample of participating customers to the entire population for the QHEC programs. We also independently verified the methods used to estimate the installation rate for program measures using phone surveys or, in limited cases, on-site visits.
- Verified NTG Ratio and Program Net Savings: Itron reviewed the survey instruments used to estimate NTG ratios for each of the target market segments and compared them to best practices in the industry. We then checked the equations used to aggregate and weight individual NTG results by savings to yield an overall NTG ratio for both programs. We discussed our reservations about the approach used with the Evaluation Team and provided our final comments on the method used in Subsection 7.6.8.
- Issues in the Estimation of Gross and Net Savings from these Programs: We identified two significant issues not addressed by the Evaluation Team and developed a recommendation to remedy these issues. The first issue was the choice of how to reflect changing baseline usage levels observed in the field with adjustments to either gross or net savings estimates. The second issue related to the use of open versus closed questions to determine what fraction of the program participants were identified as free riders and to estimate spillover savings benefits for each program.

## 7.6 Verification Findings

In this subsection, we highlight issues identified and/or discussed as part of the 2013 program verification process.

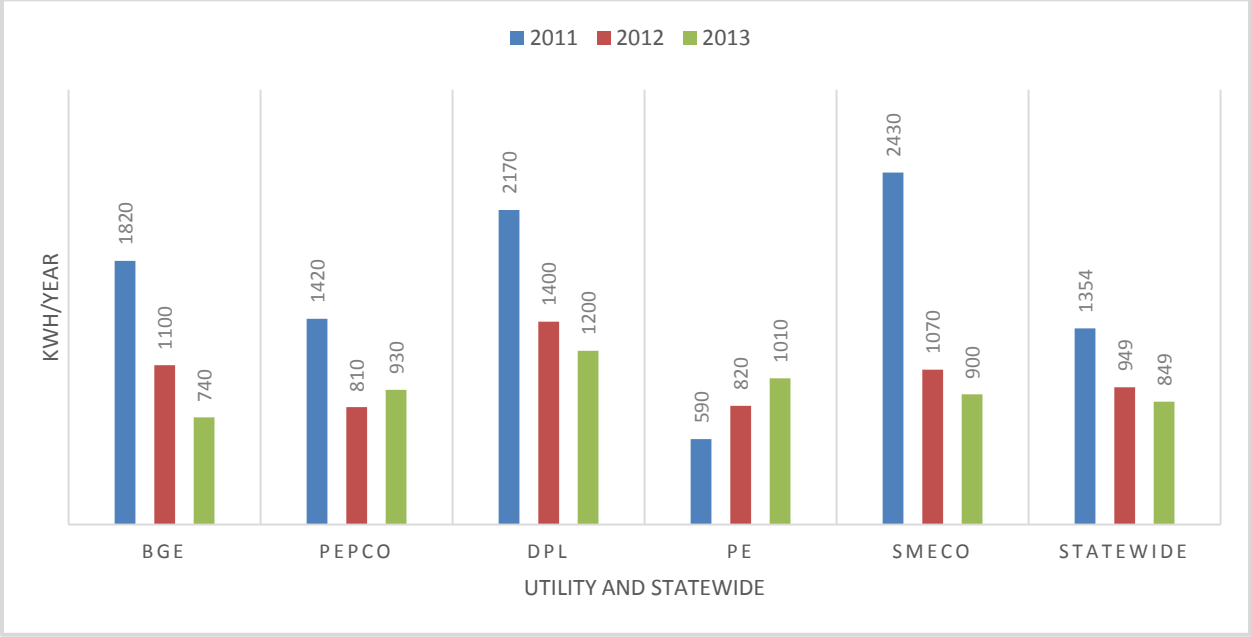
### *7.6.1 Analysis of HPwES Conversion Rates*

The QHEC programs make up well over 90% of the Retrofit programs savings, thus most of the evaluation efforts in 2011 and 2012 have been concentrated on verifying the savings from the QHEC programs. In 2013, the Evaluation Team completed their first comprehensive impact and process evaluations of the HPwES programs, which included analysis of conversion (i.e.,

installation) rates for different types of home energy or web-based audits delivered to customers in 2012 and 2013.

While this analysis provided many useful insights into the HPwES programs, there are still a number of questions pertaining to both the HPwES and QHEC programs that could use further investigation. For example, the energy savings per household for four of the utilities' HPwES programs declined dramatically between 2011 and 2013, as shown in Figure 7-1.[2]

**Figure 7-1: HPwES Average kWh Savings per Participant Household (2011-2013)**




The driving factors behind these decreases in savings could be related to increased rigor in measure savings estimates, extension of the program to customers with lower savings potential (the number of HPwES participants increased by 300% over the last two years), or changes in the mix of measures recommended or offered to customers. The reasons behind these declines in average kWh savings should be explored in greater detail as part of the 2014 evaluation, especially given concerns about the cost effectiveness of the HPwES programs.

Sampling plans and the final sample were found to be adequate to meet the confidence and precision levels that reflect potential sampling error associated with verifying installations of measures in this program. The Evaluation Team targeted 100 QHEC program participants across

[2] These totals include utility-reported savings from direct install measures and any measures installed in response to recommendations in the audit. These estimates do not include potential savings from participants who decided they did not have enough cash to finance certain measures at time of the audit, but may have decided to install these later outside of the program reporting window.

the five utilities and ended up completing phone verification surveys of a total of 105 participants (27 BGE participants, 31 PHI participants, 26 PE participants, and 16 SMECO participants).

The Evaluation Team also performed 50 on-site surveys to verify the installation of free measures (CFLs, low-flow showerheads, smart strips, etc.) installed during the QHEC audit and confirm background data collected related to baseline temperature conditions for water heaters. The sample sizes were sufficient to meet the 90/10 confidence targets when combined with the results from the telephone surveys for the majority of QHEC direct install measures at the statewide level.[3] Itron supports the use of the estimated in-service rates developed by the Evaluation Team for these programs.

The Evaluation Team also estimated the standard error and relative precision of the evaluated energy and peak savings estimates for the QHEC programs based on some Monte Carlo simulations of the factors used to estimate savings for CFLs, which represent the vast majority of savings from these programs. Itron concurs with their estimates of the relative precision of the energy savings estimates as plus or minus 4% for QHEC, plus or minus 3% for the HPwES for the BGE service area and, for the peak demand savings estimates, plus or minus 11% and 6% for the QHEC and HPwES programs, respectively.[4]

### *7.6.2 HPwES Billing Analysis*

The Evaluation Team performed an enhanced billing analysis on all utilities offering HPwES programs and increased the sample size significantly. A census of participants that installed whole house measures were screened to include only those customers who had a minimum of one year of pre and post billing data to verify the savings achieved using billing analysis. This screening step resulted in a loss of 22 participants for Delmarva, SMECO, and PE, reducing the usable set of participant billing records to 89.

The savings estimates for BGE and PEPCO look particularly robust, but the savings estimate for DPL, PE, and SMECO suffer from relatively high levels of relative precision induced by low sample sizes due to low participation levels. These low small sample sizes and resulting low

[3] The Evaluation Team was not able to collect a sufficient sample size to estimate installation rates for temperature turn down, pipe insulation, and tank wrap. In the absence of data, the Evaluation Team assumed an installation rate of 1 for these measures. Itron suggests it would be more reasonable to use an average measure installation rate of 0.80 or even 0 rather than assuming a default verification rate of 1 in the absence of data to verify installation.

[4] For a discussion of the details of these estimates for all MD utilities, see: *EmPOWER Maryland Final Eval Report EY4: Residential Retrofit Programs,* pages 98-108, Appendix F.

precision were, however, unavoidable given that these three utilities all had a low number of participants and there were some significant problems with respect to missing billing data values.

We concur with the Evaluation Team's reasoning that these point estimates are the best estimates available given the circumstances and, as a result, verify these gross savings estimates for the billing portion of the HPwES programs.

### *7.6.3 Adequacy of Data Collection Effort to Verify Measure Installation*

For the on-site verification data collection, the Evaluation Team pulled separate samples of 100 participants from the QHEC programs and had relatively high completion rates. The QHEC program sample received onsite visits to verify measure installations while HPwES measure installations were verified by phone. The results from the QHEC effort are robust (i.e., they meet the confidence and precision targets set in the evaluation plan), but, as might be expected, the verification rates from the telephone survey were lower due to the possibility that respondents did not have the knowledge needed to confirm measure installation or differentiate between measures installed through the HPwES program versus other EmPOWER programs.

Table 7-3 shows the results of the on-site verification work performed for direct install measures by the Evaluation Team for the QHEC programs. These results were primarily derived from site visits to single-family homes where the number of installations in the tracking systems matched quite well with the on-site verification results. These results seem intuitively correct because it makes sense that smart strips would have the lowest installation rate, followed by CFLs and then the other water saving measures, which are much harder to remove once installed.

Some discrepancies were found in the multi-family sector, where the number of measures in the tracking system did not track with the recollection of a few of the property managers about the quantity or quality of measures installed. Itron concurs with the Evaluation Team that these discrepancies in the multi-family segment should be explored in more depth in the next evaluation.

**Table 7-3: Summary of Site Visit Installation Rate Results–QHEC Programs**

| Utility | CFL | Smart Strip | Faucet Aerator | Shower Head | Pipe Wrap | Tank Wrap | Temp Turndown |
|---|---|---|---|---|---|---|---|
| BGE | 0.83 | 0.74 | 0.98 | 0.89 | 0.95 | N/A | N/A |
| PEPCO and DPL | 0.98 | 0.74 | 0.98 | 0.89 | 0.95 | N/A | N/A |
| PE | 0.90 | 0.74 | 0.98 | 0.89 | 0.95 | N/A | N/A |
| SMECO | 1.06 | 0.74 | 0.98 | 0.89 | 0.95 | N/A | N/A |

Source: *EmPOWER Maryland Final Eval Report EY4: Residential Retrofit Programs*, page 57, Table 44.

The Evaluation Team also performed telephone surveys to verify installations of whole house measures installed as part of the HPwES programs. The Evaluation Team was able to confirm by telephone high installation rates (greater than 0.90) for attic insulations, air sealing, and duct improvements, but found relatively low installation rates for HVAC systems (0.58) and water heaters (0.11).[5] Sample sizes for the latter two measures were relatively low—31 sites for HVAC and nine for Direct Hot Water (DHW). We agree with the Evaluation Team that relatively large discrepancies between reported and actual installations of large measures will need to be further investigated in the evaluation cycle.

We note that some evaluators might have chosen to discount or reduce the savings from the HVAC and DHW systems/measures based on these differences in reported versus customer-confirmed measure savings, but we cannot be sure whether these installation rates are accurate, given the low sample sizes. We are also not sure whether the potential errors in reporting installations lie in the tracking system or in the customer self-report surveys for major measure installations. Given the high uncertainty and low level of savings represented by the HVAC and DHW systems, we do not see the need for any discounting at this time. If these uncertainties persist and are identified in the next year's tracking system, we almost certainly will recommend a complete disallowance of savings from these two measures.

### *7.6.4 Adequacy of Phone Survey Instruments*

Itron reviewed the questions used to verify measure installations in the phone survey instruments and found no apparent reasons to think the questions were biased with respect to determining if measures were installed and baseline conditions present. We also determined that the information collected through the phone surveys was sufficient for verifying installation and the resulting energy savings for projects that did not involve site visits.

### *7.6.5 Adequacy of Engineering Reviews for Deemed Savings Measures*

The Evaluation Team conducted engineering reviews for many measures, including: lighting, efficient flow showerheads, faucet aerators, DHW tank wrap, pipe insulation, and smart strips.

The Evaluation Team updated savings estimates for CFLs based on additional metering and research conducted on HOU and WHF as part of the Residential Lighting program evaluation and improved some of the algorithms used to estimate savings from hot water savings devices, such as low-flow showerheads and faucet aerators. The Evaluation Team also improved the savings estimates for smart strips by utilizing more recent data on the penetration of television home entertainment centers and computers.

[5] Source: *EmPOWER Maryland Final Eval Report EY4: Residential Retrofit Programs,* page 64, Table 47.

We found discrepancies in the reported savings for similar CFL bulbs between the QHEC and HPwES programs for the same bulb types for one utility. The Evaluation Team pledged to look into these differences for EY5 but chose to stay with the current higher savings estimates for CFLs installed as part of the HPwES absent a better explanation of the cause of these differences. We note that the Evaluation Team continues to use higher DWM (3.2X on average) than those approved in version 2 of the Mid-Atlantic TRM (2.4 and 1.8 for specialty bulbs), but we support this application because the Evaluation Team's multipliers are based on more recent primary data collected in Maryland.

Itron concludes that the engineering reviews were satisfactory and moving in the right direction over time by collecting higher resolution data on structural variables that have a direct effect on savings estimates, such as the number of occupants per house and the number of lighting sockets.

### *7.6.6 Summary of Accuracy of Gross Energy and Peak Savings Calculations*

The gross energy and peak savings estimates for the QHEC programs are adequate and well-documented. However, no metering has yet been conducted for HPwES over the first three years of evaluations. Currently, the peak savings from these programs are inferred from other evaluations or building energy simulations of end use load shapes. A lack of on-site data collection is not as important for some of the minor savings contributors such as smart strips or water saving devices, but may become more important for estimates of peak savings from whole house measures. We recommend careful consideration of the potential value of performing some limited on-site metering for HPwES participants who have installed major measures and whose impacts should be able to be assessed using a metering approach in the next evaluation cycle. Alternatively, if these users had smart meters installed, it should be much easier to remotely estimate average load shapes for these homes before and after measure installation to determine actual peak demand savings.

### *7.6.7 Energy Savings Persistence Study Not Completed*

The persistence of savings from CFLs is an important topic that the Evaluation Team planned to research in 2012 and again in 2013 but it appears that a lack of budget precluded the ability to complete this study. Savings from CFL installations represent over 80% of the savings from QHEC programs and roughly 70% of savings from all Residential Retrofit programs. Given this fact, Itron recommends that a CFL persistence study be carried out in the next evaluation cycle and be given a higher priority in the budgeting process.

### *7.6.8 Net Savings Analysis of the QHEC Programs*

The Evaluation Team piloted a new approach to estimating both free ridership and spillover using a method that relies on Bass Diffusion Curves and information about the number of measures that had already been installed before the program intervention and participant self-reports about participants' intentions to purchase similar measures over the next 12 months. To illustrate how this method works, we describe the method used just to estimate free ridership for CFLs.

This method required the Evaluation Team to make two potentially arbitrary judgments about:

1. The relationship between the numbers of CFLs installed prior to the program and the likelihood that a specific customer would be a free rider or would have purchased the program CFL in the absence of the program.
2. The relationship between a customer stating that they intend to purchase more CFLs and the likelihood they would have purchased the program measure in the absence of the program.

The Evaluation Team constructed diffusion curves to estimate these changing probabilities of free ridership over time as a function of previous bulbs installed. This was then combined with customer survey responses about their intentions to buy more CFLs in the future.

We were unable to determine how the ends of the CFL penetration scale were set to determine when a customer might become a complete or 100% free rider. As an example of the judgment that needs to be exercised, the Evaluation Team determined that a customer with 14 or more CFLs already installed was a complete free rider (100%), while a customer who had installed nine bulbs was an 80% free rider and a purchase history of three CFL bulbs reduced the chances of being a free rider to 10%. These scores were then integrated with a simple scoring procedure for future purchases: Yes = 100% free rider; Maybe = 50% free rider; and No = 0% free rider; yielding a savings weighted average free ridership score of 41.8% or an NTG ratio of 58.2%.

This exercise requires a considerable amount of judgment to decide how the number of prior CFL installations per household should be directly translated into distinct free ridership probability estimates. Use of this method produced estimates of free ridership that varied from 13.5% for smart strips to 41.1% for CFLs.

This compares to the assumed level of free ridership at the program level that varied from 10% to 20% for the first three years of program operation. As expected, the range of free ridership is larger than the range assumed at the program level. Gathering measure-level data sends important signals to Program Managers about the need to begin to design exit strategies for measures where free ridership is approaching 50%.

On top of the free ridership estimates, the Evaluation Team added their assessment of the likelihood that the Residential Retrofit programs had achieved spillover savings in excess of the reported savings for these bulbs in EY3. Spillover impacts were determined by a single question related to the relative influence of the programs on their intended future decisions to buy additional energy savings measures in the future. Respondents who rated the programs as highly influential, e.g., they ranked the program influence with a score of 8, 9, or 10 on a scale of 1 to 10, were credited with spillover savings equal to the number of measures they planned to install within the next 12 months. Relying on the accuracy of customer-stated intentions to install measures in the next 12 months and then one potentially biased program influence question leads to overall NTG estimates that are highly uncertain at the least.

The bias in the responses to these net survey instruments comes from how specific questions have been phrased and framed with respect to the programs being the only plausible causal agent out of many possible causes. This one-sided framing is likely to bias their responses toward higher levels of program influence. The Evaluation Team agreed with this observation and promised to work on the surveys for the next round of evaluations.

Good practice in attribution research suggests that questions about how and why customers have purchased specific goods should almost always begin as an open ended dialogue, with openers such as: "Are you intending to purchase any more CFLs in the near future? If so, could you describe the factors that were the most important in leading to that decision?" If, and only if, the customer mentions a program influence or element as part of the response, is it reasonable to ask them to rate program influence on their future CFL purchases. Starting out with a scalar question on the level of program influence presumes the programs had an influence and we are only here to find out how strong the influence is. If the respondent is a free rider, this is a completely incorrect assumption. Itron and the Evaluation Team discussed these concerns, and the Evaluation Team agreed to incorporate these recommendations into future research in this area.

Use of this method resulted in a spillover estimate of 17.5% for CFLs which, when combined with the free ridership estimate of 41% (or an NTG ratio of 59%), yielded an overall NTG ratio for CFLs installed as part of the QHEC program of 76.5%. This free ridership estimate is similar to the results found in neighboring states for CFL programs, but the spillover estimate seems low given the cumulative effect of running CFL program for the last five years in Maryland. Spillover results for other measures ranged from 13.5% to 36.8% and we would have expected a priori that the CFL program would register the highest level of spillover savings given the enormous amount of bulbs rebated compared to the overall sales of CFLs in the market place over the same time period.

### *7.6.9 Net Savings Analysis for the HPwES Programs*

The Evaluation Team estimated the net savings from the HPwES programs for the first time this year based on surveys of both single-family and multi-family owners in Maryland. As with the QHEC programs, the Evaluation Team used a bass curve diffusion analysis to estimate net savings from the HPwES programs.

Different surveys were performed for direct install and whole house measures. Free ridership estimates were generally higher for the direct install measures, ranging from 32% for pipe insulation to 55% for CFLs. Free ridership estimates for HPwES direct install measures were higher than the free ridership rates estimates for comparable QHEC measures. Both teams were unsure why these rates were higher for the HPwES program, but the use of the new saturation-based method—where NTG varies as a function of CFLS installed in the household—may be part of the reason.

Spillover savings were estimated as 4.7% of reported savings across all of the direct install measures in the HPwES program. Combining free ridership and spillover resulted in a range of overall NTG ratios for direct install measures from 50% for CFLs to 74% for pipe insulation. Future spillover research should focus on estimating long-term spillover effects over many years, rather than limiting the analysis to effects for up to 12 months after program participation.

Free ridership rates were found to be lower for the larger and more expensive whole house measures. Respondents indicated there was little or no likelihood that the program had stimulated any spillover savings for similar measures after 12 months after the home performance audit was delivered. Table 7-4 illustrates these concepts by showing representative NTG results for BGE. These estimates seem reasonable and are within the range of NTG ratios reported for these types of whole house measures in other jurisdictions.

**Table 7-4: BGE—HPwES Whole House Free Ridership, Spillover, and NTG Summary**

| Measure Type | Free Ridership | Spillover | NTG | Percent of Total Gross Evaluated Population kWh Savings | Whole-House Population Weighted NTG Ratio |
|---|---|---|---|---|---|
| HVAC / HW | 29% | 0% | 0.71 | 3.0% | 81% |
| Air Sealing | 16% | 0% | 0.84 | 35.2% | |
| Weatherization | 20% | 0% | 0.80 | 61.7% | |

Source: *EmPOWER Maryland Final Eval Report EY4: Residential Retrofit Programs*, page 154, Table 128.

NTG results from direct install measures were combined with the results from the whole house measures to derive an overall NTG ratio of 0.66 for the HPwES program. This finding compares

to the use of deemed NTG ratios, which ranged from 0.80 to 0.90 for the five utility programs in the first three years.

Free ridership values based on primary data collection resulted in higher overall NTG ratios for the QHEC programs and lower NTG ratios for the HPwES programs, relative to the deemed NTG values used in the first three years for these programs. On balance, the NTG analysis provided considerable value and new insights into options for improving program designs; however, there is still some work to do to refine the survey instruments for estimating free ridership and spillover impacts. Itron will continue to discuss our critique and suggestions for improvement of the net savings method as part of the EY5 evaluation planning process.

## 7.7 Recommendations

- The Evaluation Team should consider performing at least limited on-site metering for a sample of HPwES participants.
- Low verification rates for HVAC and DHW measures should be investigated in the upcoming evaluation cycle. If these low verification rates persist, savings claims may need to be adjusted.
- The Evaluation Team should ensure that the wording of future NTG survey batteries does not bias survey respondents toward confirmation that the program was the principal factor in their decision to install a measure.
- The Evaluation Team should conduct a CFL persistence study. This should be carried out in the next evaluation year because of the need to develop more accurate estimates of lifecycle savings for CFLs, which represent over 70% of the savings for the Residential Retrofit programs.
- The Evaluation Team should investigate the use of smart meter data to develop estimates of peak savings from the HPwES programs. The average load shape profiles currently being used are not likely to be representative of the participant program and are less accurate than using actual metered data.
- A more intensive evaluation effort should be conducted in EY5 to identify the reasons for declining savings per household and lower conversion rates. At the same time, Program Administrators should be encouraged to pilot test new sales strategies or program designs, given the lack of success in improving the effectiveness of these programs over the last three years.

# 8

# Residential New Construction Programs

## 8.1 Verification Summary

This section presents the results of Itron's effort to verify the savings claims for the RNC programs that were implemented in Maryland during EY4, comprising activity from June 1, 2012 through May 31, 2013. The savings estimates were developed by the Evaluation Team using a different approach than was used for EY3; however, like previous years, the evaluation approach was again constrained by budget and by the fact that the programs represented a small, but growing fraction of the overall portfolio kWh savings,[1] up from 0.04% in 2011 and 0.55% in 2012 to 0.77%[2] in 2013.

The RNC market is showing signs of renewed strength and is poised for a rebound.[3] These programs are vital to the EmPOWER portfolio so as to be prepared to mitigate the impacts of a possible construction boom on the Maryland utilities. The energy usage of homes is firmly linked to the construction characteristics that are established by the standards in effect during the design phase, and these characteristics are unlikely to be modified during the first 15 years after construction. For example, the choice of fuel used for space and water heating is very difficult to change because of the interrelationship between indoor air quality in tight home construction practices promoted by the program, and the need to provide appropriate venting and combustion air intake for natural gas furnaces and water heaters. Therefore, continued vigilance in strength and stringency of the implementation of these programs is vital to set Maryland on a good foundation for continued load growth that has the least environmental impacts and greatest cost-effectiveness.

Table 8-1 presents a high-level comparison of the evaluated savings estimates from the Evaluation Team and verified by Itron for the RNC programs for all of the EmPOWER utilities that implemented such programs in EY4.

---

1 Navigant, *EmPOWER Maryland 2013 Final Evaluation Report: Overview of Findings*, Draft, April 21, 2014.

2 *EmPOWER Maryland Final Impact Eval Report: CY2013*.

3 U.S. Census Bureau, U.S. H.U.D., Housing Starts website, http://www.census.gov/starts

### Table 8-1: Summary of Evaluated and Verified Gross Savings–Residential New Construction for Evaluation Year 4

| | Gross Impacts | | | | | | Net Impacts | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Evaluated | | | Itron-Verified | | | Evaluated | | | Itron-Verified | | |
| Utility | MWh | Utility kW* | PJM kW | MWh | Utility kW** | PJM kW | MWh | Utility kW* | PJM kW | MWh | Utility kW** | PJM kW |
| BGE | 4,567 | 1,641 | 866 | 3,996 | 2,535 | 758 | 3,837 | 1,379 | 728 | 3,357 | 2,129 | 637 |
| PEPCO | 504 | 226 | 92 | 441 | 280 | 81 | 423 | 190 | 77 | 370 | 235 | 68 |
| DPL | 249 | 58 | 60 | 217 | 138 | 52 | 209 | 49 | 50 | 183 | 116 | 44 |
| SMECO | 1,340 | 571 | 305 | 1,172 | 743 | 267 | 1,125 | 480 | 256 | 984 | 624 | 224 |
| PE | 1,093 | 444 | 139 | 957 | 607 | 122 | 919 | 373 | 117 | 804 | 510 | 102 |
| **State-wide** | **7,753** | **2,943** | **1,462** | **6,783** | **4,302** | **1,297** | **6,512** | **2,471** | **1,228** | **5,698** | **3,614** | **1,074** |

* Utility Coincident Peak Savings. Source: Utility Semi-Annual Reports, and Navigant, *EmPOWER Maryland Final Evaluation Report Evaluation Year 4 (June 1, 2012 – May 31),* Revised June 23, 2014, Tables 5 and 6.

** Residential New Construction Measure Coincident Peak Savings. Source: Itron analysis.

Table 8-2 provides the evaluated and verified RRs and NTG ratios for each program. Note that, in accordance with their PY 2013 evaluation plan, the Evaluation Team did not conduct NTG analysis this year; therefore, statewide NTG ratios reported in previous years were used, as shown in Table 8-2. The remainder of this document discusses findings from Itron's review of the gross impact evaluation findings and makes recommendations for improving future implementation and evaluation efforts.

**Table 8-2: Summary of EY4 Gross Realization Rates and Net-to-Gross Findings– Residential New Construction for Evaluation Year 2013**

| Utility | Gross Realization Rates | | | | | | Net-to-Gross Ratios | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Evaluated* | | | Itron-Verified | | | Evaluated* | | | Itron-Verified | | |
| | MWh | Utility kW** | PJM kW | MWh | Utility kW** | PJM kW | MWh | Utility kW** | PJM KW | MWh | Utility kW** | PJM kW |
| BGE | 1.12 | 0.75 | 0.53 | 0.98 | 1.16 | 0.52 | 0.84 | 0.84 | 0.84 | 0.84 | 0.84 | 0.84 |
| PEPCO | 1.12 | 0.75 | 0.43 | 0.98 | 0.93 | 0.42 | 0.84 | 0.84 | 0.84 | 0.84 | 0.84 | 0.84 |
| DPL | 1.12 | 0.75 | 0.78 | 0.98 | 1.76 | 0.76 | 0.84 | 0.84 | 0.84 | 0.84 | 0.84 | 0.84 |
| SMECO | 1.12 | 0.75 | 0.60 | 0.98 | 0.98 | 0.59 | 0.84 | 0.84 | 0.84 | 0.84 | 0.84 | 0.84 |
| PE | 1.12 | 0.75 | 0.28 | 0.98 | 1.02 | 0.27 | 0.84 | 0.84 | 0.84 | 0.84 | 0.84 | 0.84 |
| **Statewide Average** | **1.12** | **0.75** | **0.50** | **0.98** | **1.10** | **0.50** | **0.84** | **0.84** | **0.84** | **0.84** | **0.84** | **0.84** |

* Source: Utility Semi-Annual Reports, and *EmPOWER Maryland Final Evaluation Report EY4*, Revised June 23, 2014, Tables 5 and 6.

** Utility Coincident Peak Demand Savings

Most of the issues identified in the 2012 (EY3) verification issues were addressed in the EY4 evaluation. The limited billing analysis and engineering review in 2012 was eliminated in favor of a re-calculation of the simulation results for a sample of 30 homes. One of the evaluation challenges for 2011 and 2012 was that the Program Administrator refused to provide simulation input files because of concerns about releasing proprietary information. For EY4, the Program Administrator[4] provided "scrubbed" simulation input files and details on the savings calculations within the Beacon Predictive Savings Tool (PST). Furthermore, the utilities provided all available billing usage data for a census of program participants to aid in the analysis of potentially unoccupied homes. The billing usage data was also used to calibrate the simulation models. As in previous years, however, a limited evaluation budget precluded the collection of primary or on-site data and relied on data from the implementers to complete the evaluation. Verified results for kWh savings differ slightly from the evaluated results due to discrepancies found in the calculation of savings, however; the evaluated kW demand reduction impacts are about half of the verified values due to the manner in which the sample results were scaled to the participant population. Itron will address the appropriate formulation of the peak coincident demand factor in a separate memo.

Itron generally supports the more rigorous simulation-based evaluation approach used for the EY4 evaluation. The sampling design and the calibrated savings calculation approach partially address the issue related to differential savings from single-family versus multi-family and

[4] All of the utility companies included in the evaluation plan for PY 2012 contracted with ICF International to administer their RNC programs. ICF is the developer of the Beacon PST.

attached single-family townhomes by ensuring that the sample includes both detached- and attached-style homes. However; there are ongoing issues related to the baseline in the simulation modeling approach for multi-family homes and Itron continues to be concerned over the lack of any primary data collection. The RNC programs represent a small, but growing, fraction of the statewide portfolio. The findings and recommendations in this report are important to consider for future evaluation cycles in order to reduce the uncertainty of the savings estimates and lay the foundation for these programs to realize significant savings as the economy recovers. The survey of builders conducted as part of the ICF RNC Baseline Study[5] found some confusion among builders regarding the more stringent building energy efficiency standards, which were adopted statewide in 2012. The changes suggested in this report are expected to increase the durability and indoor air quality of the homes constructed by these programs and will increase the certainty for homeowners that the promised energy savings will actually materialize. The recommendations below are intended to increase the net value (benefits-costs) of the evaluation effort to Maryland ratepayers. Itron highlights the following recommendations for the next gross impacts evaluation and additional recommendations are discussed in the Recommendations subsection:

1. Primary on-site data collection: Include some form of on-site data collection and/or verification activities in every program evaluation cycle and coordinate with process evaluation efforts to take advantage of the opportunity to review QA/QC procedures.
2. Revise program incentive structure: Adopt an alternate incentive structure which pays incentives based upon annual energy savings instead of HERS scores, which are poorly correlated with per-home energy savings (See Figure 8-3).
3. Continue to update the reference home rule set:[6] The Program Administrator should periodically review and adjust, and/or update the reference home rule set to reflect current ENERGY STAR Reference Home Guidelines, applicable IECC requirements, findings from the Baseline Study, and Federal Appliance Standards. Enforcement of the 2015 NAECA water heater standards are on the horizon and, as discussed, this will have a significant impact on the ability of builders to comply with program requirements.

- Natural gas reference home fuel type: Address the potential for program requirements and reference home rules to de-incentivize electric resistance water heaters when natural gas is available.

Recommendations related to the ICF Baseline Study include:

5 ICF, Draft *2012 IECC Code: Maryland Residential New Construction Baseline Study*, January 2014.

6 This is the rule set that is used to define the configuration of the "reference home" that is used as the baseline for energy efficiency savings calculations. for new construction.

4. ICF Baseline Study: Increase the sample size in the underrepresented jurisdictions and revise the findings of the study to accurately reflect the typical construction characteristics found, even if those findings suggest that any specific home construction characteristic, e.g., windows, are better than the code requirements.
5. REM/*Rate*™ updates: The Program Administrator should engage with the developer of REM/*Rate*[7] to ensure that it includes all appropriate as-built and baseline specifications for the upcoming program cycles for lighting and appliances.

Statewide policy issues that need to be addressed include:

6. Consistent treatment of baselines: The baseline should include the effect of above-code as well as below-code construction characteristics in determining the baseline home construction. For example, currently the ICF Baseline Study ignores its finding of greatly improved window efficiency in the general home construction market.
7. Address treatment of above or below-code compliance: If the Baseline Study finds that non-program (i.e., baseline) homes do not meet current code requirements, should this divergence between standard practice and minimum code requirements be accounted for in the net savings estimates or the gross savings estimates?
8. Peak coincident demand: The peak coincident demand factor should be based on the program savings shape, not on the overall residential load shape. The savings load shape for RNC was shown to be twice as large as the Evaluation Team's estimate using the residential load shape. While the load shape issue was not used in the verified savings, the verified peak demand differs from the evaluated demand because it should be weighted according to ex ante population (not sample) kWh savings in each utility.
9. Verify occupancy: Implementers should collect billing usage data for the participant population to verify occupancy before incentives are paid or adjust savings with an occupancy adjustment factor or adjustment to the effective useful life of the RNC program.

The Evaluation Team consulted with Itron on the evaluation plans for 2014 and learned they are proposing to scale back efforts related to the engineering review of the simulation models. Itron sees this as a reduction in the stringency of the evaluation efforts and can support this change only if balanced with greater stringency in some other aspect of the program evaluation plan. Namely, Itron strongly recommends on-site data collection efforts that focus on improving the quality of HERS rater verification and certification efforts and prepares them for enforcement of the IECC 2012 codes and standards. As such, the on-site data collection activities should be

[7] Software developed by Architectural Energy Corporation: http://www.archenergy.com/products/remrate

more focused on improving the stringency of the HERS rater processes and less focused on documenting each and every light bulb. Instead, the evaluation efforts should document when shortcomings are found in a sample of homes and track what happens to those homes as the issues are hopefully addressed and remedied by the builder.

Another issue that may prove challenging for builders in coming years is the NAECA 2015[8] water heater standards, which are significantly increasing the energy efficiency for gas and electric water heaters with storage capacity between 65 and 120 gallons. These changes will require these larger water heaters to meet the efficiency of a heat pump for electric water heaters and forced draft combustion for natural gas water heaters. The RNC programs are an effective way to push these new technologies into the market by creating early adopters of upcoming changes to the codes and standards. The NAECA 2015 water heater efficiency standards are a great opportunity to increase energy efficiency and transform the local market for more efficient water heaters.

## 8.2 Program Summary

The RNC programs comprised 2% of the CY 2013 EmPOWER Maryland kW savings and 1% of the MWh savings. The 2013 RNC programs provided incentives to builders for constructing single-family housing units that met the energy efficiency thresholds defined by the HERS and which were consistent with the Environmental Protection Agency's ENERGY STAR New Homes program version 3.0. To qualify their homes, builders must incorporate energy-efficient appliances, lighting, building shell features, and high-performance building techniques designed to reduce the energy requirements in homes.

As in previous years, the front lines of construction quality control and program savings were assured through inspections by independent HERS raters who performed field verification and diagnostic testing of a sample of homes to verify construction characteristics and to score the homes' overall annual energy use. The ratings for the tested homes were applied by the HERS rater to similar participating homes constructed by the same builder to determine program participation and incentives. The Program Administrator aggregates the energy ratings for the tested and "sampled" homes (where the results from tested homes are used instead of actual tested results) and uses their proprietary Beacon PST to determine overall program savings. All utility companies implemented their RNC programs during the 2012 and 2013 program years using the same program master implementer (ICF), and the programs' designs were very similar.

[8] For this report, "NAECA 2015" refers to the changes to the National Appliance Energy Conservation Act of 1997, which are going into effect in April 16, 2015.

## 8.3 Evaluation Summary

For EY4, Itron focused its efforts on reviewing the tracking data, sampling method, the simulation modeling of building performance, and the spreadsheet calculation of lighting and appliance savings. For the first time, hourly kW savings were available from the EnergyPro simulations, which Itron used to develop a custom peak coincident load factor based upon the programs' energy savings load shape. This approach was not used to calculate program peak demand because it diverged from statewide practice. Instead, Itron used the utility peak demand values weighted by the distribution of savings between the utilities to calculate the program peak coincident demand impacts. The Evaluation Team's focus was to verify that the Beacon PST savings were reasonable and used the savings simulations and calculations for the 30 sample homes to adjust program electricity savings. The effort to estimate natural gas savings was notably less rigorous than the effort to estimate electricity impacts. While the evaluation only looks at electricity savings, it is not clear if the report contains sufficient information about natural gas savings for the Evaluation Team to complete the upcoming cost-effectiveness analyses. The Evaluation Team effort also included a sensitivity analysis of building orientation requested by Itron. Unfortunately, the effort utilized only the four cardinal orientations of the home, which is insufficient to fully explore the direction and magnitude of this issue.

Given the breadth of information provided and the depth of the analysis conducted by the Evaluation Team, Itron's objectives were a) to determine how well the sample of homes used in the calculation approach represented the participant population, b) to verify the baseline and as-built inputs for the EnergyPro simulations, c) to verify that the calculations of lighting and appliance savings were accurate, and d) to develop a more accurate method of estimating coincident peak demand. Itron's review findings based on these activities are discussed in the following subsection.

Upon receipt of program information from the Evaluation Team, Itron did the following:

1. Reviewed the utility companies' tracking databases for anomalies.
2. Reviewed the sampling methods and the samples represent the participant populations.
3. Reviewed the savings spreadsheet used to estimate savings for lighting and appliances. (This is required because EnergyPro does not allow user input on these types of loads.)
4. Reviewed the reasonableness of the simulation software results by re-calculating the energy savings and demand reduction using the hourly kW values from EnergyPro.
5. Reviewed and commented on the Draft Process Evaluation Memo conducted by the Evaluation Team.
6. Reviewed and commented on the Draft Baseline Study report conducted by the Implementation Team.

The findings based upon the above activities are discussed in greater detail in the next subsection. Itron had planned, but was not able to a) verify that the appropriate weather normalization and billing usage data calculations were carried out, or b) conduct a more thorough analysis of single-family versus multi-family home types, as discussed in the PY2011 report. Time limits also prevented Itron from completing an in-depth review of the sensitivity analysis of building orientation.

Given that evaluation funds are allocated in proportion to each program's share of statewide savings claims, on-site data collection is not feasible on an annual basis. However, the importance of evaluating the programs based on primary on-site data collection cannot be overstated, but the relatively small fraction of the portfolio savings does not allow for this level of rigor. After delivery of data in response to a second data request, Itron discussed with the Evaluation Team an alternative evaluation approach that does not rely as heavily upon engineering reviews, simulations, or billing data analyses and instead focuses includes on-site data collection to support the process evaluation of the Program Administrator's quality control procedures every other year. Itron will explore these methods with the Evaluation Team as part of the 2014 evaluation planning process.

## 8.4 Verification Findings

This subsection discusses Itron's review of the program information and analysis methods used by the Evaluation Team to evaluate the reported impacts for the RNC programs. The evaluation of savings involved determining a statewide RR based upon simulation models of a sample of program participants and applying that RR equally to each utility's ex ante claimed program savings. Itron identified several issues including incorrect baseline assumptions for three sample homes, incorrect modeling of dishwasher savings, estimates of lighting savings that assumed HOU different than the TRM deemed value and did not include heating and cooling interactive effects, and an incorrect method of scaling the coincident peak demand impacts from the sample to the participant population. In terms of program design, Itron also found that the sampling approach accomplishes its certainty and precision but that it did not make the best use of the sample points, that energy savings do not correlate with HERS scores, and that large-capacity water heaters are being used to meet program energy efficiency targets. Itron discussed these issues with the Evaluation Team and some issues were addressed in the final report, as discussed in greater detail below. This subsection concludes with a discussion of alternative evaluation activities that may be a more efficient use of the evaluation budget in 2014.

### *8.4.1 Status of PY 2012 Verification Recommendations*

Recommendations from the PY2012 Verification Report, along with Itron's review of the Evaluation Team's responses, are summarized below:

1. **Recommendation:** Future evaluation efforts should include on-site data collection and verification activities of a statistically significant sample of participant homes according to the Strategic Evaluation Plan if the RNC programs savings are expected to reach 5% of the statewide portfolio savings.

   *Status:* Not fully implemented due to budget considerations and program impacts less than 5% of statewide portfolio. No on-site data collection.

2. **Recommendation:** The utility billing usage data for a census of homes from the current PY should be collected to verify occupancy during future evaluation efforts if on-site data collection is not feasible; also, freely available online aerial photography should be used to verify building orientation if on-site data collection does not take place.

   *Status:* This was partially implemented. Billing data were collected and used in an evaluation of the impact of occupancy. Further policy direction is needed with respect to the occupancy findings and how those should be used in future evaluation findings. Building orientation was not trued-up to actual construction due to lack of on-site data collection; a sensitivity study on orientation was conducted but did not address southwest and southeast orientations as needed to fully address the issue.

3. **Recommendation:** Future evaluations should use billing and weather data in REM/*Rate* (or similar simulation software) to adjust home savings and to supplement the use of 8760-hourly usage profiles generated by the Beacon PST. Such data should also be used to independently verify the DOE-2 simulation inputs and results.

   *Status:* This was fully implemented, but a manual calibration process and EnergyPro was used instead of REM/Rate. Some time savings and accuracy improvements may be found by using the built-in utility bill calibration features of EnergyPro or REM/Rate.

4. **Recommendation:** The Evaluation Team should use features of REM/*Rate*[9] that allow for the mass export of the building characteristics and savings results for the reference (baseline) and rated (as-built) home building characteristics if the Program Administrator does not provide the Beacon PST input files.

   *Status:* This was not required for EY4 because the program administrator provided the Beacon PST input files. Nevertheless, this was partially implemented in that the REM/Rate files were used to inform the EnergyPro simulations and external spreadsheet adjustment for lighting and appliances. Itron would like to discuss an alternate approach that assesses the REM/Rate files for the entire participant population, not just the sample.

5. **Recommendation:** The Program Administrator should immediately correct, adjust, and update the baseline reference home to reflect current ENERGY STAR Reference Home Guidelines, applicable IECC requirements, and Federal Appliance Standards. The Program Administrator should also address the potential for program requirements and ambiguous reference home rules to encourage electric space and water heating when gas is available.

   *Status:* This was partially implemented. The baseline home characteristics for the sample were mapped to either IECC 2006 or IECC 2009, depending upon the date the home was submitted for a permit and the jurisdiction in which it was located.[10] However, no attempt was made to review the availability of natural gas for the sample homes or to address program design issues that might influence the builder to select an electric water heater. Itron believes that it is inconsistent with statewide policy when a new program is being proposed to motivate homeowners to replace electric water heaters with natural gas water heaters, while this program implicitly promotes electric water heaters by allowing them to be built in homes where natural gas is available.

   Furthermore, Itron believes that the large electric storage water heater baseline issue allows the programs to be gamed by allowing other energy efficiency features to be eliminated that would otherwise be required to meet program rules. This results in higher energy bills for the customer, fewer electricity savings from the programs, and higher rebates to be paid for homes that might not otherwise meet program rules or which might otherwise employ other energy efficiency features to make up the deficit caused by an

---

9 Architectural Energy Corporation, *REM/Rate*, http://www.archenergy.com/products/remrate. See also Sentech, Inc., *Review of Selected Home Energy Auditing Tools*, January 2, 2010, http://apps1.eere.energy.gov/buildings/publications/pdfs/homescore/auditing_tool_review.pdf for a review of the different types of rating tools.

10 While the current statewide residential energy efficiency code is IECC 2012, the code applicable to each home depends upon the date its construction drawings were submitted to the local building department for approval. The participating homes in this program were submitted prior to the effective enforcement date for IECC 2012.

electric water heater when a natural gas one should have been installed. The NAECA 2015 appliance standards plug the water heater loophole by using an electric heat pump water heater as the baseline efficiency for water heaters above 50 gallons. Evidence that the programs could be increasing, rather than decreasing, electric consumptions includes: a) natural gas is available to the home as evidenced by the existence of a natural gas-fired space heating system, and/or b) a large electric storage water heater is installed in a home where the size or number of bedrooms in the home does not suggest that a large storage water heater of any type is required to meet hot water requirements.

### *8.4.2 Review of Tracking Databases*

As part of its verification effort, Itron obtained the tracking database from the Evaluation Team with the goal of independently verifying the quality of the dataset. However, the EY4 data seems to have some of these same issues found in EY2011 (EY2)—including zero savings homes and negative savings homes—and a different format for the PEPCO dataset, with some data fields missing all together. At this point, it is probably inappropriate to alter the dataset to account for these issues because the presumption is that the dataset represents the actual participation and performance of the programs. The zero and negative savings homes in program participation counts toward program goals but these cases need to be explained. For example, did the HERS rater determine zero savings through the normal one-in-seven testing rule and what are the process evaluation findings that would remedy this situation, if any? If the utilities were billed for rebates paid for new homes that did not save energy relative to the baseline home, then that fact should be included in the cost-effectiveness of the programs. However, in the normal course of reviewing the tracking database, the Evaluation Team and the Program Administrator need to have a discussion about any anomalies and come to an agreement about how to treat any uncorrectable anomalies that are found.

### *8.4.3 Review of Sampling Methods*

In previous evaluations, due to lack of data for the RNC programs, the Evaluation Team modified the sampling and evaluation approach to fit the available information. This year, the Evaluation Team benefitted from a more complete response to their data request and was able to pursue their objective to verify the reasonableness of the Beacon PST savings estimates. The Evaluation Team selected a random sample of 30 homes and accounted for the potential bias of multi-family versus single-family homes in the programs by selecting a number of samples of each home type in proportion to the prevalence of that type of home in the population—15 single-family and 15 multi-family homes. The sample size of 30 was based upon achieving an overall result with 15% precision with 85% confidence assuming a coefficient of variation in the population of 0.50. The sample design was not intended to provide statistically significant results at the utility level nor at the stratum-level for single-family and multi-family home types

separately. A total of only 1.73% of the programs' overall electricity savings was represented by the sample.

To verify that the sample is an adequate representation of the population, Itron analyzed the sample and population data and created plots to illustrate its findings. Figure 8-1 below plots the program participants' electricity savings (kWh impact) sorted along the x-axis from least to greatest savings per home. The blue bars identify the sample homes within the population of program participants shown in red.

**Figure 8-1: Distribution of kWh Savings in Sample and Population**

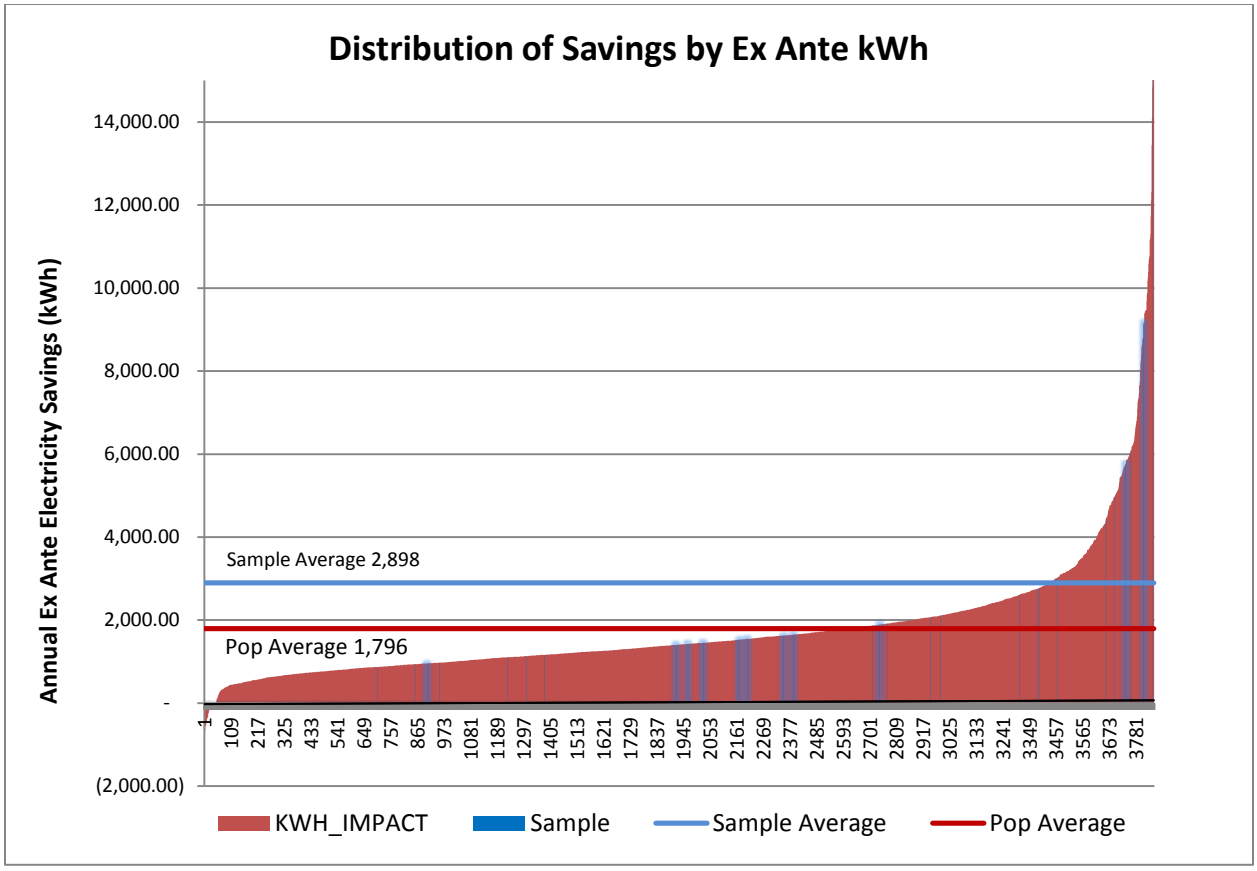
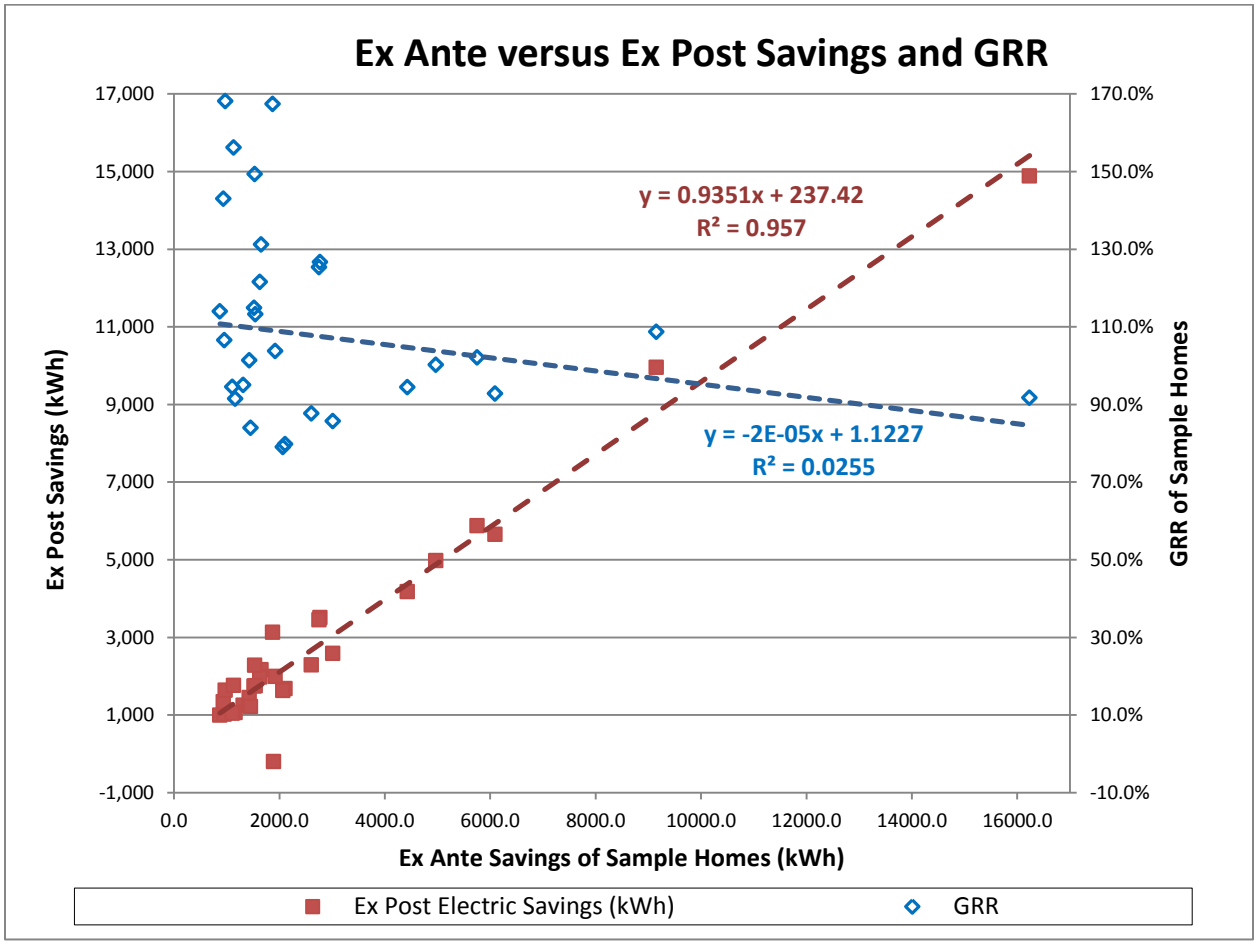



Figure 8-1 shows the skewed distribution of savings per home in the direction of the lower energy-saving homes, causing a minority of the high energy-using homes in the program to account for a majority of the savings. Usually this indicates that a stratified sampling approach would provide greater predictive strength with fewer sample points; however, the sample appears to include a larger percentage of these larger homes than the smaller homes. The average of the 30 sample points is 63% larger than the population average because the overall RR is used to adjust the claimed savings for the program overall. This does not introduce a bias in the results because the RR of the sample homes does not vary with savings per home. Figure 8-2 below explores this issue by plotting the relationship between energy savings per home and the RR.

**Figure 8-2: Realization Rate and Ex Post Savings versus Ex Ante Savings**




The red squares and dashed trend line in Figure 8-2 above show a very strong relationship between ex ante and ex post electricity savings ($R^2$ = 0.96), and the blue diamonds and short-dash trend line show a very weak negative relationship between ex ante savings and the RR ($R^2$ = 0.025). Since the RR does not vary significantly across the range of savings per home, the sample is likely to yield an adequate but slightly conservative estimate of savings for the programs.

The evaluation effort achieved precision of 0.0828 with a confidence of 85%, but the statewide sampling approach cannot provide robust results for sub-populations. If the program is expected to reach 5% of the statewide portfolio, future evaluation efforts should use a stratified sampling and estimation approach to ensure that the sample is representative of the various sub-populations such as each utility company, multi-family versus single-family home types, homes with larger savings, jurisdictions with different code requirements, regions where homes remain unoccupied longer, etc. Other factors where a stratified sampling approach would be indicated include multiple Program Administrators, high seasonal vacancy rates in certain regions, and

other factors that might vary from the assumption that all homes are constructed under the same program rules and under identical program rule enforcement regimes.

### *8.4.4 Review of the Quality of the Data for the Sample*

The Evaluation Team provided the final ex post savings spreadsheet that combines the simulation of space heating, space cooling, and water heating savings with the lighting and appliance savings for homes in the sample and calculates the sample RR. The issues included calculation worksheets that were "de-linked" from the other datasets, which fed into the final calculations. Also, the calculation of demand was not included in response to initial data requests. This introduced unnecessary delays in re-connecting these datasets and delayed the completion of the verification effort. It is consistent with best practices for datasets to be provided with all linkages intact, wherever possible, and an explanation of the basis for any values that are hard-coded had to be de-linked. To facilitate the verification efforts, Itron recommends that this procedure be followed more closely in the future.

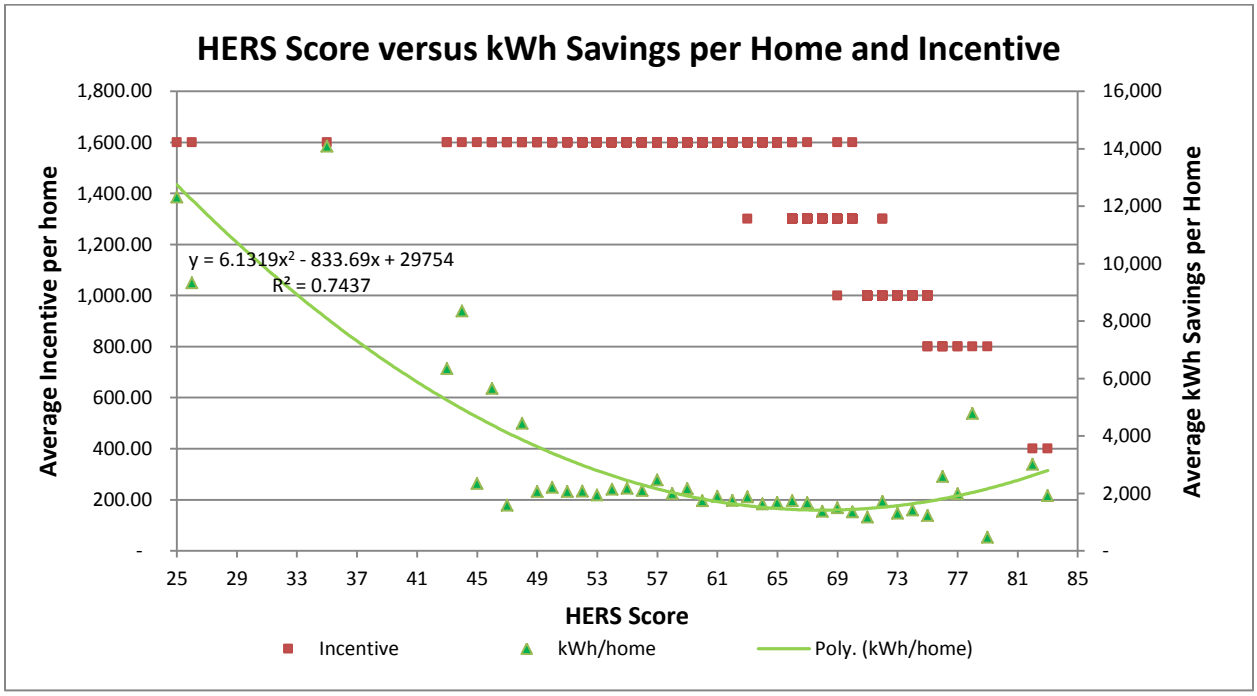
During the review, Itron noted that three sample points did not have the full complement of monthly billing usage data. In addition, the spreadsheet contained comments about the sample homes discussing issues that might indicate the sample includes some outliers, explaining some non-standard data inputs for the sample—including one home with no CFL lighting fixtures (100% incandescent)—and indicating some homes whose simulation shows excessive domestic hot water usage. Itron investigated these issues and found the possible error introduced by these omissions and outliers to be within reasonable limits.

### *8.4.5 Review of Incentive Levels and HERS Score*

A 2009 NYSERDA report on the evaluation of RNC programs in New York[11] noted concerns with structuring program incentive levels based upon the HERS score. The scope of this report does not allow for a full discussion of the topic, but the underlying issue relates to the way that the HERS score was designed to be "fuel neutral." The NYSERDA report found that energy savings per home did not significantly vary with HERS score; i.e., the better (lower) HERS scores did not equate to greater electricity or natural gas savings per home. To investigate the degree to which the EmPOWER RNC programs suffer from this issue, Figure 8-3 below plots the distribution of savings per home averaged for the group of homes of the same HERS score.

[11] NYSERDA, *New York Energy $Mart[SM] Program Evaluation and Status Report 2008*, March 2009, www.neep.org/Assets/uploads/files/emv/emv-library/2009-3_NY_EnergySmart_Program_Eval_Status_2008.pdf, page 4-30.

**Figure 8-3: Distribution of Savings and Incentives by HERS Score**




The horizontal axis of the plot in Figure 8-3 above is the HERS score with the highest potential energy savings homes (lowest HERS score) on the left and deceasing energy savings potential per home on the right. The average ex post savings of each HERS score bin is shown with the green triangles, and the incentive per home is shown with red squares. As the stringency of the energy efficiency construction requirement increases from right to left, the average energy savings per home is mostly flat between 50 and 77 HERS score range; however, the incentive amount increases dramatically from $1,000 to $1,600 across that same range. The extra incentives do not appear to be buying a proportionate increase in energy savings. This issue was discussed in a paper presented to the ACEEE 2010 Summer Study Conference.[12] Itron recommends that the Program Administrators adopt an alternate incentive structure that uses the HERS score for ease of program targeting, but pays incentives based upon annual energy savings as calculated by Beacon PST, REM/*Rate* fuel summary report, or similar. This issue may be related to the issue of large storage water heaters noted above.

### *8.4.6 Review of Lighting Savings Calculations*

Itron reviewed the inputs for the lighting savings calculations for the sample homes and noted inconsistencies that change the results and necessitated small adjustments to the RR for the programs. These issues included HOU for lighting that was inconsistent with statewide policy

[12] Khan, Haider and William Blake, "EUI: A Metric for Energy Savings for New Homes with ENERGY STAR Programs." Paper presented at ACEEE Summer Study 2010 Conference.

for residential bulbs, a baseline issue for one sample point affecting lighting savings, and a lack of an adjustment for heating/cooling interactive effect (WHF) to account for lighting savings that are calculated outside of the EnergyPro simulation software.

The lighting savings spreadsheet assumed 2.4 hours per day usage for indoor lighting end uses instead of the statewide approved value of 3.0 hours per day. Updating this value would usually result in a linear increase in energy savings, if not for the fact that one of the sample points was a home with negative savings because it was found to contain 100% incandescent bulbs. This was noted in the final savings spreadsheet with a comment stating that this finding was consistent with the REM/*Rate* input file. Increasing lighting HOU for the entire sample resulted in an increase of 25% of lighting energy savings and 2.4% increase in overall program-level electricity energy and demand savings. The Evaluation Team reviewed this finding but believes that 2.4 hours per day is reasonable for new construction homes where, on average, the locations within the home where the bulbs are installed are not expected to be lit for as long as the average bulb replaced in an existing home.

One sample home constructed in 2013 was indicated in the Draft Evaluation Report as using the IECC 2009 baseline in Table 18 but the IECC 2006 baseline in Table 26, indicating possible confusion about the applicable code requirements. The sample point was located in Gaithersburg, a city with a building department that enforced the IECC 2009 energy efficiency requirements beginning in July 2012, but within Montgomery County that did not enforce a building energy efficiency code at the time. A home must be constructed to meet the more stringent requirements when located within a jurisdiction whose enforcement or code stringency exceeds its "parent" jurisdiction's requirements. This determination was reinforced by the fact that the simulation input values for the EnergyPro baseline simulation model included construction features that were subject to change between IECC 2006 and IECC 2009 and which were consistent with the IECC 2009 requirements. This adjustment reduced lighting savings by 2.6% and overall savings by 0.12%.

The adjustment in lighting savings was accomplished with spreadsheet calculations outside of the simulation model. EnergyPro and other whole building thermal modeling tools automatically include lighting and HVAC interactive effects, and so its omission could skew the results. Itron used the results of a recent study in California and applied a 23% WHF[13] to lighting savings. Itron re-calculated the lighting savings for the sample and applied this to the program savings estimate, resulting in a 5.3% reduction in overall program kWh savings.

[13] A 14% WHF is consistent with the Mid-Atlantic TRM and is appropriate for use with deemed savings approaches. Whole-house simulation software such as REM/*Rate*, EnergyPro, and Beacon PST effectively use a higher value by directly calculating the interactive effects, so 23% is appropriate here because it is more comparable to the simulation-based results.

### *8.4.7 Review of Appliance Savings Calculations*

Itron reviewed the inputs for the appliance savings calculations for the sample homes and noted inconsistencies that changed the results and necessitated small adjustments to the RR for the programs. The appliance (dishwasher) savings algorithm divided the number of wash cycles by the dishwasher Energy Factor (EF) to determine savings. Itron reviewed other documents to verify savings and found multiple other sources showing much smaller savings, including one source that showed an increase in natural gas usage for homes with natural gas water heaters. Savings for dishwashers went down by 85% after using a look-up table to calculate savings instead of the algorithm. This change had an overall effect of a 3.8% reduction in the overall program electricity savings. The Evaluation Team reviewed this finding and concluded that their method was adequate for the current Evaluation Report and stated that any differences between their method and the methods documented in the reference document[14] will not significantly impact program savings. The large difference in the verified savings for the sample contradicts this statement.

### *8.4.8 Review of Coincident Peak Demand*

The Evaluation Team provided a spreadsheet containing the residential load shapes for four of the five utilities and a calculation of the peak coincident demand factor for each utility. According to page 50 of the Evaluation Report, these data were used in the calculation of utility coincident peak demand savings. These peak coincident demand factors were applied to the 30-home sample electricity savings, depending upon the utility to which each sample belongs, to determine the programs' statewide utility coincident peak demand impact RR. Applying the utility-specific peak coincident demand factor to each sample's kWh savings is incorrect because the per-utility sample weights are different than the population sample weights and because the 30-home sample is not significant at the utility level. Instead, Itron used the per-utility peak coincident demand factors, weighted them by the proportion of each utility's ex ante kW savings in the participant population, then multiplied these factors by the total verified sample kWh savings to determine the statewide peak demand kW savings. The ratio of the sum of the verified kW sample to ex ante kW sample total is the peak kW demand RR for the statewide program. Each utility's peak coincident demand savings is that utility's percent share of the population ex ante kW savings times the statewide kW total savings. Correcting the weighting of the peak coincident demand factor reduces the peak coincident demand savings by 5.6% and correcting the application of the ratio to the overall kW savings increases the peak coincident demand by 14.4% for a total adjustment compared to the ex ante by 7.9%. Overall, the effect is to increase peak kW demand by 46% over the evaluated values. The Evaluation Team reviewed

[14] Pennsylvania TRM: http://www.puc.pa.gov/filing_resources/issues_laws_regulations/act_129_information/technical_reference_manual.aspx

this finding and agreed that peak demand kW impacts should be adjusted to account for the population weights but could not revise their final report due to time constraints.

### *8.4.9 Review of EnergyPro Simulation Modeling*

The PY2013 evaluation plan for RNC programs specified that the Evaluation Team was to conduct an engineering review of the implementation contractor's software using DOE-2 or another more rigorous simulation tool. The primary objective of the Evaluation Team's effort was to assess the reasonableness of the Beacon PST savings estimates. This year, the effort was able to be carried out because the Program Administrator provided all of the requested information about the sample homes including "scrubbed" versions of its Beacon PST input files and REM/*Rate* files and the utility-provided billing usage data. The Evaluation Team used the EnergyPro v5 software—which employs the DOE-2 whole building thermal modeling and simulation engine to estimate envelope, heating, and cooling energy savings—and used a custom spreadsheet to estimate lighting and appliance savings.

While the close agreement between the ex ante and ex post savings for the sample homes lends support for the accuracy of the ex ante estimates and Beacon PST, Itron cannot verify the overall success of this effort. The sample was designed to estimate program savings, not to exercise the vulnerabilities of the software. The evaluation succeeds in showing that the ex ante savings are within a reasonable margin of error and the Evaluation Report correctly adjusts program electricity savings based upon these results.

Itron identified issues with the simulation inputs as well as the lighting and appliance savings spreadsheet. Each of these issues were corrected in isolation, where possible, to determine if each issue is significant and to determine an adjustment factor to be applied to the overall savings, if warranted.

The primary concern with the overall approach is the calculation of the lighting and appliance savings outside of the simulation of overall building thermal performance, rather than using a software simulation tool that can appropriately account for all of the impacts associated with lighting and appliance energy savings. This creates three significant issues for the evaluation effort, including: 1) the lighting and appliance savings do not play a part in the simulation's built-in capability to address lighting, heating and cooling interactive effects;[15] 2) without the lighting and appliance savings included in the model, the calibration of the models to the billing usage data is less accurate; and 3) the baseline models cannot benefit from the calibration of the as-built models. The first and second issues have reasonable workarounds as discussed elsewhere in this report, but the third issue means that the utility bill calibration built-in to

[15] Itron applied a WHF to lighting and appliance savings to produce a reasonable estimate of the heating and cooling interactive effects.

EnergyPro cannot be used to calibrate the model to daily or monthly usage patterns, and instead the calibration process is limited to the application of a linear scalar factor to the annual savings results. It is very difficult to assess the magnitude or direction of this potential source of error without a more sophisticated simulation tool that includes lighting and appliance savings in the estimates of home energy usage. At this point in time, the tool that comes the closest to achieving that goal is REM/*Rate*; however, it too suffers from a lack of fine-grained control over the lighting inputs and from the inability to calculate an hourly savings load shape or coincident peak demand. REM/*Rate* passed the BESTEST[16] battery of simulation results and can be relied upon for accurate annual energy savings estimates that are calibrated to weather and billing usage data in a manner as robust as the effort conducted by the Evaluation Team using EnergyPro.

### *8.4.10 Review of Electric End Uses and Natural Gas Availability at the Site*

Maryland is considering adding a new fuel switching program to the portfolio targeting electric water heaters. Itron reviewed the equipment specifications for the sample simulation models and found six homes that used electric water heaters but had a central furnace that used natural gas. This indicated that these homes had access to natural gas and that the selection of an electric water heater was for other reasons besides the availability of natural gas.[17] In some states, such as California, the building energy efficiency standards specify that the baseline home shall use natural gas fired water heater and space heating equipment if it was available "at the curb" for homes submitted for a new construction permit. California further stipulated that energy savings was to be determined using source energy (not site energy) that was adjusted for distribution and transmission losses. Because the site-to-source factor for electricity is almost three times larger than the value for natural gas, any home that specified an electric water heater in the as-built model created a huge deficit in the water heater energy budget for the home that could not be overcome without much more expensive construction features to compensate. This had the effect of eliminating electric resistance water heaters for most new homes unless gas was not available or if the home used a "non-depletable" fuel source such as wood for heating purposes.

Itron believes it is consistent with statewide policy objectives to create a similar market driver for RNC in Maryland by adjusting the program rules for the RNC programs to encourage early adoption of statewide policy. Future program cycles should assume natural gas water heating and space heating equipment in the baseline home if natural gas is available to the home. Enforcement of this rule can be improved by contracting with the makers of REM/*Rate* (and possibly other software makers) to develop a special rule set to be used for comprehensive

[16] Building Energy Simulation Test method: http://www.nrel.gov/docs/legosti/old/6231.pdf

[17] The EY3 Verification Report discusses the motivations to install electric water heaters, which may be the result of unintended consequences of the indoor air quality requirements of the ENERGY STAR New Homes program.

residential energy efficiency programs in Maryland, such as the RNC programs and possibly the Home Performance with ENERGY STAR program. Note that the rule must be more nuanced to address existing homes that may have natural gas available at the site, but not piped into the home, space limitations, exhaust gas considerations, and safety considerations due to interactions with other appliances and equipment in the home. Under this new rule, if a home is proposed to include an electric water heater even though natural gas is available, the baseline equipment efficiency is increased to the minimum efficiency of an equivalent natural gas water heater. This rule is viable; however, it does not fully reflect the impacts of electric water heaters on the grid. Using a time-dependent valuation of energy or a site-to-source EF that more heavily weights electricity savings over natural gas savings is advised to account for the extra transmission and distribution losses that would occur. Calculating energy savings in this way is effective because it does not "target" specific types of equipment that are likely to draw attention from some stakeholders.

To assess the magnitude of the effect the water heater baseline change (without implementing the site-to-source EF) would have on the RNC programs, Itron re-calculated savings for the six homes in the sample that had electric water heaters and gas space heating equipment using a natural gas water heater of the same capacity. The baseline efficiency was set to the NAECA 2015 minimum EF, as shown in Table 8-3, and the as-built efficiency was set to the average efficiency of the sample homes within each capacity bin, as shown in Table 8-4.

**Table 8-3: NAECA 2015 Water Heater Efficiency Standards**

| Capacity (Gallons) | Electric (EF) | Natural Gas (EF) |
|---|---|---|
| 40 | 0.917 | 0.594 |
| 50 | 0.904 | 0.575 |
| 55 | 0.897 | 0.566 |
| 60 | 0.891 | 0.556 |
| 65 | 0.884 | 0.547 |
| 75 | 0.871 | 0.528 |
| 80 | 0.864 | 0.518 |

**Table 8-4: Average Water Heater Efficiency for Sample Homes**

| Capacity (gallons) | Baseline (EF) | As-built (EF) |
|---|---|---|
| 40 | 0.594 | 0.610 |
| 50 | 0.547 | 0.638 |
| 65 | 0.546 | 0.660 |
| 75 | 0.487 | 0.800 |
| 80 | 0.487 | 0.800 |

Itron determined the potential impact of requiring natural gas water heaters where natural gas is available by changing the water heater fuel type from electricity to natural gas for the six sample homes where the space heating equipment is powered by natural gas. Table 8-5 below compares the fuel consumption of the sample homes under these two scenarios. In summary, the overall program water heating electricity usage and peak demand decreases by 3.2% and overall electricity usage and peak demand decreases by 0.55%, while the natural gas usage increases by 7.5% overall.

**Table 8-5: Impact of Requiring Natural Gas Water Heaters**

| Energy Impacts of Natural Gas Water Heaters | Electricity Use for Water Heating (kWh) | Electric Peak Demand for Water Heating (kW) | Total Household Electricity Use Total (kWh) | Total Household Electric Peak Demand Total (kW) | Total Household Natural Gas Total (Therms) |
|---|---|---|---|---|---|
| Electric or gas water heating equipment as found in sample | 9,572 | 5.57 | 54,831 | 31.89 | 5,585 |
| Only Change electric water heating to gas storage water heating when where gas is available (natural gas furnace) | 9,268 | 5.39 | 54,528 | 31.71 | 6,007 |
| Difference | -304 | 0 | -303 | -0.18 | 422 |
| Percent Difference | -3.2% | -3.2% | -0.55% | -0.55% | 7.5% |

Itron recommends changes to the RNC programs to address electric water heaters where natural gas is available to the home and to take advantage of the opportunity to push early adoption of the NAECA 2015 water heater standards.

### *8.4.11 Baseline Building Characteristics for Simulations*

Itron reviewed the baseline building characteristics in the sample homes to verify that the baseline energy use matched code minimum requirements, including the applicable IECC building energy efficiency and NAECA appliance standards. In the 2012 Verification Report, Itron reviewed in detail the IECC baseline and ENERGY STAR 3.0 Guidelines document. This year, the baseline review focused on the modeling inputs and was limited to the water heating equipment—the largest single end-use category of electricity and natural gas in residential construction. Itron found two sample homes where the water heater baseline efficiency was incorrect; one of these two sites also used a different input capacity for the baseline water heater versus as-built model.

According to the NAECA 2015 water heater efficiency standards[18] shown in Table 8-3 above, the minimum efficiency for an 80-gallon electric water heater was 0.864 EF, but the baseline model for BGPRPS1526230078 specified 0.850 EF. Additionally, the NAECA 2015 minimum efficiency for a 75 gallon gas water heater was 0.528 EF but the baseline model for DPNAPS1526269002 used 0.447 EF. Furthermore, for this sample, the as-built simulation model water heater used an input capacity of 300,000 Btu/hour, whereas the baseline input capacity was 65,000 Btu/hour. The larger input capacity value put the water heater into a commercial regulatory category with different baseline requirement than the baseline water heater. The higher capacity value appeared to be a data entry error because the water heater model number in the EnergyPro simulation file matches a water heater[19] that actually has a 75,000 input capacity.

Itron corrected the simulation models by assigning the EF originally specified in the EnergyPro model and adjusting the input capacity to the value matching the model number from the simulation file. The change in savings was based on the proportional change in savings from the re-calculated simulation models. These changes resulted in a 0.5% reduction in electricity savings for sample BGPRPS1526230078 and a 50% reduction in natural gas savings for sample DPNAPS1526269002. The impact on overall program savings was not significant for electricity and not possible to determine for natural gas because no ex ante or ex post natural gas savings were provided for this sample point. Itron did not apply these adjustments to the overall program savings because without on-site data to verify the water heater information, uncertainty remains regarding what was actually installed. This finding underscores the need for onsite data collection.

### *8.4.12 Potential Bias in Saving Estimates for Multi-family Buildings*

In the Verification Reports for 2011 and 2012, Itron discussed the potential bias related to multi-family buildings or homes with adjoining outside walls. In EY4, the Evaluation Team used a weighted sample to ensure that multi-family and attached single-family buildings are included in the sample in the same proportion as the population. In addition, because the baseline homes were modeled with the same distribution of windows and walls on each facade as the as-built models, the simulation-based savings calculation addressed some of the concerns related to the way baseline homes are modeled. The baseline home is usually adjusted by software such that the distribution of window and wall areas on each side of the home is equal and the window area is increased to the maximum allowed by code. Because the REM/*Rate* is used to determine the HERS score and program eligibility, its calculations will show greater savings for homes with adjoining walls and these homes will require fewer upgrades to meet program targets as

[18] http://www1.eere.energy.gov/buildings/appliance_standards/product.aspx/productid/27

[19] Manufacturer data for Bradford White model M-2-T2-75T6BN/SX.

compared to single-family homes with no adjoining walls. This is because homes with a lower percentage of walls and windows exposed to the outdoor environment will have lower heat gains and losses and require less energy from the building heating and cooling systems to maintain comfort.

The programs continue to enroll a large number of attached single-family, row-house style homes that are inherently more energy-efficient than detached single-family homes. As discussed in prior Verification Reports, these homes typically produce fewer savings per home but get the same incentives. Itron recommends that the programs develop program features to level the playing field for these two different types of home construction, considering both energy savings estimates and incentives. Using a sample weighting process can only go so far to correct for error caused by limitations in the modeling software.

### *8.4.13 Adequacy of Metering Data Collection and Weather Normalization*

The 2011 and 2012 Verification Reports for the RNC programs recommended the use of metering, billing usage, and other on-site data collection activities consistent with the associated evaluation plans and budget. This recommendation was based upon an understanding of the programs' relatively small proportion of statewide portfolio savings. Billing usage data were subsequently employed to evaluate program savings for 2012 and EY4. The EY4 Report benefits from extensive billing usage data for the participant population and weather normalization. The data were used to determine the length of time that new homes remained unoccupied (see the next subsection) and to adjust the billing usage data to match the typical meteorological year for comparison with the simulation results.

An alternative approach that uses the built-in weather calibration features of the modeling software is preferred, but requires the software to automatically create the baseline model. EnergyPro has the capability to calibrate the model to weather data, but lacks the ability to automatically create a baseline model appropriate for the IECC codes and the user cannot specify different lighting and appliance loads for the baseline model. This necessitates the use of two simulation models for each home; one with the baseline construction characteristics and another for the as-built home. This approach requires a more painstaking process to calibrate the baseline model to billing usage that is probably not cost-effective.

Based upon a review of publically available information on the Beacon software, Energy Plus, and based upon first-hand use of the REM/*Rate* software and EnergyPro, these software programs provide the ability to incorporate monthly usage data and local weather for the matching time period to calibrate the simulation results. Appropriate specifications for lighting and appliances are missing from most of these software tools. These features, when fully implemented, will greatly improve the quality of the evaluation effort without incurring significant data collection or analysis costs. Itron recommends that the Program Administrator

engage with the developer of REM/*Rate* to ensure that it includes all appropriate as-built and baseline specifications for the upcoming program cycles for lighting and appliances.

To provide an additional level of assurance that the simulation approach is providing reliable results, Itron conducted a partial billing analysis, without weather adjustment, using the data provided for BGE because it was the only utility with a significant number of participant homes with billing usage data. The Itron review found savings for those homes with exactly 12 months of billing data was 12% higher than the evaluated savings. This result is within the range we would expect from weather adjustment and close to the error of the evaluated savings for BGE. This exercise provides further evidence that the simulation results are reliable.

### *8.4.14 Adequacy of Analysis of Unoccupied Homes*

Partially in response to the 2012 Verification Report's discussion of potentially unoccupied homes, the EY4 Evaluation Report includes a detailed review of unoccupied homes in the programs. The importance of this effort is to appropriately adjust program savings to account for the period when no one is living in the home. In addition, the market for new homes is changing. During the economic downturn, builders typically were not constructing a home until there was a known buyer. This means occupancy of the homes was likely to occur soon after the home was completed. It appears that speculative home construction is returning as a common practice. This will likely mean that the delay between completion of the home and the beginning of occupancy will increase.

The report estimates the number of homes that were unoccupied for each utility service territory and finds that only one home was unoccupied for all months of the year. There is considerable variability in the findings across the utilities but there is no discussion about why occupancy might vary in those locations.

For future evaluation efforts, Itron recommends that billing usage data for a statistically significant sample of program participants continue to be used to determine the delay between construction completion and occupancy. The report should include an estimate of average duration of vacant homes and discuss the market factors that could influence these findings. If warranted, Itron recommends than an occupancy adjustment factor or other process to account for the period of vacancy should be included in the calculation of program savings.

Not all of the issues discussed above have a direct impact on the savings estimates for the program. In the next subsection we review those issues which had a significant impact on program savings.

## 8.5 Adjustments to Evaluated Savings

This subsection summarizes the adjustment to evaluated savings, discusses Itron's review of the Process Evaluation Memo, reviews the main issues identified in our review of the Draft Baseline Study conducted by the Program Administrator, and addresses the adequacy of the net savings estimates. The following subsection summarizes Itron's recommendations for future evaluation cycles.

### *8.5.1 Sensitivity Analysis of Findings that Affect Savings*

The following sub-sections discuss Itron's review of NTG adjustment to savings, process evaluation findings, and the baseline study of residential home construction practices.

### *8.5.2 Review of Net Savings Estimates*

Itron is unaware of any rigorous NTG evaluation efforts of the RNC programs. Previous evaluations have relied upon secondary research. The 0.84 NTG ratio is within the range found in evaluations of RNC programs run in other states. Since the programs were implemented by the same Program Administrator, it is reasonable to apply the same value for all utility company service territories. Maryland has only recently started implementing energy savings programs after a long hiatus, so it would be expected for free ridership to be on the low side, contributing to a higher NTG ratio for the programs. However, due to the recent adoption of the IECC 2012 baseline, we recommend that free ridership and spillover be addressed and that the recently completed Baseline Study by ICF, with adjustments, be the starting point for these efforts.

Some of the findings from the focus group effort in 2012 suggest that a more rigorous NTG evaluation effort is warranted. Some builders stated that they did not find ENERGY STAR 3.0 home designs to be any different than ENERGY STAR 2.5 (paraphrasing). The Evaluation Team found that participant homes sold for a significantly higher price than non-participant homes (approaching 15%), while the costs to participate were in the range of $1,000 to $4,000 per home. Itron recommends that future NTG evaluations include a more detailed review of the incremental costs to build participant homes as compared to standard construction. Itron also recommends that the NTG evaluation involve a greater number of non-participant builders and that the survey be expanded to include more questions on free ridership.

In future verification efforts, Itron may no longer be able to support the 0.84 net evaluated savings ratio reported by the Evaluation Team and included in Table 8-2. Itron also identified a number of issues that emerged from our review of the baseline study that should be addressed during the current evaluation planning cycle.

### *8.5.3 Review of Draft Baseline Study*

Itron reviewed the Draft Baseline Study by ICF and provided comments in an annotated PDF document to the Evaluation Team. The study appears to be overly constrained by budget, resulting in a sample of only 60 homes that did not closely follow the sample design, was conducted immediately after the IECC 2012 code was adopted, and shows bias in its approach to estimating the current building industry standards of construction. Itron recommends revising the Baseline Study's conclusions after conducting additional field research. Also, it is consistent with evaluation best practices to include market baseline issues as part of the gross savings calculations, not the net impacts.

The Baseline Study sample design called for a sample size of 60 homes weighted by construction volume with a target minimum of one sample point for each of Maryland's 24 counties. The study ultimately sampled 64 homes but was limited to the 13 counties where there was construction greater than 3% of the statewide construction volume. The final distribution of sample points appears to differ substantially from the initial design; for example, Baltimore County and St. Mary's County—each with 6% of the overall volume of construction—were treated very differently. Baltimore County was slated for four homes but 15 were included in the final sample, whereas St. Mary's county was slated for five sample points but none was included in the final sample. These two counties appear to be very different in terms of the socioeconomic background and weather, two factors that would be expected to influence building construction practices.

A total of 27 builders were included in the on-site survey but the report does not discuss the criteria used to determine if non-participant homes constructed by ENERGY STAR builders were excluded from the sample. The report states that data collection efforts were constrained by limited availability of non-ENERGY STAR builders and noted that builders were predisposed to permit their homes under the IECC 2009 standards. Itron questions if the study's timeframe is appropriate to capture a set of construction features that would be typical of Maryland's building practices after sufficient diffusion of the new code requirements into the new construction market. Did all of the permitting agencies enforce the new code to the same level of compliance or are any counties not expecting builders to meet the new code until sometime in the future? While the study found that builders in the sample were aware of the IECC 2012 code, did they have enough time to review the code and develop a complete awareness of the new code's implications? On average, how many homes did each builder construct under the new code requirements?

The issues with sample of homes and builders suggest that the baseline data collected by the survey would tend to favor higher levels of energy efficiency, and another respected organization in the building industry appears to agree. The Baseline Study references an ACEEE State Energy Efficiency Scorecard that found Maryland ranked 9$^{th}$ in the country for overall energy

efficiency and 19th in terms of code enforcement. Given this evidence, the Baseline Study ignores evidence that the homes are in fact being constructed somewhat better than code in one important respect; namely, the finding that the weighted study average window performance (0.33 U-factor, 0.27 SHGC) is significantly better than the code requirements (0.35 U-factor, 0.40 SHGC). Instead, the study finds that the baseline homes will consume an average of 800 kWh more than code-compliant homes, but does not discuss the relative performance improvement of the homes if the weighted study average window was used in the simulations.

## 8.6 Recommendations

Itron concludes that the evaluated savings estimates for the RNC programs are reasonable but recommends minor adjustments.

Itron offers the following recommendations for the next gross impacts evaluation:

- Improve review of tracking database: After a thorough review and cleaning of the tracking database, the Evaluation Team and the Program Administrator need to agree upon what constitutes the final tracking database for the programs. If this includes anomalies such as zero-saving or negative-savings homes, then an explanation for each of these cases should be provided if they are to be included in the final tracking database.
- Provide evaluation datasets with links to data and/or reference documents: Provide evaluation datasets with all linkages intact, wherever possible, and provide an explanation of the basis for any values that are hard-coded or that had to be de-linked. To streamline our complementary evaluation and verification efforts, Itron recommends that this procedure be rigorously followed in the future.
- Primary on-site data collection: Future evaluation efforts should include some form of on-site data collection and/or verification activities. These can be conducted using a ride-along sampling method for a portion of the sample of prototype homes to distribute the effort over the implementation cycle and to minimize the impact on the builder, Program Administrator, and homeowner. Coordination with the process evaluation effort is essential to take advantage of the opportunity to review QA/QC procedures and to promote diffusion of the program goals to the largest audience.
- Verify and update building orientation: Update the orientation sensitivity study to include southeast and southwest orientations. If orientation is found to be a significant factor, include home orientation in the final simulation of home energy savings.
- Revise calculation of peak coincident demand: Since this evaluation is based upon a ratio of savings between the sample of 30 homes and the participant population of homes, the calculation of peak coincident demand impacts must correspondingly be based upon the

statewide, population-weighted ex ante kW impacts. More careful attention and QA/QC review of the calculation methods is needed.

- Revise program incentive structure: Itron recommends that the Program Administrators adopt an alternate incentive structure that uses the HERS score for ease of program targeting, but pays incentives based upon annual energy savings. This is needed to address findings by Itron and other evaluators of RNC programs that the HERS score is poorly correlated with per-home energy savings.
- Continue to update the reference home rule set: The Program Administrator should periodically review and adjust, and/or update the reference home rule set to reflect current ENERGY STAR Reference Home Guidelines, applicable IECC requirements, findings from the Baseline Study, and Federal Appliance Standards. The more stringent requirements of these three sources should be used to determine the reference home construction and operating characteristics. Itron notes that the NAECA 2015 changes to the efficiency requirements for large storage water heaters represents an opportunity for program enhancement to lay the groundwork for early adoption of the new standards statewide after a period of introduction through the RNC program.
- Calibrate models using billing data and recalculate savings: Future evaluation efforts should periodically include re-calculation of the simulation results using any of the BESTTEST certified simulation tools. Itron recommends that billing usage data and matching local weather data be used within the simulation tool to calibrate home savings. If the tool does not support the output of 8760 hourly data, use of Beacon PST-generated 8760-hourly usage profiles or the use of a representative sample of typical construction is adequate.
- Natural gas reference home fuel type: For future evaluation efforts, the Evaluation Team should address the potential for perverse program influences as a result of program requirements and inappropriate reference home rules. The programs' focus on tight home construction combined with using the as-built home construction characteristics to determine the baseline fuel type for space heating and water heating could lead to increased installation of electric resistance water heaters in order to mitigate indoor air pollution issues. The NAECA large storage water heater baseline issue further compounds the severity of this issue. Instead, the baseline fuel type for space heating and water heating should be based upon the availability of natural gas to the site and a site-to-source or other program structure used to appropriately account for transmission and distribution losses to promote natural gas water heaters. The Evaluation Team may need to conduct on-site data collection activities when it is not clear whether or not natural gas is available from the utility bills.
- Correct for multi-family/attached home savings bias: The Evaluation Team should continue to address the potential for upward bias in the impact estimates for multi-family homes by stratifying the sample according to building type (single-family attached,

single-family detached, multi-family) and including a custom reference home that is appropriate for multi-family dwellings. The possibility of poorer construction characteristics in non-single-family detached homes being traded-off against other building features—and the corresponding increased utility bills—should be mitigated through adjustments to the reference home model. The occupants of multi-family homes are the ones who can least afford higher utility bills as a result of lower energy efficiency traded-offs in the name of construction cost reductions.

Recommendations related to the process and NTG evaluation effort include:

- ICF Baseline Study: Increase the sample size in the underrepresented jurisdictions and revise the findings of the Baseline Study to accurately reflect the typical construction characteristics found, even if those findings suggest that any specific home construction characteristic, i.e., windows, are better than the code requirements.
- Net savings evaluation: Future evaluation efforts should include data collection and verification activities of participant and non-participant builders. Itron found that the NTG ratio of 84% used by the Evaluation Team to adjust gross savings estimates is consistent with evaluated NTG ratios found for similar programs in other service territories. However, due to the recent adoption of the IECC 2012 baseline, we recommend that free ridership and spillover be addressed and that the ICF Baseline Study be the starting point for these efforts. In future verification efforts, Itron may no longer be able to support the 0.84 net evaluated savings ratio.
- RESNet QA/QC Procedures: In lieu of recalculation of simulation savings, the Evaluation Team should verify that RESNet and HERS rater quality control procedures are being followed by the Program Administrator and participating HERS raters. The evaluation efforts should document when shortcomings are found in a sample of homes and track what happens to those homes as the issues are hopefully addressed and remedied by the builder.
- Software baseline adjustments: The Program Administrator should ensure that REM/*Rate* includes all appropriate as-built and baseline specifications for the upcoming program cycles for lighting and appliances.

Statewide policy issues that need to be addressed include:

- Consistent treatment of baselines: The baseline should include the effect of above-code as well as below-code construction characteristics, either in the gross savings estimate or as a contributing factor in the calculation of net savings. For example, currently the Baseline Study ignores its finding of greatly improved window efficiency in the general home construction market. If the Baseline Study finds that non-program (i.e., baseline) homes do not meet current code requirements, should this divergence between standard

practice and code be accounted for in the net savings estimates or the gross savings estimates?

- Verify occupancy: Implementers should collect billing usage data for the participant population to verify occupancy before incentives are paid or adjust savings with an occupancy adjustment factor or adjustment to the effective useful life of the RNC programs.

# 9

# Commercial & Industrial Prescriptive and Small Business Programs

## 9.1 Verification Summary

This subsection discusses the results of Itron's verification review of the Evaluation Team's Commercial and Industrial Prescriptive (C&IP) and SB Programs 2013 Evaluation Report.[1] The goal of this effort was to review and verify the methods used to estimate gross and net energy impacts for these two programs, identify issues, and provide recommendations to improve future evaluations. This Verification Report covers the EY4 program activity, which covers the period from June 1, 2012 through May 31, 2013. CY2013 evaluation results[2] for *all* programs were discussed in Section 2.

### *9.1.1 Summary of Gross and Net Impacts*

Table 9-1 and Table 9-2 present a summary of the Evaluation Team's evaluated and Itron-verified EY4 gross and net impacts for the C&IP and SB programs. Because no adjustments were made by Itron, the evaluated and verified values are the same.

1 Navigant, *EmPOWER Maryland Final Impact Evaluation Report, Evaluation Year 4 (June 1, 2012 – May 31, 2013), Commercial & Industrial Prescriptive and Small Business Programs*, March 31, 2014.

2 *EmPOWER Maryland Final Impact Eval Report: CY2013.*

### Table 9-1: Summary of Evaluation Year 4 Evaluated & Verified Gross and Net Impacts for the C&I Prescriptive Programs

| | Gross Impacts | | | | Net Impacts | | | |
|---|---|---|---|---|---|---|---|---|
| | Evaluated | | Itron-Verified | | Evaluated | | Itron-Verified | |
| **Utility** | **MWh** | **kW*** | **MWh** | **kW*** | **MWh** | **kW*** | **MWh** | **kW*** |
| BGE | 81,185 | 13,088 | 81,185 | 13,088 | 58,453 | 9,423 | 58,453 | 9,423 |
| PEPCO | 32,423 | 6,017 | 32,423 | 6,017 | 23,345 | 4,332 | 23,345 | 4,332 |
| DPL | 6,885 | 1,036 | 6,885 | 1,036 | 4,957 | 746 | 4,957 | 746 |
| PE | 5,374 | 510 | 5,374 | 510 | 3,869 | 367 | 3,869 | 367 |
| SMECO | 3,449 | 788 | 3,449 | 788 | 2,483 | 567 | 2,483 | 567 |
| **Statewide** | **129,317** | **21,439** | **129,317** | **21,439** | **93,108** | **15,436** | **93,108** | **15,436** |

* Utility Coincident Peak Savings

Source: *EmPOWER Maryland Final Impact Eval Report EY4: C&IP & SB Programs*

### Table 9-2: Summary of Evaluation Year 4 Evaluated and Verified Gross and Net Impacts for Small Business Programs

| | Gross Impacts | | | | Net Impacts | | | |
|---|---|---|---|---|---|---|---|---|
| | Evaluated | | Itron-Verified | | Evaluated | | Itron-Verified | |
| **Utility** | **MWh** | **kW*** | **MWh** | **kW*** | **MWh** | **kW*** | **MWh** | **kW*** |
| BGE | 24,743 | 5,276 | 24,743 | 5,276 | 18,310 | 3,904 | 18,310 | 3,904 |
| PEPCO | 9,963 | 2,595 | 9,963 | 2,595 | 7,373 | 1,920 | 7,373 | 1,920 |
| DPL | 3,817 | 658 | 3,817 | 658 | 2,825 | 487 | 2,825 | 487 |
| PE** | 29 | 6 | 29 | 6 | 21 | 4 | 21 | 4 |
| PE (EE Kits) | 8,608 | 1,755 | 8,608 | 1,755 | 4,218 | 860 | 4,218 | 860 |
| SMECO | 852 | 232 | 852 | 232 | 630 | 172 | 630 | 172 |
| **Statewide** | **48,012** | **10,522** | **48,012** | **10,522** | **33,377** | **7,348** | **33,377** | **7,348** |

* Utility Coincident Peak Savings

** The SB direct install (non-kit) projects were evaluated as part of the PE Prescriptive program evaluation due to its small size.

Source: *EmPOWER Maryland Final Impact Eval Report EY4: C&IP & SB Programs*

For EY4, savings for the PE SB programs were dominated by the Energy Efficiency Kits (EE Kits) program component. Because the EE Kits were the predominant measure for the PE SB program, this element was evaluated separately from the main PE SB program. In addition, the much smaller PE SB direct install component was rolled into the PE Prescriptive evaluation. However, commercial EE Kits were discontinued after 2012.

### *9.1.2 Key Recommendations to Improve Next Year's Program Evaluation Plan*

The key recommendations resulting from Itron's review of the Evaluation Report are summarized below and discussed in more detail in Subsection 9.7:

- The tracking data and the evaluation should use consistent descriptions of building types because building types are a critical parameter for estimating savings for these programs. Selection of the appropriate building type is one of the key evaluator decisions that has a significant effect on C&I savings estimates. This issue should be addressed in the next evaluation cycle.
- Review and consider incorporating the results of the NEEP Study when it becomes available in 2014, but also consider conducting primary M&V.[3] Unfortunately, the NEEP Study was not available in time for this effort, but as the Evaluation Team's detailed analysis clearly details, the savings approach for this measure should be cleaned up and standardized across utilities.
- Review the approach and underlying self-reported lighting hours of operation that were used to develop the CSRR to determine if these values should be estimated separately for each utility. Each utility currently uses a slightly different method for estimating the customer self-reported lighting operating hours, but the evaluation treated these methods equally in calculating the overall CSRR.
- Consider revising and updating the NTG ratio values, except for PE SB EE Kits which were developed under the EY4 evaluation, as the programs are well established and the current values are several years old now.
- Consider the lighting early retirement baseline issue for the T12/T8 linear fluorescent phase-out. The Mid-Atlantic TRM provides an approach for the phase-out, and the Evaluation Team should determine if utility calculators and savings calculations are consistent with the TRM approach. If they are not, the utilities should be advised on the changes that are needed.
- Estimates of savings for the PE SB EE Kits should use the HOU for CFLs specified in the latest version of the Mid-Atlantic TRM rather than HOU from the most recent metering study. The Evaluation Team used the HOU value from the lighting metering study, but this value is likely more representative of non-CFL lighting fixtures, and there can be a significant difference in the operating hours between CFLs and linear lighting fixtures for some business types. However, as EE Kits have been discontinued, this is a minor issue.

[3] NEEP is considering a few drastic revisions: One option being discussed is dropping the building types completely and using deemed values by application, and increasing the size range up to 200HP. However, the data and analysis are still under development but are expected to be posted July 2014.

## 9.2 Overview of the 2013 C&IP and SB Programs

Together, the C&IP and SB programs account for about 32% of the overall EmPOWER CY2013 portfolio savings. The largest share is contributed by the C&IP programs at 20% of statewide portfolio energy savings, and the SB program is about 12%.[4] For both the C&IP and SB programs, the largest share of EY4 energy savings is contributed by BGE at 62.8 % and 52%, respectively.[5] The majority of program energy savings are still from lighting, HVAC, variable speed drive (VSD) measures, and EE Kits, although unlike previous years, the Evaluation Report did not provide measure-level savings totals. A brief overview of each program is provided below.

### *9.2.1 Overview of the 2013 C&I Prescriptive Program*

The C&IP programs implemented by the EmPOWER utilities support small and large C&I customers in identifying and implementing cost-effective energy efficiency opportunities by offering incentives for measures that span various technology end uses. All five EmPOWER utilities offer incentives for HVAC equipment, variable frequency drives (VFDs), lighting fixtures, lighting control measures, and commercial kitchen measures. The PHI, BGE, and SMECO programs also include refrigeration equipment incentives and new construction, performance-based lighting incentives.

For PY 2013, PE also began offering a new water heating program and a new food service/commercial kitchen program. In addition, PE offers incentives for multi-family efficient appliances in multi-family buildings, and both PE and DPL offer incentives for small business specialized controls. The incentives and reported savings of each utility are based on predefined savings values and calculation methods, which are often incorporated into savings calculators.

The C&I Prescriptive programs comprised 18% of the calendar year 2013 EmPOWER Maryland MWh savings and 19% of the MW savings.

### *9.2.2 Overview of the 2013 Small Business Programs*

The C&I SB programs implemented by the EmPOWER utilities support small business customers in identifying and implementing cost-effective energy efficiency opportunities by offering incentives for measures that span various technology end uses. Direct install measure programs are offered to customers with a simplified application and installation process. Only BGE, SMECO, and PE offered direct install programs to their customers during EY4. All of the EmPOWER utilities offer some kind of technical assistance to their small business customers. The technical assistance may include an energy audit or energy assessment of the small business

[4] CY results from *EmPOWER Maryland Final Impact Eval Report: CY2013.*

[5] EY results from *EmPOWER Maryland Final Impact Eval Report EY4: C&IP & SB Programs.*

facility. A significant change for PE, starting and ending in 2012, was the mass offering of EE Kits to commercial customers. These kits include CFLs, low-flow water aerators (only if electric water heating), and a Smart Strip. Six different kit configurations were sent out to business customers, based on the business size and whether or not they had electric water heating. For PE, as shown in Table 9-2, the EY4 SB program savings are almost entirely from the mailing of EE Kits.

The SB Direct Install programs comprised 12% of the CY 2013 EmPOWER Maryland MWh savings and 19% of the MW savings.

## 9.3 Evaluation Summary

The Evaluation Team developed an estimate of the evaluated savings for the EY4 period from June 1, 2012 through May 31, 2013. CY savings were not addressed at all in the program-level Evaluation Report, but are instead addressed in an Overview section of the report, as previously mentioned. The sampling plan was designed on a one-tailed 90/20 confidence interval and precision level for all study aspects of the C&IP programs, as well as the SB evaluation. As in previous years, engineering reviews, phone surveys, on-site verifications, metering data, new metering, and results from previous metering efforts were used. Specific study components of this year's evaluation include the following efforts:

- **Lighting Metering Study–TRM Parameters.** The goal of this aspect of the evaluation effort was to generate improved assumptions for the hours of use, coincidence factors, and CSRR[6] factors by leveraging all metering data collected in previous evaluation years (2010-2012) along with new data from this year (EY4), and performing the analysis at a "space-type" level (e.g., restroom, office, hallway, kitchen, retail sales floor). The Evaluation Team used the same M&V approach that was used in previous years, including extrapolation of a few weeks of logger data to a full year. Although the data were collected and developed at the space-type level, it was ultimately aggregated to the building-type level. Both the building-type and space-type level values are presented in the Evaluation Report. For the new study, 62 sites yielded lighting usage data from 323 lighting loggers, bringing the total metering sample for all previous years to 172 sites.
- **PE SB EE Kits Verification.** Savings from the PE SB EE Kits have not been evaluated in the past, but represent a majority of the EY4 PE SB program savings, and also represent a significant portion of the overall SB program energy savings across all

[6] The CSRR is the ratio of logged operating hours to operating hours reported by the customer on customer applications. Three different methods are used by the utilities to gather customer-estimated hours of operation: 1) A single annual hours estimate, 2) An average weekly hours estimate (which gets expanded to annual hours), and 3) A three-parameter method: Hours per day, the number of days per week of that day type per week, and the number of weekdays per year that the lights are off. These values are also expanded to annual hours.

EmPOWER utilities (18%). The evaluation of this program component investigated the customer installation rates for the kits after delivery by mail, and deemed savings values per measure mailed. The Evaluation Team conducted interviews with participating customers to verify quantities of EE equipment they received and what they did with the equipment. The Evaluation Team also asked utility staff to verify the equipment that was sent out and the detailed characteristics (CFL wattages, quantities, etc.) of that equipment. A 90/15 one-tailed confidence and precision were targeted for the gross savings estimated for this program. The ex post savings were calculated based on participant reported use/installation rates for the kits by building type, and estimated lighting hours of use from SB by explicit building type or "Other" building type, as determined from the EY4 lighting metering study. The ex post savings estimates were compared to tracked savings estimates for the building in the sample and used to calculate an overall realization rate.

- **Tracking Database Augmentation and Upgrades.** The Evaluation Team continued to work with all utilities and implementers in EY4 to upgrade their tracking databases and add additional data fields that will help support future evaluation efforts. In past years, the Evaluation Team was limited in the parameter updates that could be performed on the entire tracking system, due to the multiple and different data systems used by the utilities. Currently, many of the key details and parameters are stored in calculators (Excel workbooks), and often the savings results output by the calculators are aggregated values. For example, the savings for *multiple* lighting fixtures with different configurations might be aggregated and reported as a *single* utility measure, and a *single* record in the tracking data. As an additional complication, customers often input measure and usage data into various vintages of calculators with differing parameters (from the Mid-Atlantic TRM, Maryland Evaluation Report, NEEP, or other sources), so the Evaluation Team could not assume that all projects used the same parameter assumptions within a particular program year. For example, if measure savings are calculated using a TRM algorithm, the parameters in the algorithm will include measure-specific inputs and/or stipulated values. Under the *parameter update* approach, the stipulated values from the TRM could be updated with new values determined from the evaluation effort, which were then used to recalculate savings for all projects in the automated and manual tracking data extraction samples. This effort is important because key parameters, such as those that are currently only available in calculators, could be made more accessible to evaluators, reduce the cost of verifying savings calculations, and simplify the savings reporting process. These improvements are also expected to lead to a more standardized tracking data format across utilities.
- **Installation Verification Study.** The Evaluation Team conducted verification activities that covered all measures for all utilities. The field technicians evaluated the quantity of functioning installed equipment, and collected detailed/contextual data needed to refine

estimates of savings such as equipment size and efficiency, building type, space type, HVAC equipment type, application, and control type. The Evaluation Team designed samples to meet a target of 90/20 one-tailed confidence and precision for program savings estimates each utility. All sites that implemented projects during the evaluation period were included in the sample frame.

- **VFD Study.** VFDs represent approximately 11% of the statewide reported C&IP programs energy savings. The Evaluation Team intended to leverage the NEEP[7] VSD Load Shapes Research Study, which includes measurement and verification of HVAC equipment usage controlled by VFDs. The study was not completed in time for use by the EY4 evaluation. The Evaluation Team also received preliminary VFD data from NEEP; however, those data could also not be used without further refinement for the EY4 impact evaluation. Instead, the Evaluation Team used an interim approach to estimate savings by making adjustments to key parameters based on a detailed review of a sample of implementer or utility calculators used to estimate VFD savings and verification site visits.
- **NTG or Free Ridership Studies -** The NTG ratio values from prior efforts were carried over from PY2012 and applied to the EY4 results for the C&IP and the SB programs. The NTG ratio values used were 0.72 for the C&IP and 0.74 for SB. A new NTG ratio evaluation effort was conducted for the PE SB EE Kits, and a value of 0.49 was determined.

## 9.4 Verification Approach

Itron verification focused on the key sampling elements and assumptions used by the Evaluation Team to produce the energy and peak savings estimates for these programs. As with previous years' verification efforts, it was a cooperative and complimentary effort between the Evaluation Team and Itron for both the Evaluation and Verification Reports. Itron's verification review of the C&IP and SB programs' EY4 evaluation consisted of the following steps:

- Review of the Evaluation Plan in May and June of 2013.
- Review and comment on the Draft Evaluation Reports, which identified the potential need for data requests to review and or verify key calculation steps.
- Meetings and conference calls with the Evaluation Team to discuss evaluation issues revealed in our review.
- Data requests to spot check or verify calculations used to estimate savings at the site level and/or extrapolate the savings from the sample to the general population.

[7] Northeast Energy Efficiency Partnerships, *Study RE12-1: Load Shape Research - VFD Study*.

- Identification of strategies to improve evaluation approaches in next year and potential program design improvements.
- Provide the Evaluation Team with a draft of the verification analysis and hold several calls to further discuss questions and issues.
- Produce a final verification analysis that incorporates the Evaluation Team comments.

We describe the key findings and analysis performed in each step that were used to ultimately verify savings, identify some issues, and recommend changes in the evaluation approach.

## 9.5 Verification Findings

Below we review the adequacy of the analysis methods and data collection methods used to estimate savings for these programs.

### *9.5.1 Review of the Lighting Metering Study*

Itron conducted a detailed review and analysis of the Metering Study results, which included several calls with the Evaluation Team, and the review of supplemental data sets. The Lighting Metering Study used the previous year's approach and the general approach was sound.

However, Itron did identify several issues that should be explored and addressed if warranted in future years. As shown in Table 61 of the Evaluation Report (shown below in Table 9-3), the first issue is that the targeted sample size of 66 sites was not met (only 62 sites were metered); however, the target sample did include a 10% contingency for logger failure. The most significant issue is the building type distribution of the Actual sites visited versus the Targeted sites. For example, the table shows a large shift in sites from the 2013 targeted, *explicit* building types (Retail, School, Health, Grocery) to the Other building type and to *non-targeted* building types (Warehouse/Industrial and Office). This change in sample distribution, along with comments in the Evaluation Report,[8] indicate that the building types entered into the utility tracking data are often different (in this case almost half the time) than the actual activity type found during the phone or on-site visit.

[8] This passage from the Evaluation Report is the most revealing of the issue: "The discrepancies between the target sample sizes and actual sample sizes arose because the building type listed in the tracking database, which was used to draw the target sample, sometimes differed from the actual building type found when conducting the site visit. For example, a site may have been categorized as an "Other" building type and drawn as one of the sites in the "Other" sample, but the site visit revealed that it was a retail store, and therefore it was categorized as "Retail" in the analysis."

**Table 9-3: Metering Sample Sizes (Sites Completed)**

| Building Type | Sample Size from Previous Years | Target Sample Size in 2013 | Actual Sample Size in 2013 | Total Sample Size |
|---|---|---|---|---|
| Warehouse/Industrial | 35 | 0 | 3 | 38 |
| Office | 23 | 0 | 2 | 25 |
| Retail | 18 | 5 | 3 | 21 |
| School | 0 | 15 | 10 | 10 |
| Health | 0 | 15 | 7 | 7 |
| Grocery | 7 | 5 | 3 | 10 |
| Other* | 27 | 26 | 34 | 61 |
| **Total** | **110** | **66** | **62** | **172** |

* The "Other" category refers to all projects that do not fall in the first six explicit building type categories, including projects that have no specified building type.

Source: *EmPOWER Maryland Final Impact Eval Report EY4: C&IP & SB Programs*, page 83, Table 61.

Incorrectly assigned building types in the tracking data could be an issue for the evaluation effort if lighting study metering results are mapped back to the tracking data records; for example, as part of the parameter update process. It is definitely an issue for utility-reported/ex ante lighting savings because building type is used to obtain hours of operation and coincidence factor. Furthermore, in reviewing the tracking data provided by the evaluation team, Itron also found that the building type definitions used by the utilities are not consistent, are sometimes missing, and are building types such as "Multi-family" that are not even included in the list of commercial building types contained in the Mid-Atlantic TRM. Given these issues, the Evaluation Team should look into why entries in the tracking system are often different than the actual building types discovered during the site visits, and also why business type is missing for some records. One possible reason for this systematic error could be that the NAICS code for the corporate business is used to classify the building type rather than the actual function/activity found at the physical site. For example, a "Retail" building type might be used for a large distribution warehouse for a retail chain store. Some related options to also consider are:

- Standardize the list of building types used across all efficiency measures, wherever it makes sense. The Evaluation Team should also examine the utility tracking data and identify business types that are unique to Maryland (if there are any), then provide guidance on how to map some of these *utility* business types to *TRM* and/or Evaluation Report business types. In addition, relabeling these categories as business types or business segment instead of *building* type might also help obtain better initial values. This change should be considered for future program years, not for the 2013 evaluation.
- Consider providing savings input factors for an "Unknown" building type that provides a conservatively low savings estimate (e.g. lowest hours of operation, CF, CSRR, etc.). These values would be used for the evaluation when the building type in the tracking data

is missing, and it would hopefully encourage the use of an explicit building type rather than leaving the field blank.

There are additional building type issues: The "Other" building type for the Evaluation Report has a different definition than the "Other" building type in the NEEP Lighting Load Shape Study and the TRM, so the values from the metering study are not comparable with the NEEP and TRM values. The Evaluation Team also uses a composite "Warehouse/Industrial" building type that *combines* two building types that are typically separate in other references. Another issue that was mentioned during discussion with the Evaluation Team was that of lighting in parking garages. The hours of operation from the metering studies should ideally only include indoor lighting, but some metered sites included lighting for parking garages (with 8760 operation). There were only five parking garage sites in the metering study, and any fixtures that were located on the roof of the structure were treated as *outside* lighting. However, the lighting for the other floors of the parking structure was treated as *inside* lighting. The Evaluation Team should consider splitting the Warehouse/Industrial building type for consistency with the TRM, and consider treating all lighting in parking garages as outside lighting or creating a Parking Garage building type category.

The Evaluation Team has only developed lighting parameters for five of the 14 explicit (i.e., not "Other") business types used in the TRM, and interactive HVAC factors for only four business types. A large lighting logger study to develop hours of operation for more building types should be considered. Given the size of the lighting logger sample used to estimate hours of use for the Evaluation Team's "Other" building type, it seems probable that lighting evaluation parameters could have been developed for at least one other explicit building type, but the Evaluation Team explored this issue and it was not possible.

Another issue to consider is the phase-out of T12 and 700-series T8 linear fluorescents. T12s can no longer be manufactured as of July 2012, and in July 2014 the same new standard or requirement applies to 700-series T8s. This issue was not mentioned in the Evaluation Report, but it has already been addressed in the Mid-Atlantic TRM. The TRM shows T12s being disallowed as a valid baseline system and being replaced by an 800-series T8 system in 2017.[9] What this means is that in 2017, T12 conversions will still be allowed, but the *savings assumed for those systems* must reflect the new 800-series T8 lamp and appropriate ballast. This change has been implemented in other states by making modifications to the Standard Fixture Wattage table that is incorporated into utility savings calculators—the fixture wattage for T12 systems would be revised to reflect an 800-series T8 system instead of the T12 system. Although 2017 is still several years away, the Evaluation Team should start considering the changes that will be needed in utility calculators and advise the utilities on those changes.

9 Shelter Analytics, *Mid-Atlantic Technical Resource Manual, Version 3*, prepared for NEEP, February 2013, page 205.

One more possible issue discovered during our review is the development of the CSRR. CSRRs are developed from lighting logger data and customer-reported business hours. However, the utilities use at least three different methods to report lighting hours of operation: 1) a single annual hours estimate, 2) an average weekly hours estimate (which gets expanded to annual hours), and 3) a three-parameter method—hours per day, the number of days per week of that day type per week, and the number of weekdays per year that the lights are off. Only the annual hours are used to estimate the CSRR, but the annual hour values developed from each of these three approaches could be quite different. However, the evaluation approach assumes that these values are all essentially the same. Itron recommends that the Evaluation Team address this issue in the 2014 evaluation by examining the variation in the annual hours of operation by business type for the various methods and across utilities. If there is a significant difference in the estimated annual hours from each method, then the use of a standard approach across all utilities should be encouraged.

### *9.5.2 Review of the PE Small Business Energy Efficiency Kits Evaluation*

Hours used to estimate savings for the CFLs in the PE EE Kits were likely too high because the Evaluation Team used metering study results that were not CFL-specific. The Metering Study results likely represent non-CFL lighting, as the predominant area lighting type is linear fluorescent lighting. The assumption that business hours are likely to be the same as the hours of operation for lighting fixtures is often not true for CFL lamps, which are used for task-specific applications and, as such, are not on for the duration of business hours.

This issue is illustrated in Table 51 (shown below in Table 9-4) from the Evaluation Report. The metered HOU from the Evaluated Data is compared to the TRM values, and the presented TRM values are hours of operation for CFLs. As shown, the Evaluated Data values are typically slightly higher than the TRM V3 CFL values, which could lead to systematically overestimating the savings from the CFLs installed by these programs. However, 60 to 70% of the kWh savings were for "Other" building types, where the Evaluated and TRM V3 hours of operation are within 3% of each other. The hours for almost all building types except Grocery are actually very close.

### Table 9-4: Business Type Hours Comparison

| Business Type (Hours of Operation) | Evaluated Data | TRM V3 | TRM V2 |
|---|---|---|---|
| Grocery | 7134 | 5010 | 3879 |
| Health | 3909 | 2849 | 1888 |
| Office | 2950 | 3516 | 2478 |
| Retail | 4926 | 4413 | 3043 |
| School | 2575 | 2513 | 1670 |
| Warehouse/Industrial | 3799 | 3571 | 2063 |
| Other | 4573 | 4444 | 1871 |

Sources: Mid-Atlantic TRM version 2 & version 3 and Navigant, *Lighting Metering Study*; *EmPOWER Maryland Final Impact Eval Report EY4: C&IP & SB Programs* page 69, Table 51.

Another issue relates to the evaluation of savings from the PE SB program that was not covered by EE Kits. It is not very clear in the Evaluation Report that SB direct install projects (i.e. everything other than the EE Kits) were grouped and evaluated with the PE *Prescriptive program*. The Evaluation Team should have clearly called out how the savings from these measures were evaluated by adding a footnote to explain this situation to Table 20 (shown below as Table 9-5) in the Evaluation Report. However, the savings for this element of the program are very small.

### Table 9-5: PE 2012-2013 Evaluation Year 4 Ex Ante Tracked and Ex Post Evaluated Gross Annual Savings—Small Business Programs

| | Ex Ante Tracked Gross Savings* | Ex Post Evaluated Gross Savings | Gross Realized Savings Ratio** |
|---|---|---|---|
| **PE Small Business** | | | |
| PJM Coincident Peak Demand Savings (kW) ** | 6 | 5 | 0.84 |
| Utility Coincident Peak Demand Savings (kW) | 6 | 6 | 0.94 |
| Annual Energy Savings (kWh) | 29,920 | 29,496 | 0.99 |
| **PE Small Business Kits** | | | |
| PJM Coincident Peak Demand Savings (kW) ** | 1792 | 1713 | 0.96 |
| Utility Coincident Peak Demand Savings (kW) | 1800 | 1755 | 0.98 |
| Annual Energy Savings (kWh) | 8,863,598 | 8,607,614 | 0.97 |

* Reported savings reflects program tracking database values.

** PJM demand savings excludes all non-lighting, schools lighting, outdoor lighting, and lighting controls.

Source: *EmPOWER Maryland Final Impact Eval Report EY4: C&IP & SB Programs*, page 31, Table 20.

The PE SB EE Kits analysis in the Evaluation Report could more clearly show *all* of the kit configurations distributed to customers, including the type and quantity of each item in each kit.

Table 50 of the Evaluation Report only presents four of the kits but states that larger businesses received *double the quantities* distributed to smaller customers. A more complete version of the Evaluation Report table is provided below.

**Table 9-6: PE Small Business Kits Contents**

| Small Business Customer Size | Number of 20 Watt CFLs | Number of 26 Watt CFLs | Number of Smart Strips | Number of Faucet Aerators | Number of Aerator Adaptors | Total Items in Kit |
|---|---|---|---|---|---|---|
| Small Non-Electric (SNE) | 6 | | 1 | | | 7 |
| Large Non-Electric (LNE) | 5 | 5 | 1 | | | 11 |
| Small Electric (SE) | 6 | | 1 | 4 | 4 | 15 |
| Large Electric (LE) | 5 | 5 | 1 | 4 | 4 | 19 |
| Non-Electric Double (2x SNE) | 12 | | 2 | | | 14 |
| Electric Double (2x SE) | 12 | | 2 | 8 | 8 | 30 |

Sources: Program Implementer and Tracking Database; *EmPOWER Maryland Final Impact Eval Report EY4: C&IP & SB Programs,* page 67, Table 50.

### *9.5.3 Review of the VFD Parameter Update*

Itron conducted a detailed review of Appendix E.2 of the Evaluation Report. The VFD analysis was very detailed and sound, in spite of not being able to use the results of the NEEP VFD Study. Itron completely concurs with the Evaluation Team's assessment that this analysis of this measure should be a high-priority target for improvement in 2014. In addition, the Evaluation Team should consider the use of Advanced Metering Infrastructure (AMI) interval metered data for the evaluation of these measures, especially for VFDs associated with HVAC systems.

### *9.5.4 Review of the Database and Tracking System Upgrades*

The Evaluation Team has done an excellent job in pursuing this objective, and it appears that many improvements have been made or are underway. They were even able to use sampling combined with a manual calculator review process to take care of some unexpected issues with savings estimates that came from calculators due to custom modifications. Given the preliminary steps taken, and that the amount of effort (and cost) for some of the steps in the process that could only be guessed at previously are now known, 2014 might be a good time to consider if this effort should be continued. Given the issues encountered so far, this may not be the best route for achieving a faster, more effective/efficient and more accurate evaluation and cost-effectiveness analysis.

### *9.5.5 Review of the Installation Verification Study*

The verification effort for non-lighting measures included the typical range of activities, namely phone surveys, on-site surveys, and desk reviews of algorithms and assumptions. Because this

effort was essentially the same as in previous years, Itron performed only a quick review of this effort, and no significant issues were found. However, the Evaluation Team should consider adopting as standard practice the use of electric (and possibly gas) consumption data from AMI smart meters for evaluation project sample sites. The whole-building interval metered load data could provide a better sanity check on savings estimates. These data could be used to verify the assumed general facility and equipment hours of operation, and weather-sensitive cooling/heating savings at the site level. If it is considered too costly to use these data for all sample sites, then a pilot test should be conducted on a smaller set of sites to work out the process and assess the evaluation value of this additional data.

### *9.5.6 Review of the Net to Gross Ratio Estimates*

For the PE SB EE Kits program, Itron reviewed the NTG ratio phone survey form, data tables, and survey responses, which were included in an appendix of the Evaluation Report. No significant issues were found with this analysis. The NTG ratio values used for other programs were carryovers from previous evaluations and the data were gathered in late 2011, so Itron recommends that the NTG ratio values be re-evaluated in 2014.

## 9.6 Final Verified Gross and Net Impacts

Table 9-7 and Table 9-8 show the final evaluated and Itron-verified EY4 gross and net impacts for the C&IP) and SB programs. Because no adjustments were made by the Itron Verification Team, the evaluated and verified values are the same.

**Table 9-7: Summary of Evaluation Year 4 Evaluated & Verified Gross and Net Impacts for the C&I Prescriptive Programs**

| | Gross Impacts | | | | Net Impacts | | | |
|---|---|---|---|---|---|---|---|---|
| | Evaluated | | Itron-Verified | | Evaluated | | Itron-Verified | |
| Utility | MWh | kW* | MWh | kW* | MWh | kW* | MWh | kW* |
| BGE | 81,185 | 13,088 | 81,185 | 13,088 | 58,453 | 9,423 | 58,453 | 9,423 |
| PEPCO | 32,423 | 6,017 | 32,423 | 6,017 | 23,345 | 4,332 | 23,345 | 4,332 |
| DPL | 6,885 | 1,036 | 6,885 | 1,036 | 4,957 | 746 | 4,957 | 746 |
| PE | 5,374 | 510 | 5,374 | 510 | 3,869 | 367 | 3,869 | 367 |
| SMECO | 3,449 | 788 | 3,449 | 788 | 2,483 | 567 | 2,483 | 567 |
| **Statewide** | **129,317** | **21,439** | **129,317** | **21,439** | **93,108** | **15,436** | **93,108** | **15,436** |

* Utility Coincident Peak Savings

Source: *EmPOWER Maryland Final Impact Eval Report EY4: C&IP & SB Programs*

**Table 9-8: Summary of Evaluation Year 4 Evaluated and Verified Gross and Net Impacts for Small Business Programs**

| | Gross Impacts | | | | Net Impacts | | | |
|---|---|---|---|---|---|---|---|---|
| | Evaluated | | Itron-Verified | | Evaluated | | Itron-Verified | |
| Utility | MWh | kW* | MWh | kW* | MWh | kW* | MWh | kW* |
| BGE | 24,743 | 5,276 | 24,743 | 5,276 | 18,310 | 3,904 | 18,310 | 3,904 |
| PEPCO | 9,963 | 2,595 | 9,963 | 2,595 | 7,373 | 1,920 | 7,373 | 1,920 |
| DPL | 3,817 | 658 | 3,817 | 658 | 2,825 | 487 | 2,825 | 487 |
| PE** | 29 | 6 | 29 | 6 | 21 | 4 | 21 | 4 |
| PE (EE Kits) | 8,608 | 1,755 | 8,608 | 1,755 | 4,218 | 860 | 4,218 | 860 |
| SMECO | 852 | 232 | 852 | 232 | 630 | 172 | 630 | 172 |
| **Statewide** | **48,012** | **10,522** | **48,012** | **10,522** | **33,377** | **7,348** | **33,377** | **7,348** |

* Utility Coincident Peak Savings

** The SB direct install (non-kit) projects were evaluated as part of the PE Prescriptive program evaluation due to its small size.

Source: *EmPOWER Maryland Final Impact Eval Report EY4: C&IP & SB Programs*

For EY4, savings for the PE SB programs was dominated by the EE Kits program component. Because the EE Kits were the predominant measure for the PE SB program, this element was evaluated separately from the main PE SB program. In addition, the much smaller PE SB direct install component was rolled into the PE Prescriptive evaluation.

## 9.7 Recommendations

Itron offers the following recommendations:

- Increased and more frequent coordination of the Evaluation and Itron Verification Teams. These meetings could include informal reviews of working data and systems, not formal presentations. This should enhance the already excellent working relationship between Itron and the Evaluation Team, and help to identify and resolve any issues a lot sooner, so they might even be addressed and communicated to the Program Administrator in real-time.
- Emphasize the need to use consistent and accurate building (business) types by for implementation and evaluation. For commercial sector programs, building (or business) type is one of the most critical parameters for most measures, yet the Evaluation Report and the underlying data (which were examined by Itron) illustrate there are many issues with getting accurate and complete building types. An accurate building type is one that reflects the actual business activity that occurs at the site.

- Determine why incorrect building (business) types are used for implementation and evaluation. The Evaluation Team should investigate further to determine why entries in the tracking system are often different than the actual building types discovered during the site visits, and also why business type is missing for some records. One possible reason for this systematic error could be that the NAICS code for the corporate business is used to classify the building type rather than the actual function/activity found at the physical site. Once the reasons are determined, then a solution can be assessed.
- Revisit VFDs in 2014: Consider incorporating NEEP Study when available and consider conducting M&V. The portion of C&IP program savings for VFDs is significant, yet the estimated savings for this measure appears to be quite uncertain, and in Maryland there are many inconsistencies between the methods used across utilities. The evaluation of this measure, and providing a consistent approach for estimating savings, should be made a priority in 2014.
- Develop lighting CSRR for each utility due to differences in customer-reported hours of operation. CSRRs are developed from lighting logger data and customer-reported annual hours of operation. As reported earlier, there are three different methods used to report and develop annual lighting hours of operation. However, the evaluation approach assumes that these values are all essentially the same. Itron recommends that the Evaluation Team address this issue next year by examining the variation in the annual hours of operation by business type for the different approaches used, which may show a need to develop CSRRs versus the self-report method used. Another option is to encourage the use of a single approach across all utilities.
- Revisit and update the NTG values. The values currently being used for C&IP and SB programs (except the EE Kits) are now several years old (surveys were fielded in 2011). As most of the programs are well established now, Itron recommends that the NTG values be re-evaluated in 2014.
- Lighting early retirement, T12/T8 linear fluorescent baseline issue. T12s can no longer be manufactured as of July 2012, and in July 2014 the same new standard or requirement applies to first generation T8s. The TRM provides a method for phasing these lamp types out as a valid baseline system by 2017.[10] The utilities and the Evaluation Team should start working to ensure that these changes are reflected in the utility calculators and the ex post savings calculations by 2017.
- The Evaluation Team should consider using AMI/smart meter data for evaluation. This is a carryover suggestion made previously in 2012. The Evaluation Team should consider adopting as standard practice the use of electric (and possibly gas) consumption data from AMI smart meters for evaluation project sample sites. The whole-building interval metered load data could provide a better sanity check on savings estimates,

[10] *Mid-Atlantic Technical Resource Manual, Version 3*, page 205.

business hours, equipment hours of operation, and weather-sensitive cooling/heating savings at the site level. As a minimum, a pilot test should be conducted on a smaller set of sites to work out the process and assess the evaluation value of this additional data.

- PE SB EE Kits should use TRM Hours of Operation for CFLs from the TRM. If EE Kits are used again in the future, the CFL savings estimates for PE EE Kits may need to be revised to use HOU found in the Mid-Atlantic TRM. The hours and factors developed from the Evaluation Team metering study were used. Those results are likely biased to represent non-CFL lighting, which typically has higher hours of operation than CFL lighting. However, for the building types represented in the EE Kits program, the metered hours of operation are only slightly higher than the TRM values, so the overestimated savings are likely small. Furthermore, the distribution of EE Kits was discontinued at the end of 2012.

# 10

# C&I Custom and Retrocommissioning Programs

## 10.1 Verification Summary

Table 10-1 and Table 10-2 present a high-level comparison of the evaluated and verified gross savings impacts for all of the EmPOWER utilities that operated C&I Custom and RCx programs for EY4 (June 1, 2012 through May 31, 2013). Table 10-1 focuses on the absolute level of energy and peak savings estimated for the large C&I Custom and RCx programs in each service area by the Evaluation Team and the verified savings resulting from this review by Itron. Table 10-1 also presents high and low estimates of likely program savings calculated using the Evaluation Team's one-tailed 90/10 confidence and precision estimates.

**Table 10-1: Summary of Evaluated & Verified Gross Impacts—C&I Custom & RCx Programs for Evaluation Year 4**

| | Annual Gross Impacts | | | |
|---|---|---|---|---|
| | Evaluated* | | Itron-Verified | |
| Utility | MWh | kW** | MWh | kW** |
| BGE *** | 33,434 | 3,217 | 33,434 | 3,217 |
| PEPCO *** | 22,136 | 2,638 | 22,136 | 2,638 |
| DPL | 4,460 | 667 | 4,460 | 667 |
| SMECO | 1,074 | 85 | 1,074 | 85 |
| PE | 10,561 | 1,503 | 10,561 | 1,503 |
| **Total Statewide** | **71,665** | **8,110** | **71,665** | **8,110** |

* Source: *EmPOWER 2013 Draft Final Evaluation Report*, March 28, 2014

** Utility Coincident Peak Demand Savings

*** C&I Custom includes RCx

Itron performed an in-depth review for 12 out of the 80 sites within the Evaluation Team's telephone-supported engineering review and on-site M&V sample, and identified one site that needed a revision to the evaluation engineering calculations. These sites were selected for review for one of two reasons; either each site represented a large amount of savings or the calculation methods and data inputs for a specific site warranted further review. The Evaluation Team adjusted the ex post energy and peak demand savings for this site in response to issue found by Itron. The impact of this adjustment was minor, and did not affect the overall statewide

GRSR. These adjusted savings numbers were used by the Evaluation Team in their Final Evaluation Report. Therefore, both the overall verified and evaluated GRSR and gross impacts are in agreement at the statewide level.

Table 10-2 provides the evaluated and verified RRs and NTG ratios for each program. Note that the Evaluation Team did not conduct NTG analysis this year; therefore, statewide NTG ratios reported in the previous year were used for all five utility programs (see statewide NTG ratios in weighted average row in Table 10-2 below).

### Table 10-2: Summary of Evaluation Year Gross Realization Rates and NTG Findings

| | Realization Rates | | | | Net-to-Gross Ratios | | | |
|---|---|---|---|---|---|---|---|---|
| | Evaluated* | | Itron-Verified | | Evaluated* | | Itron-Verified | |
| **Utility** | **MWh** | **kW**** | **MWh** | **kW**** | **MWh** | **kW**** | **MWh** | **kW**** |
| BGE*** | 0.62 | 0.31 | 0.60 | 0.31 | -- | -- | -- | -- |
| PEPCO*** | 1.11 | 1.03 | 1.11 | 1.03 | -- | -- | -- | -- |
| DPL*** | 1.11 | 1.03 | 1.11 | 1.03 | -- | -- | -- | -- |
| SMECO | 0.83 | 1.28 | 0.83 | 1.28 | -- | -- | -- | -- |
| PE | 0.95 | 0.87 | 0.95 | 0.87 | -- | -- | -- | -- |
| **Wtd Average Statewide** | **0.79** | **0.53** | **0.79** | **0.53** | **0.69** | **0.61** | **0.69** | **0.61** |

* Source: *EmPOWER 2013 Draft Final Evaluation Report*, March 28, 2014

** Utility Coincident Peak Demand Savings

*** C&I Custom includes RCx

## 10.2 Overview of the 2013 Custom Programs

The Custom programs implemented by the EmPOWER utilities support large C&I customers by identifying and implementing site-specific and unique cost-effective energy efficiency opportunities through measures not addressed by the prescriptive rebate programs. These programs offer customized cash incentives for more complex and site-specific measures and projects, ranging from complex commercial HVAC projects to industrial process improvements. Custom projects must be able to show specific and verifiable energy savings and costs, based on a walk through audit and subsequent analysis typically developed by a third-party firm. Any measure that improves the electric energy efficiency of a customer's building or facility is eligible provided that it is cost-effective.

The RCx programs implemented by BGE and Pepco Holdings, Inc. (PHI–PEPCO and DPL) are designed to help customers identify and implement low-cost tune-ups and adjustments that improve the efficiency of existing building operating systems by returning them to their intended

operation or design specification. These programs focus on promotion and installation of building control systems used to fine-tune the performance of HVAC systems.

The C&I Custom programs comprised 9% of the CY 2013 EmPOWER Maryland MWh savings and 5% of the MW savings.

## 10.3 Evaluation Summary

The Evaluation Team designed the sampling plan to achieve one-tailed 90/20 confidence interval and precision at the utility level. The Evaluation Team conducted Telephone-Supported Engineering Reviews (TSERs) for sampled projects at all utilities and on-site M&V in addition to TSERs for BGE and PHI programs.

Based on comparison between utility-reported savings for the sample of sites and the Evaluation Team's estimates of savings from the same sites, a sample GRSR was calculated. The EY sample gross realized savings was extrapolated to the utility-reported savings for the entire program population using a ratio estimation method (double ratio estimation method for BGE and PHI and a standard ratio method for SMECO and PE) to calculate the overall (population) savings for the C&I Custom programs.

## 10.4 Verification Approach

Itron designed its verification process to focus on the key sampling elements and assumptions used by the Evaluation Team to produce the energy and peak savings estimates for this program. Our verification approach to confirm evaluated savings from the 2013 programs was very similar to the approach used for the 2011 and 2012 programs. The Evaluation Team provided all the supporting data and documentation that Itron requested for a selected sample of projects necessary to facilitate the savings verification process. Our review of the C&I Custom programs included the following key elements.

- **Evaluation Plan Review:** Itron reviewed the evaluation plan provided by the Evaluation Team for the Custom programs and determined that it conformed to the proposed evaluation sampling plan and the (gross and net) impact evaluation methodology.
- **Evaluation Report Review:** Itron reviewed the Draft Evaluation Report and provided comments to the Evaluation Team. We discussed the comments and changes that needed to be made in the Final Evaluation Report with the Evaluation Team.
- **Sample Design**: Itron reviewed the sampling plan and confirmed that the sample design and selection met the confidence and precision levels prescribed in the Evaluation Plan.

- **Engineering Review:** Itron conducted two levels of reviews: in-depth reviews and basic reviews. The in-depth reviews consisted of verifying calculation models and all supporting documentation for the sites with significant variation in the energy and demand savings. A total of 12 sites were selected for the in-depth reviews. Additionally, Itron conducted basic reviews of the data collected and analysis conducted for all 45 BGE sites. For basic reviews, Itron reviewed the on-site and TSER reports provided by the Evaluation Team.
- **Verified Sample Savings and GRSR:** Itron noted issues that were identified for the selected sites and discussed them with the Evaluation Team. Based on these discussions, the Evaluation Team recalculated savings for selected sites based on Itron's findings.
- **Program Savings and GRSR:** Itron verified the ratio estimation method used to extrapolate the energy and demand sample GRSR of 78% and 53%, respectively, to the entire population of utility reported savings, to estimate the overall savings from the C&I Custom programs.
- **Precipitous Year to Year Changes in Gross Realization Rates:** As shown in Table 10-3, BGE, which accounts for about 55% of the statewide savings, had a significant drop in their energy and demand RRs relative to the rates reported in the 2012 evaluation. A major priority of the Itron verification was to identify and understand factors that may have led to this drop in RRs and providing recommendations to improve program tracking and performance.

### Table 10-3: Comparison of Realization Rates between 2012 & 2013 Custom Programs Evaluation Results

| | Gross Realization Rates | | | |
|---|---|---|---|---|
| | 2013 Evaluated* | | 2012 Evaluated** | |
| **Utility** | **Energy** | **Demand** | **Energy** | **Demand** |
| BGE | 0.60 | 0.31 | 0.87 | 1.42 |
| PEPCO | 1.11 | 1.03 | 0.64 | 1.24 |
| DPL | 1.11 | 1.03 | 0.64 | 1.50 |
| SMECO | 0.83 | 1.28 | 0.40 | 0.65 |
| PE | 0.95 | 0.87 | 0.99 | 0.84 |
| **Wtd Average Statewide** | **0.78** | **0.53** | **0.70** | **0.88** |

Note: Demand RR is the Utility Coincident Peak Demand. These values do not include the PEPCO O&M Training program, but do include the RCx programs

* Source: *EmPOWER 2013 Draft Evaluation Report*, March 28, 2014

** Source: EmPOWER Impact Report Overview_25Mar2013_Final

The Evaluation Team communicated with Itron throughout the program year to discuss site-specific issues as needed. Itron and the Evaluation Team discussed complex calculation issues

(e.g., regression-based estimates of site savings using billing data) and baseline selection issues (e.g., fuel switching sites) and determined the best method to estimate gross savings. In addition, the Evaluation Team organized a call at the end of the EY to discuss all the sites with potential issues. In general, this process gave Itron and the Evaluation Team an opportunity to work closely to resolve any site-specific issues prior to Itron's verification. As a result, Itron's verification process was more efficient for everyone involved and focused on identifying opportunities to improve program performance.

## 10.5 Verification Findings

Below we review the adequacy of the analysis methods and data collection methods used to estimate savings for these programs. The key verification steps are listed below:

- **Adequacy of Baseline Operating Conditions and Energy Usage Analysis:** Verification of baseline selection methodology to ensure it is based on best available data, consistent with standard practice, and consistent across projects and across reviewers, for the purposes of estimating gross savings.
- **Adequacy of On-Site Data Collection Effort and Phone Survey Instruments to Verify Measure Installations:** Itron reviewed the results of the phone survey and the data collected during the on-site surveys for the 12 detailed site evaluations to confirm the sufficiency of the data collection effort.
- **Adequacy of Engineering Reviews:** Review of the engineering methods used for estimating gross savings for selected sites.
- **Accuracy of Energy Savings Calculations and Specific Site Issues:** Adjustments to the evaluated energy savings calculations based on Itron's findings and discussion of sites identified with evaluation gross savings estimates.
- **Review of Net Savings Estimates:** Verifying the NTG analysis and reporting issues identified with the net savings estimates.

The details and findings from each of these verification steps are discussed in the following subsections.

### *10.5.1 Adequacy of Baseline Conditions Description and Estimated Energy Usage*

Itron reviewed a sample of sites to verify the selection methodology used to estimate baseline usage for the purposes of estimating gross savings. From the sample of sites reviewed, we found that the baseline adjustments made by the Evaluation Team were the main reason for BGE's low RR. The Evaluation Team adjusted the baseline reported by the Program Implementer for several retrofit and new construction projects. For retrofit projects, the ER baseline claim was

adjusted to a Replace-On-Burnout (ROB) claim or vice versa. The Evaluation Team has since provided specific guidelines to utilities for what project type (ROB or ER) various measures should fall under.[1] For New Construction projects, the code year for the applicable commercial buildings code was adjusted to match prevailing code when the facility obtained the permit date. Itron discussed these adjustments with the Evaluation Team and agreed that the use of applicable based on permit code is an acceptable approach for New Construction projects.

### *10.5.2 Adequacy of On-site Data Collection Effort and Phone Survey Instruments*

Itron reviewed a sample of sites to understand the evidence and rationale used by the Evaluation Team to select the baseline system characteristics. These characteristics are used to estimate baseline usage for the purposes of estimating gross savings. From the sample of sites reviewed, we conclude that the Evaluation Team method for verifying the installed measures and collecting baseline and post-usage data conditions for calculating energy savings is consistent with industry standard practices.

### *10.5.3 Adequacy of Engineering Reviews*

Itron reviewed the both the engineering methods or equations used to calculate savings for the sites visited by the Evaluation Team in the BGE territory, and the TSERs completed by the Evaluation Team. The Evaluation Team provided site reports to facilitate this review. Itron performed in-depth reviews of the assumptions used to estimate savings for 12 sites (combination of on-sites and TSERs), selected because either the site represented a large amount of savings, or the calculation methods and data inputs for the site needed further review. Our review included assessment of baseline selection methods, data collection and management procedures, consistency and accuracy of algorithms and values used in calculations, and documentation for each of these sites. The Evaluation Team used sound engineering analysis for estimating gross impacts (energy and peak savings) for custom energy efficiency projects included in the program.

One of the biggest barriers for the Evaluation Team was the lack of information provided by the implementers to support the accuracy of their ex ante savings calculations for each site. This lack of supporting information often resulted in the need for the Evaluation Team to create their own method of estimating savings based on best-available data, rather than allowing them to verify the implementer's calculation method. This practice often resulted in differences in savings assumptions and methodologies and, ultimately, differences between the reported savings results and the evaluated savings results. In Subsection 10.7, we offer recommendations to help reduce this gap between reported and evaluated savings.

[1] These guidelines have been specified below in the Recommendations subsection.

### *10.5.4 Accuracy of Energy and Peak Savings Calculations and Specific Site Issues*

Itron's site-specific reviews identified potential issues in the estimation of gross savings for seven projects. These issues included the methods of baseline selection process and calculation and modeling inputs used by the Evaluation Team. Itron and the Evaluation Team discussed these seven sites and determined that only one site required engineering calculations adjustments. Itron agreed with the responses provided by the Evaluation Team for the six projects to support their savings estimates and determined that no changes were needed to these site specific analyses. Itron requested that the Evaluation Team include the baseline issue as one of the top drivers for their Findings and Recommendations section in the Evaluation Report. According to the Evaluation Team, differences in the selection of baseline condition for estimating savings between the implementers and the Evaluation Team was found to be a pattern across the sites in each utilities program; therefore, any recommendation to improve these practices would apply to all utilities.

Based on Itron's detailed review, site-specific engineering calculations were adjusted for one project (CU-10-0944).

- **CU-10-0944:** Itron recommended that the Evaluation Team adjust the occupancy rate for the building, which was originally calculated using 100% occupancy because evidence collected during the onsite visit found an occupancy rate of 66%. Due to a lack of available information about this site that could be feasibly collected within the prescribed evaluation budget, the Evaluation Team agreed that 100% occupancy should not be used. The Evaluation Team researched standard occupancy rates for the office building market in that county, and came up with a rate of 70.9%, which was then used to recalculate savings.

The impact on the GRSR was a reduction of 14% for energy and 3% for demand for the site. The Evaluation Team made these adjustments to the impacts reported in the Final Evaluation Report.

Table 10-4 and Table 10-5 show verified energy and demand savings for site CU-10-0944, along with the site-level GRSR (i.e., the ratio of verified savings to utility-reported savings).

**Table 10-4: Summary of Site-Specific Adjusted Annual Energy Savings***

| Utility | Project ID | Evaluation Review Plan | Tracking Energy Savings (kWh) | Evaluated Energy Savings (kWh) | Evaluated Gross Realized Savings Ratio | Verified Energy Savings (kWh) | Verified Gross Realized Savings Ratio |
|---|---|---|---|---|---|---|---|
| BGE | CU-10-944 | On-site | 324,104 | 151,094 | 0.47 | 107,600 | 0.33 |

* The Evaluation Team adjusted the evaluation savings in response to issues found by Itron for this site. These adjusted savings numbers were used by the Evaluation Team in their Final Evaluation Report.

**Table 10-5: Summary of Site-Specific Adjusted Peak Savings***

| Utility | Project ID | Evaluation Review Plan | Tracking Demand Savings (kW) | Evaluated Demand Savings (kW) | Evaluated Gross Realized Savings Ratio | Verified Demand Savings (kW) | Verified Gross Realized Savings Ratio |
|---|---|---|---|---|---|---|---|
| BGE | CU-10-944 | On-site | 517.5 | 112.18 | 0.22 | 98.9 | 0.19 |

* The Evaluation Team adjusted the evaluation savings in response to issues found by Itron for this site. These adjusted savings numbers were used by the Evaluation Team in their Final Evaluation Report.

### *10.5.5 Review of Net Savings Estimates*

An NTG analysis is performed by the Evaluation Team every other year. Therefore, the NTG analysis from last year was used to estimate the net savings for this year. Statewide NTG ratios, rather than utility-specific ratios, were produced last year due to a lack of utility-specific responses to the survey instruments fielded in 2012. The statewide Utility Coincident Peak Demand Savings NTG ratio was 0.61, and the Annual Energy Savings NTG ratio was 0.69. Itron recommends the Evaluation Team estimate NTG ratios at the utility level, rather than the statewide level, as part of the 2014 evaluation. This will provide individual Program Administrators better guidance and the ability to adjust program design practices to address any free ridership issues.

## 10.6 Final Verified Gross Savings Estimates

Table 10-6 shows the final verified statewide gross savings and GRSR. Based on the changes to the RR of the one Custom program site discussed above, the verified statewide GRSR and energy and demand savings are in agreement with the corresponding evaluated GRSR and savings estimates.

**Table 10-6: Summary of C&I Custom Programs Evaluation Year Savings Verification**

| | Gross Realized Savings Ratio | | | | Gross Impacts | | | |
|---|---|---|---|---|---|---|---|---|
| | Evaluated* | | Itron-Verified** | | Evaluated* | | Itron-Verified** | |
| Utility | kWh | kW | kWh | kW | MWh | kW | MWh | kW |
| BGE Custom | 0.65 | 0.30 | 0.65 | 0.30 | 28,482 | 2,699 | 28,482 | 2,699 |
| BGE RCx | 0.43 | 0.41 | 0.43 | 0.41 | 4,392 | 456 | 4,392 | 456 |
| PEPCO Custom & RCx | 1.11 | 1.03 | 1.11 | 1.03 | 22,136 | 2,638 | 22,136 | 2,638 |
| DPL Custom & RCx | 1.11 | 1.03 | 1.11 | 1.03 | 4,460 | 667 | 4,460 | 667 |
| SMECO | 0.83 | 1.28 | 0.83 | 1.28 | 1,074 | 84.84 | 1,074 | 84.84 |
| PE | 0.95 | 0.87 | 0.95 | 0.87 | 10,561 | 1,503 | 10,561 | 1,503 |
| **Total** | **0.78** | **0.53** | **0.78** | **0.53** | **70,879** | **7,558** | **71,105** | **7997** |

* Source: *EmPOWER 2013 Draft Evaluation Report*, March 28, 2014

** The Evaluation Team adjusted the evaluation program savings in response to issues found by Itron. These adjusted program savings numbers were used by the Evaluation Team in their Final Evaluation Report.

## 10.7 Recommendations

Based on the findings and analysis above, Itron offers the following recommendations to improve the accuracy of future reported savings estimates and ensure progress is made toward reducing the observed differences in reported compared to evaluated savings estimates:

- The Evaluation Team should provide guidance to the utility implementation staff about the baseline selection process used by the Evaluation Team to ensure implementation team baseline selection methods become more consistent with the selection processes and ultimate baseline condition choices of the Evaluation Team.
  - For ER claims, the Evaluation Team should provide guidance on the best practices used to determine RUL period based on the information collected through site visits and interviews.
  - The Evaluation Team should inform the implementers that the Evaluation Team will only accept ER claims for those projects with a RUL of two years or greater.
  - For New Construction projects, the Evaluation Team should provide the process selecting the applicable code using the facility permit date.
- The Evaluation Team should develop estimates of NTG for these programs at the utility level next year. This will provide the Program Administrators with better guidance and the ability to adjust program design practices to address and even reduce the incidence of free riders.

- The Evaluation Team should ensure that the savings calculations for a given project are based on representative conditions found at the facility. When the "as found" site conditions are not able to be considered as valid representation of typical site conditions, the evaluation assumptions of typical site conditions should be based on well-grounded sources (e.g., customer agreements to confirm increase in occupancy or production rates, benchmarked industry averages for the measure/facility type, etc.).

# List of Acronyms

| | |
|---|---|
| AMI | Advanced Metering Infrastructure |
| ASHP | Air Source Heat Pump |
| BEopt | Building Energy Optimization |
| BGE | Baltimore Gas & Electric |
| C&I | Commercial and Industrial |
| CAC | Central Air Conditioner |
| CDF | Coincident Demand Factors |
| CF | Coincidence Factor |
| CSRR | Customer Self-Report Ratio |
| CY | Calendar Year |
| CY2013 | Calendar Year 2013 |
| DEER | Database for Energy Efficient Resources |
| DHW | Domestic Hot Water |
| DWM | Delta Watts Multipliers |
| DPL | Delmarva Power & Light |
| ECM | Electronically Commutated Motors |
| EE Kits | Energy Efficiency Kits |
| EER | Energy Efficiency Ratio |
| EERE | DOE Office of Energy Efficiency & Renewable Energy |
| EF | Energy Factor |
| EFLH | Equivalent Full Load Hours |
| EISA | Energy Information and Security Act |
| EPA | Environmental Protection Agency |
| ER | Early Retirement |
| EY4 | Evaluation Year 4 |
| GRSRs | Gross Realized Savings Ratios |
| GSHP | Ground Source Heat Pumps |
| HERS | Home Energy Rating System |
| HOU | Hours of Use |
| HP | Heat Pump |
| HPwES | Home Performance with Energy Star |
| HSPF | Heating Seasonal Performance Factor |
| IECC | International Energy Conservation Code |
| ISR | In-service Rate |
| MMMF | Master-metered Multi-family |
| NAECA | National Appliance Energy Conservation Act |

| | |
|---|---|
| NAICS | North American Industry Classification System |
| NEEP | Northeast Energy Efficiency Partnerships |
| NTG | Net-to-Gross |
| PE | Potomac Edison |
| PEPCO | Potomac Electric Power Company |
| PHI | Pepco Holdings, Inc. |
| PRM | Price Response Modeling |
| PSC | Public Service Commission |
| PST | Predictive Savings Tool |
| PY | Program Year |
| QHEC | Quick Home Energy Checkup |
| RCx | Retrocommissioning |
| RNC | Residential New Construction |
| ROB | Replace-On-Burnout |
| RR | Realization Rate |
| RUL | Remaining Useful Life |
| SB | Small Business |
| SEER | Seasonal Energy Efficiency Ratio |
| SHGC | Solar Heat Gain Coefficient |
| SKU | Stock-keeping Unit |
| SMECO | Southern Maryland Electric Cooperative |
| TAF | Temperature Adjustment Factor |
| TRM | Technical Reference Manual |
| TSERs | Telephone-Supported Engineering Reviews |
| UMP | Uniform Methods Project (DOE) |
| WHF | Waste Heat Factor |
| WHFd | Waste Heat Factor for Demand |
| WHFe | Waste Heat Factor for Energy |
| UMP | Uniform Method Protocols |
| U.S. DOE | U.S. Department of Energy |
| VFD | Variable Frequency Drive |
| VSD | Variable Speed Drive |