GDS Associates, Inc.
Engineers and Consultants

Nexant

EVALUATION FRAMEWORK

FOR PENNSYLVANIA ACT 129 PHASE II ENERGY EFFICIENCY AND CONSERVATION PROGRAMS

Prepared by **THE STATEWIDE EVALUATION TEAM:**

GDS Associates, Inc.
Nexant, Inc.
Research Into Action, Inc.
Apex Analytics, LLC

CONTRACTED UNDER THE PENNSYLVANIA PUBLIC UTILITY COMMISSION'S RFP 2012-8 FOR THE STATEWIDE EVALUATOR

June 1, 2014
FINAL VERSION

research into action inc

APEX ANALYTICS, LLC

PREPARED BY:

SALIL GOGTE, JESSE SMITH – NEXANT

DICK SPELLMAN, ANDREA JESTER – GDS ASSOCIATES

JANE PETERS – RESEARCH INTO ACTION

SCOTT DIMETROSKY – APEX ANALYTICS

Table of Contents

List of Tables

List of Figures

List of Acronyms

B/C Ratio: Benefit/Cost Ratio

BTUh: BTU-hours

BTUS: Bureau of Technical Utility Services [formerly the Conservation, Economics, and Energy Planning (CEEP)]

CDO: Commercial Date of Operation

CEEP: Conservation, Economics, and Energy Planning [now called the Bureau of Technical Utility Services (BTUS)]

CFL: Compact Fluorescent Light

CPITD: Cumulative Program Inception to Date

Cv: Coefficient of Variation

DLC: Direct Load Control

DR: Demand Response

DSM: Demand Side Management

EC: Evaluation Contractor

ECM: Energy Conservation Measure

EDC: Electric Distribution Company

EE: Energy Efficiency

EE&C Plan: Energy Efficiency and Conservation Plan

EER: Energy-Efficiency Resource

EM&V: Evaluation, Measurement, and Verification

FPC: Finite Population Correction Factor

HVAC: Heating, Ventilating, and Air Conditioning

ICSP: Implementation Conservation Service Provider

IMP: Interim Measure Protocol

IPMVP: International Performance Measurement and Verification Protocol

ISD: In-Service Date

kW: Kilowatt

kWh: Kilowatt-Hour

MMP: Mass Market Protocol

M&V: Measurement and Verification

NPV: Net Present Value

NTG: Net-to-Gross Savings

NTGR: Net-to-Gross Ratio

PEG: Program Evaluation Group

PUC: Pennsylvania Public Utility Commission

PY: Program Year

SEM: Simple Engineering Model

SSMVP: Site-Specific M&V Plan

SWE: Statewide Evaluator

SWE Team: Statewide Evaluation Team

TOU: Time-of-Use

TRC: Total Resource Cost Test

TRM: Technical Reference Manual

TWG: Technical Working Group

UMPSP: Uniform Methods Project Sampling Protocols

VFD: Variable Frequency Drive

VOI: Value of Information

Appendix A contains a glossary of terms.

1 Introduction and Purpose of the Evaluation Framework

This Evaluation Framework includes guidelines and expectations for the seven Pennsylvania electric distribution companies (EDCs) whose energy efficiency and conservation (EE&C) program plans were approved by the Pennsylvania Public Utility Commission (PUC) to promote the goals and objectives of Act 129. The EDCs are: Duquesne Light Company, Metropolitan Edison Company, PECO Energy Company, Pennsylvania Electric Company, Pennsylvania Power Company, PPL Electric Utilities Corporation, and West Penn Power Company.

Through a Request for Proposal (RFP) process initiated on November 30, 2012, the PUC contracted with a Statewide Evaluation Team (SWE Team), led by GDS Associates, Inc. (the Statewide Evaluator, or SWE) that following March. The SWE Team's objective is to complete a comprehensive evaluation of the Phase II EE&C programs (program years ending in 2014, 2015, and 2016) implemented by the seven EDCs in Pennsylvania subject to the requirements of Act 129. GDS led the SWE Team for Phase I as well. As in Phase I, the SWE contract in Phase II is funded by a proration from the EDCs. The other members of the SWE Team are Nexant, Inc., Research Into Action, Inc., and Apex Analytics, LLC.

The SWE Team proposed a scope of work that met all of the requirements for tasks and deliverables in the PUC's RFP, including the level of verification described in the RFP and at the pre-bid meeting. The approach involves auditing verifications completed by EDC evaluators.

To conduct these activities, the SWE Team will collaborate with the seven EDCs, their evaluation teams, and the PUC staff in order to develop appropriate, effective, and uniform procedures to ensure that the performance of each EDC's EE&C programs are verifiable and reliable and meets the objectives of the Act 129 under which the programs were developed.

In accordance with the RFP and the scope of work for the Statewide Evaluator, the SWE Team's tasks are to:

- Develop the Evaluation Framework, specifying:
    - Expectations and technical guidance for evaluation activities
    - Standard data to be collected by implementation conservation service providers (ICSPs) and verified by evaluation contractors (ECs) under contract to the EDCs
    - Audit activities to be conducted by the SWE to confirm the accuracy of EDC-reported and -verified savings estimates
- Perform ongoing impact and cost-effectiveness audits of each EDC's EE&C Plan
- Complete statewide studies and documents, including:
    - Annual updates to the Technical Reference Manual (TRM)
    - Statewide Baseline Study to characterize the market and assess equipment saturation and energy efficiency levels
    - Statewide Market Potential Study to provide estimates for additional electric energy and load reductions for Phase III of the Act 129 programs

To prepare for the program year beginning June 1, 2013, the PUC set the start date for the Phase II SWE contract at April 3, 2013; the contract will end on February 28, 2017. By beginning work in April 2013, the Phase II SWE was able to develop plans and prepare for its responsibilities that began on June 1, 2013. This timing allowed the Phase I SWE Team to complete the annual audit of EDC programs and a review of the Phase I EE&C programs. More information on the responsibilities of the Phase II SWE are provided in the PUC's August 2012 Phase II Implementation Order.[1]

This document satisfies the SWE's first task for Phase II: development of an Evaluation Framework to guide the three-year program evaluation process. The Evaluation Framework outlines the metrics, methodologies, and guidelines for measuring performance by detailing the processes that should be used to evaluate the Act 129 Phase II programs sponsored by the EDCs throughout the Commonwealth of Pennsylvania. It also sets the stage for discussions among a Program Evaluation Group (PEG) of the EDCs, their evaluators, the SWE Team, and the PUC. During these discussions, the PEG will clarify and interpret the TRM, recommend additional measures to be included in the TRM, and define guidelines for acceptable measurement protocols for custom measures in order to mitigate risks to the EDCs. This will require clear and auditable definitions of kWh/yr and kW savings and sound engineering bases for estimating verified gross energy savings.

Specifically, the Evaluation Framework addresses the following:

- Savings protocols
- Metrics and data formats
- Guidance and requirements on claiming savings
- Guidance and requirements on gross impact evaluation procedures
- Guidance and requirements on process evaluation procedures
- Guidance and requirements on net-to-gross (NTG) analysis
- Guidance and requirements on cost-effectiveness analysis
- Guidance and requirements on statistics and confidence/precision
- Required reporting formats
- Data management and quality control guidelines and requirements
- Guidance and requirements on data tracking and reporting systems
- SWE Team public website
- Statewide studies
- Description and schedule of activities the SWE Team will conduct to audit evaluations performed by each EDC's evaluation contractor and assess individual and collective EDC progress toward attainment of Act 129 energy savings targets
- Criteria the SWE Team will use to review and assess EDC evaluations

[1] http://www.puc.state.pa.us/filing_resources/issues_laws_regulations/act_129_information/energy_efficiency_and_conservation_ee_c_program.aspx

Per the PUC, the EDCs must adopt and implement the approved Evaluation Framework upon its release.[2] Any updates to the Evaluation Framework will clarify and memorialize decisions made through other means, such as Orders, Secretarial Letters, and Guidance Memos. The SWE Team will provide PUC-approved updates as addenda to the Evaluation Framework.

1.1 Act 129 Requirements for the Statewide Evaluation

As noted in the introduction, the SWE's services include, but are not limited to:

1. Developing an Evaluation Framework;
2. Monitoring and verifying EDC data collection;
3. Developing and implementing quality assurance processes; and
4. Defining performance measures by customer class.

The SWE is responsible for auditing the results of each EDC's EE&C plan annually and updating the overall EE&C program's goals for Phase III of Act 129. The audits will include an analysis of each EDC plan from process, impact, and cost-effectiveness standpoints. The annual audits will include an analysis of plan and program impacts (energy and demand savings) and cost-effectiveness. The SWE is to report results and provide recommendations for plan and program improvements. The RFP states that the SWE will produce an accurate assessment of the potential for energy savings through market potential assessments, and conduct metering studies to update key assumptions within the residential and non-residential lighting protocols in the TRM. The RFP also specifies that these programs must be implemented pursuant to Act 129 of 2008 and that the evaluations must be conducted within the context of the Phase II Implementation Order and Act 129.[3]

In addition, as needed, the SWE Team will conduct best practice workshops with the EDCs to encourage improvements to impact and process evaluation techniques. The SWE also will produce an accurate assessment of the potential for energy savings through a market characterization and assessment study and propose saving targets for a possible Phase III of Act 129. While all of these tasks are related, each has distinct goals:

- **Impact evaluations** seek to *quantify t*he energy, demand, and possible non-energy impacts that have resulted from demand-side management (DSM) program operations.
- **Process evaluations** seek to *describe* how well those programs operate and to characterize their efficiency and effectiveness.

[2] Exceptions are noted in Section 2.3.3. These exceptions involve portions of the Framework which are intended to align with the 2014 TRM. Consequently, these sections can be considered effective June 1, 2014.

[3] The PUC has been charged by the Pennsylvania General Assembly pursuant to Act 129 of 2008 ("Act 129") with establishing an Energy Efficiency and Conservation (EE&C) program. 66 Pa.C.S. §§ 2806.1 and 2806.2. The EE&C program requires each EDC with at least 100,000 customers to adopt a plan to reduce energy demand and consumption within its service territory. 66 Pa.C.S. § 2806.1. In order to fulfill this obligation, on August 2, 2012, the PUC entered an Implementation Order at Docket No. M-2012-2289411. As part of the Implementation Order and Act 129, the PUC issued an RFP for a Statewide Evaluator (on November 30, 2012) to evaluate the EDCs' Phase II EE&C programs.

- **Cost-effectiveness tests** seek to *assess* that the avoided monetary cost of supplying electricity is greater than the monetary cost of energy efficiency conservation measures.
- **Market characterizations and assessments** seek to *determine* the attitudes and awareness of market actors, measure market indicators, and identify barriers to market penetration.

1.2 Roles and Responsibilities

The following tables, adapted from the RFP, delineate the roles and responsibilities for the EDCs, the SWE Team, and the PUC, by tasks and deliverable, per these categories:

- Statewide Studies
- Audit and Assess EDC Phase II Programs and Results
- Databases
- Primary Data Collection and Impact Analyses
- EDC Plan Review
- Reporting (Annual and Quarterly)
- Best Practices
- Other

When appropriate, the SWE has classified tasks within the EDCs' primary responsibilities as a role of the implementation conservation service provider(s) (ICSP) or evaluation contractor (EC).

Table 1-1: Roles and Responsibilities - Statewide Studies

Task and/or Deliverable[4]	EDC	SWE	PUC
Conduct energy efficiency baseline studies to support Market Potential Study		XX	
Conduct electric energy efficiency Market Potential Study for targets to be achieved in a potential Phase III EE&C Program (6/1/16 to 5/31/21)		XX	
Conduct a Demand Response Potential Study for targets to be achieved in a potential Phase III Demand Response Program (6/1/16 to 5/31/21) (Optional at PA PUC discretion)		XX	
Conduct logging/metering studies to update the hours-of-use values in the TRM for residential and C&I lighting[5]		XX	
Review and get approval of Statewide Baseline, Market Potential, and metering studies (The PUC would get approval of these studies from the Commission.)			XX

[4] The wording of each Task and/or Deliverable in Tables 1-1 through 1-8 and responsibility assignments are copied verbatim from the contract that GDS Associates executed with the Pennsylvania PUC for the Phase II Statewide Evaluator responsibilities. Although several EDCs suggested wording and responsibility assignment changes to these tables for clarification purposes, the SWE decided not to change any of the contract language.

[5] In addition to the logging/metering studies directly conducted by the SWE, logging/metering data from EDC evaluators will also be leveraged when appropriate.

Task and/or Deliverable[4]	EDC	SWE	PUC
Initiate and coordinate annual updates to TRM and interim updates (new protocols)		XX	
Approve TRM updates			XX
Initiate, scope, and conduct/coordinate statewide site inspections, statewide site metering studies, review of data/studies from PA and other states to determine if the PA TRM appropriately estimates savings and/or to revise PA TRM protocols		XX	
Develop and conduct NTG studies	XX		
Coordinate the development of and approve the methodologies for EDC NTG studies consistent with the NTG guidance memos and white papers		XX	
Update Phase I NTG white paper as needed		XX	

Table 1-2: Roles and Responsibilities – Audit and Assessment of EDC Programs and Results

Task and/or Deliverable	EDC	SWE	PUC
Prepare EDC impact and process evaluation plans (EM&V plans), including database and reporting protocols, survey templates, and schedules	EC		
Review and approve the EDC evaluation plans submitted by EDC evaluation contractors		XX	XX
Review and update the Evaluation Framework		XX	
Provide input on the Evaluation Framework as needed	EC		
Approve the statewide Evaluation Framework and revisions			XX
Conduct impact evaluation, process evaluation, NTG analysis, and cost-effectiveness evaluation	EC		
Review/audit all EDC evaluation results, impact evaluation, process evaluation, NTG analysis, and cost-effectiveness evaluation		XX	

Table 1-3: Roles and Responsibilities - Databases

Task and/or Deliverable	EDC	SWE	PUC
Design, implement, and maintain EDC primary program tracking database(s) with project and program data[6]	ICSP		
Establish and implement quality control of EDC program tracking database(s)[7]	EC	XX	
Oversee statewide data management and quality control, including: design, implementation, and maintenance of statewide database of program, portfolio,		XX	

[6] It is highly likely that EDCs have internal program tracking database(s). The entry for responsible party is not limited to the ICSP.

[7] It is the ICSPs' and EDCs' primary responsibility for establishing and implementing QA/QC of EDC program tracking database(s). Evaluation contractors should perform QA/QC of an EDC program tracking database. The SWE audits/reviews the QA/QC performed by an EDC, ICSP, and an evaluation contractor.

Task and/or Deliverable	EDC	SWE	PUC
EDC, and statewide energy and demand savings and cost-effectiveness reporting			
Develop and maintain secure SharePoint site for maintenance and exchange of confidential data and information with EDCs		XX	

Table 1-4: Roles and Responsibilities - Primary Data Collection and Impact Analyses

Task and/or Deliverable	EDC	SWE	PUC
Collect primary data and site baseline and retrofit equipment information	ICSP/ EC		
Determine ex post verification of installation, measure operability, and energy savings	EC		
Analyze and document project, program, and portfolio gross and net energy and demand savings	EC		
Oversee quality control and due diligence, including: inspections of project sites, reviews of primary data and analyses, and preparation of claimed and verified savings	ICSP/ EC		
Audit and assess EDC evaluator contractor performance of EM&V Plans		XX	

Table 1-5: Roles and Responsibilities - EDC Plan Review

Task and/or Deliverable	EDC	SWE	PUC
Review filed EDC EE&C plans and provide advice to PUC staff on ability of plans to meet targets cost-effectively (includes cost-effectiveness analyses)		XX	
Review EDCs' EM&V plans and provide advice to PUC staff on the ability of plans to adequately measure energy savings		XX	

Table 1-6: Roles and Responsibilities – Reporting (Quarterly, Semi-Annual and Annual)

Task and/or Deliverable	EDC	SWE	PUC
Report EDC quarterly and annual energy efficiency (EE)program and portfolio net and gross impacts, as applicable, and cost-effectiveness, and EDC progress in reaching targets; conduct process evaluation	EC		
Develop the statewide semi-annual and annual report templates; review EDC reports and advise the PUC of program and portfolio results: net and gross impacts, cost-effectiveness, and EDC progress in reaching targets (prepare statewide annual and semi-annual reports to the PUC)		XX	
Review and approve SWE semi-annual and annual reports			XX
Review EDC quarterly and annual reports and SWE's semi-annual and annual reports on EE programs: net and gross savings impacts, cost-effectiveness, and EDC progress in reaching targets		XX	XX

Table 1-7: Roles and Responsibilities - Best Practices

Task and/or Deliverable	EDC	SWE	PUC
Participate in at least annual impact evaluation and process evaluation review and improvement workshops[8] as needed	EC	XX	XX
Prepare best practices recommendations for improvements to impact and process evaluation processes		XX	
Prepare best practices recommendations for program modifications and improvements	EC	XX	

Table 1-8: Roles and Responsibilities - Other

Task and/or Deliverable	EDC	SWE	PUC
Prepare materials and reports in support of PUC analysis of efficiency programs		XX	
Organize and conduct periodic stakeholder meetings on evaluation results of EE programs and proposed changes to the TRM		XX	

1.3 Research Objectives

Table 1-9 displays the Evaluation Framework research objectives for three audiences: the Pennsylvania legislature, the PUC, and the EDCs.

Table 1-9: Evaluation Framework Research Objectives

Target Audience	Impact Questions	Process Questions
Pennsylvania Legislature	• Did the EDCs meet statutory targets described in Section 2.1 of this Evaluation Framework? • Were energy and demand savings calculated via vetted protocols (PA TRM and Evaluation Framework)? • Were the EDC EE&C plans implemented in a cost-effective manner in accordance with the Total Resource Cost (TRC) Test?	• Which programs were the most successful and why? • Which programs were the most cost-effective and why? • If an EDC is behind schedule and is unlikely to meet the statutory targets, how can the EDC improve programs in order to meet statutory targets?

[8] The SWE contract for Phase II requires the SWE to conduct workshops for EDCs and their evaluation contractors where Evaluation Framework requirements and evaluation best practices are presented and discussed.

Target Audience	Impact Questions	Process Questions
Pennsylvania PUC	• What level of program energy savings was verified for each EDC and how does this compare to planning estimates and savings reported in EDC quarterly and annual reports? • What assumptions related to energy and demand savings need to be updated in the future TRM versions? • What were the largest sources of uncertainty identified by EDC evaluators related to energy and demand savings and cost-effectiveness?	• Why did planning estimates and reported gross savings differ from verified gross savings? • Considering differences in planning estimates, reported gross savings, and verified gross savings, how can program planning and reporting be improved? • What actions have the EDCs taken in response to process evaluation recommendations made by the EDCs' evaluation contractors? • What were the findings of all of the site inspections conducted by EDCs to verify equipment installation?
Pennsylvania EDCs	• What factors contributed to differences between planning estimates and reported gross savings at the program and portfolio levels? • What factors contributed to differences between *reported* gross savings and *verified* gross savings? • Are there programs or measures which exhibit high free-ridership and may warrant a plan revision? • What factors contributed to differences between planned cost-effectiveness and actual cost-effectiveness at the program and portfolio levels? • Which programs performed the best? Which programs require modification or consideration for elimination based on evaluation results?	• What changes can the EDCs adopt to minimize differences between planning estimates, reported gross savings, and verified gross savings? • What procedural changes can the EDCs adopt to influence customer awareness, satisfaction, and adoption of EE&C programs?

2 Policy Requirements

2.1 Requirements From the Phase II Implementation Order

Act 129 requires the PUC to:

- Adopt an "energy efficiency and conservation program to require electric distribution companies[9] to adopt and implement cost-effective energy efficiency and conservation plans to reduce energy demand and consumption within the service territory of each electric distribution company in this commonwealth"[10];
- Adopt additional incremental reductions in consumption if the benefits of the EE&C Program exceed its costs, and "establish the standards each plan must meet and provide guidance on the procedures to be followed for submittal, review and approval of all aspects of EE&C plans for Phase II of the program"[11];
- Evaluate the costs and benefits of the Act 129 EE&C programs in Pennsylvania by November 30, 2013, and every five years thereafter; and
- Ensure that the EE&C Program includes "an evaluation process, including a process to monitor and verify data collection, quality assurance and results of each plan and the program."[12]

Based on findings from the Phase I Market Potential Study dated May 10, 2012, the PUC determined that the benefits of a Phase II Act 129 program would exceed its costs, and therefore adopted additional required incremental reductions in consumption for another EE&C Program term (program years 2014, 2015, and 2016). In its Phase II Implementation Order, the PUC established targets for those incremental reductions in electricity consumption for each of the seven EDCs in Pennsylvania; established the standards each plan must meet; and provided guidance on the procedures to be followed for submittal, review, and approval of all aspects of EDC EE&C plans for Phase II.[13]

2.1.1 Phase II Energy Reduction Targets for Each EDC

The PUC's August 2012 Implementation Order explained that it was required to establish electric energy consumption reduction compliance targets for Phase II of Act 129. Table 2-1 contains these targets, as percentages and three-year cumulative totals in MWh/year for each of the seven EDCs.

[9] This Act 129 requirement does not apply to an electric distribution company with fewer than 100,000 customers.
[10] See House Bill No. 2200 of the General Assembly of Pennsylvania, An Act Amending Title 66 (Public Utilities) of the Pennsylvania Consolidated Utilities, October 7, 2008, page 50.
[11] Pennsylvania Public Utility Commission, *Energy Efficiency and Conservation Program* Implementation Order, at page 1, at Docket No M-2012-2289411, (*Phase II Implementation Order*), entered August 3, 2012.
[12] See House Bill No. 2200 of the General Assembly of Pennsylvania, An Act Amending Title 66 (Public Utilities) of the Pennsylvania Consolidated Utilities, October 7, 2008, page 51.
[13] Pennsylvania Public Utility Commission, *Energy Efficiency and Conservation Program* Implementation Order, at page 20, at Docket No M-2012-2289411, (*Phase II Implementation Order*), entered August 3, 2012.

Table 2-1: Act 129 Phase II Three-Year Energy Efficiency Reduction Compliance Targets

EDC	Three-Year Program Acquisition Cost ($/MWh/Yr)	Three-Year % of 2009/2010 Forecast Reductions	Three-Year MWh/Yr Value of 2009/2010 Forecast Reductions
Duquesne	$211.90	2.0	276,722
Met-Ed	$220.87	2.3	337,753
Penelec	$216.19	2.2	318,813
Penn Power	$209.20	2.0	95,502
PPL	$224.71	2.1	821,072
PECO	$227.55	2.9	1,125,851
West Penn	$209.42	1.6	337,533

2.1.2 Standards Each EDC's Phase II EE&C Plan Must Meet

The PUC requires that each EDC's plan for Phase II meet several standards, including the following:

1. EDCs must include in their filing an EE&C Plan that obtains at least ten percent (10%) of all consumption reduction requirements from the federal, state, and local governments, including municipalities, school districts, institutions of higher education, and nonprofit entities.
2. Each EDC Phase II EE&C Plan must obtain at least four-and-one-half percent (4.5%) of its consumption reduction requirements from the residential low-income sector. Act 129 also includes legislation to ensure that there are specific measures available for and provided to low-income customers. The compliance criteria for this metric are to include a number of energy efficiency measures for households at or below 150% of the federal poverty income guidelines that is proportionate to each EDC's total low-income consumption relative to the total energy usage in the service territory. The SWE has advised that EDCs should consider the definition of a low-income measure to include a measure that is targeted to low-income customers and is available at no cost to low-income customers. Alternatively, to satisfy the 4.5% savings requirement, EDCs will be provided the flexibility to include savings from general residential measures that low-income customers may have participated in, as well as savings attributed to specific low-income measures. Appendix I provides further clarification.
3. EDCs should determine the initial mix and proportion of energy efficiency programs, subject to PUC approval. The PUC expects the EDCs to provide a reasonable mix of energy efficiency programs for all customers. However, each EDC's Phase II EE&C Plan must ensure that the utility offers each customer class at least one energy efficiency program.

2.1.3 Accumulation of "Over-Compliance" Savings from Phase I

The PUC's August 2012 Implementation Order for Phase II specifies that its Bureau of Technical Utility Services (BTUS) staff will coordinate with the EDCs to add a line item regarding energy savings that exceed the Phase I targets to their Phase II quarterly and annual reports. The EDCs are allowed to use such "over-compliance" savings from Phase I toward meeting their Phase II savings reduction targets.

2.1.4 Expired Savings

In some cases, the effective useful life, or measure life, of an energy efficiency measure is shorter than the length of the Act 129 phase in which it was implemented. The Phase II Implementation Order specified how EDCs and their evaluation contractors should address this issue with respect to calculating and reporting gross verified savings estimates toward achieving the prescribed reduction targets.[14]

Based on this policy requirement, if an EDC were to implement an energy efficiency measure in PY5 of Act 129 with an effective useful life of one (1) year, savings from that measure could not be counted toward that EDC's Phase II MWh/yr reduction target because the savings from the measure would have expired before the end of the phase (in this case, May 31, 2016). If the same measure were installed in the final year of Phase II (PY7), the savings would not expire and the measure savings would count toward the compliance reduction target.

2.1.5 Exclusion of Peak Demand Goals

Demand response and demand reduction goals were not established for Phase II. As a result, this Evaluation Framework does not focus on EM&V activities pertaining to peak demand reduction programs like Demand Response (DR). Instead, the PUC established a three-year term for Phase II for several reasons. One primary reason was to enable the PUC time to evaluate the current and future peak demand reduction program design and assess the potential for DR savings in a possible Phase III EE&C program. If the PUC determines that a program design can cost-effectively achieve peak demand reductions, it will incorporate such a program into its plan for a potential Phase III. This process will enable the establishment of both energy conservation and peak demand targets and budgets for an entire phase prior to program implementation. The PUC stated that the plan for Phase II would provide the certainty needed to fund and operate the Act 129 programs throughout Phase II, and prevent program disruptions and administrative costs that would be incurred if the PUC were to insert a peak demand reduction program target during Phase II.

2.1.6 Net-to-Gross Ratio for Phase II of Act 129

The PUC's Phase II Implementation Order directs that net-to-gross (NTG) adjustments be treated the same way for Phase II as they were during Phase I. Specifically, the PUC directed that NTG research be used to direct Act 129 program design and implementation, but not for compliance purposes.[15] The PUC's Phase II Implementation Order notes that there is no requirement in Act 129 that kWh/yr and kW savings are determined on a net basis. Accordingly, the PUC directed in its Phase II Order that EDCs

[14] "The Act 129 programs are cumulative at the end of a phase such that the savings at the end of a phase must show that the total savings from measures installed during the phase are equal to or greater than the established reduction target. Therefore, if any measures are installed whose useful life expires before the end of the phase, another measure must be installed or implemented during that phase which replenishes the savings from the expired measure." Pennsylvania Public Utility Commission, *Energy Efficiency and Conservation Program* Implementation Order, at page 26, at Docket No M-2012-2289411, (*Phase II Implementation Order*), entered August 3, 2012.

[15] Pennsylvania Public Utility Commission, *Energy Efficiency and Conservation Program* Implementation Order, at page 78, at Docket No M-2012-2289411, (*Phase II Implementation Order*), entered August 3, 2012.

continue to use net verified savings for program planning purposes (e.g., program design, modifying program incentive levels and eligibility requirements) and that compliance with savings targets in Phase II be determined using gross verified savings. Section 3.6 of this Evaluation Framework contains guidance on how EDC evaluation contractors should conduct NTG research in Phase II and how the results of this research can be incorporated into program planning.

2.2 2013 TRC Order

2.2.1 Intent of the TRC Order

Act 129 of 2008, 66 Pa. C.S. § 2806.1 directs the PUC to use a TRC Test to analyze the benefits and costs of the EE&C plans that certain EDCs must file.[16] The PUC established the TRC Order to provide guidance, methodology, and formulas for properly evaluating the benefits and costs of the proposed EE&C plans. All cost-effectiveness evaluations and assessments must be conducted in accordance with the TRC Order. The TRC Test for Phase II will be applicable throughout Phase II, unless the PUC determines a need to modify the TRC during Phase II.

2.2.2 2013 TRC Order

Although much of the 2013 Phase II TRC Test Order (issued August 30, 2012) is consistent with the 2009 and 2011 TRC Orders used throughout the course of Phase I, there are some refinements and additional guidelines. In particular, the 2013 TRC Test Order adopts several enhancements to the methodology for calculating forecasted avoided costs, including those regarding: avoided energy costs; transmission, distribution, and capacity costs; and avoided costs for compliance with the Alternative Energy Portfolio Standards (AEPS).

The Phase II Order also clarifies the appropriate timing and use of updated avoided costs for program screening, the reporting of low-income savings from non-low-income sectors, and the absence of demand response from Phase II. The 2013 TRC Test Order specifies EDCs will continue to use net verified savings in their TRC test for program planning purposes and cost-effectiveness compliance in Phase II will be determined using gross verified savings. Finally, the Order addresses proper protocols for determining measure incremental costs, as discussed in the section below.

2.2.3 Incremental Costs

From a planning standpoint, one of the key elements of each measure is its incremental cost. For Phase I, energy efficiency cost calculations used only the incremental energy efficient measure costs. For new construction and existing equipment measures that have reached the end of their useful life, incremental cost is the additional cost incurred to buy an efficient device or measure, over and above the cost of the standard efficiency device or measure. For the early replacement of a functioning device, incremental cost is defined as the full cost of the new efficient measure (including installation costs) that is bought to replace the existing, installed measure.

[16] The Pennsylvania TRC Test for Phase I was adopted by PUC order at Docket No. M-2009-2108601 on June 23, 2009 (*2009 PA TRC Test Order*). The TRC Test Order for Phase I later was refined in the same docket on August 2, 2011 (2011 PA TRC Test Order). The 2013 TRC Order for Phase II of Act 129 was issued on August 30, 2012.

Phase II adopts the same incremental cost calculation method, but will not preclude an EDC from using an alternative calculation method for early retirement measures, where the incremental cost is the difference between the full cost of the efficient device (including installation) minus the present value[17] of the standard device (plus installation costs). The installation of an early retirement measure disrupts the schedule of natural retirement and replacement of the existing standard device, reducing the present value of the costs of future replacement installations.[18] If an EDC uses this alternative calculation method, it must document, in its EE&C plans and annual reports, which method it used and why.

Additionally, the Phase II TRC Order recognizes the development of a statewide incremental cost database to assist EDCs in their development of TRC ratio calculations and to promote consistency in TRC calculations. The TRC Order also clarifies that EDCs will have the flexibility to choose between values in the new SWE incremental cost database, values from the DEER database (adjusted for regional and local conditions using appropriate cost multipliers), values currently used for program planning and cost-effectiveness testing, or other pertinent data sources. EDCs are expected to document, in their annual reports, the source of incremental cost data and their reasons for choosing that source. The cost values selected by the EDCs are subject to PUC review.

2.2.4 TRC Order Schedule

The SWE will coordinate and participate in a periodic review of the TRC Test that the PUC and EDCs use to determine cost-effectiveness. The PUC issued a Final Implementation Order for the TRC Test for Phase II of Act 129 EE&C programs on August 30, 2012. Unless otherwise directed, the SWE will assist the PUC in updating the TRC Test during the first eight months of 2015 if the PUC decides to implement a Phase III Act 129 EE&C Program. The 2015 update of the TRC Test would be effective at the beginning of any Phase III programs.

2.3 PA TRM Order and TRM Manual

In implementing the AEPS Act, 73 P.S. §§ 1648.1 – 1648.8, the PUC adopted Energy Efficiency and DSM Rules for Pennsylvania's AEPS, including a Technical Reference Manual (TRM) for the State of Pennsylvania on October 3, 2005.[19] The PUC also directed the Bureau of Conservation, Economics and Energy Planning (CEEP)[20] to oversee the implementation, maintenance, and periodic updating of the TRM.[21] On January 16, 2009, in the Energy Efficiency and Conservation Program Implementation Order

[17] *Present value* refers to the future amount of money that has been discounted to reflect its current value, as if it existed today. The present value is always less than or equal to the future value because money has interest-earning potential, a characteristic referred to as the *time value of money.*

[18] Retrofit Economics 201: Correcting Common Errors in Demand-Side Management Cost-Benefit Analysis. R. Brailove, J. Plunkett, and J. Wallach. Resource Insight, Inc. 1995.

[19] Order entered on October 3, 2005, at Docket No. M-00051865 (October 3, 2005 Order).

[20] As of August 11, 2011, the Bureau of CEEP was eliminated and its functions and staff transferred to the newly created Bureau of Technical Utility Services (BTUS). See Implementation of Act 129 of 2008; Organization of Bureaus and Offices, Final Procedural Order, entered August 11, 2011, at Docket No. M-2008-2071852, at page 4.

[21] See October 3, 2005 Order at page 13.

for Phase I of Act 129's EE&C Program,[22] the PUC adopted the TRM as a component of the EE&C Program evaluation process. In the Phase I Implementation Order, the PUC also noted that, "as the TRM was initially created to fulfill requirements of the AEPS Act, it will need to be updated and expanded to fulfill the requirements of the EE&C provisions of Act 129."[23] Soon after the adoption of the EE&C Program Phase I Implementation Order, PUC staff initiated a collaborative process to review and update the TRM with the purpose of supporting both the AEPS Act and the Act 129 EE&C Program that culminated in the adoption of the 2009 TRM at the May 28, 2009 public meeting.[24] In adopting the 2009 TRM, the PUC recognized the importance of updating the TRM annually.[25] A program evaluation group (PEG)[26] was formed to, among other things, provide guidance to the SWE in clarifying energy savings measurement protocols and plans by recommending improvements to the existing TRM and other aspects of the EE&C program. In addition, the PUC convened a Technical Working Group (TWG)[27] meeting to discuss the proposed TRM updates.[28]

During Phase I of Act 129, the PUC filed and approved the 2009, 2010, 2011, and 2012 TRM Orders. The approval date and effective date of these TRM versions are presented in Table 2-2. Previous TRM orders and TRM manuals can be accessed through the PUC website.[29]

The approval date of the TRM is when the TRM Order was entered after the PUC approved it during a public meeting; this differs from the effective date of the TRM, which specifies when the TRM shall be used. The effective date typically will align with the next program year, which spans from June 1 to May 31 of the following year.

[22] See Energy Efficiency and Conservation Program Implementation Order at Docket No. M-2008-2069887, (Phase I Implementation Order), at page 13, entered January 16, 2009.
[23] Ibid.
[24] See Implementation of the Alternative Energy Portfolio Standards Act of 2004: Standards for the Participation of Demand Side Management Resources – Technical Reference Manual Update Order, at Docket No. M-00051865, (2009 TRM), entered June 1, 2009.
[25] Ibid., pages 17 and 18.
[26] The PEG is chaired by PUC staff and is comprised of representatives from the EDCs and the SWE to encourage discussions of EDC program-specific issues and associated evaluation, measurement, and verification.
[27] The TWG is chaired by PUC staff and is comprised of representatives from the EDCs, the SWE, and other interested parties to encourage discussions of the technical issues related to the EM&V of savings programs to be implemented pursuant to Act 129.
[28] The PUC held TWG meetings to provide stakeholders with the opportunity to review proposed high-impact changes to residential, commercial, and industrial measures, and also allow for a question and answer session about those changes. Additionally, stakeholders had the opportunity to propose any other changes to the TRM.
[29] See link:
http://www.puc.state.pa.us/filing_resources/issues_laws_regulations/act_129_information/technical_reference_manual.aspx

Table 2-2: Phase I Approved TRMs[30]

TRM Version	Program Year	Approval Date	Effective Date[31]
2009 TRM	**PY 1**	June 1, 2009	June 1, 2009 – May 31, 2010
2010 TRM	**PY 2**	June 8, 2010	June 1, 2009 – May 31, 2011[32]
2011 TRM	**PY 3**	February 28, 2011	June 1, 2011 – May 31, 2012
2012 TRM	**PY 4**	December 15, 2011	June 1, 2012 – May 31, 2013

For Phase II of the Act 129 EE&C program, the PUC again adopted the TRM as a component of the EE&C Program evaluation process.[33] The Phase II Implementation Order also recognized the importance of the continued use of an annual updating process for the TRM for Phase II.[34] This evaluation framework adopts the policy guidance provided by the PUC relating to the importance of the annual updating process for the Pennsylvania TRM. The PUC approved the 2013 TRM, which pertains to the Act 129 EE&C Program Phase II compliance year (June 1, 2013-May 31, 2014) through the 2013 TRM Order, and the 2014 TRM through the 2014 TRM Order.

The TRM Order represents the PUC's continuing efforts to establish a comprehensive and up-to-date TRM with a purpose of supporting the EE&C Program provisions of Act 129. The PUC will continue to use the TRM to help fulfill the evaluation process requirements contained in the Act. By maintaining up-to-date information, the PUC assures that Act 129 monies collected from ratepayers are reflecting reasonably accurate savings estimates.

The TRM is organized into several chapters. The first chapter provides guidance and overarching rules regarding use of the TRM. The second chapter contains TRM protocols, or measure-specific methodologies for estimating energy and demand savings, for residential measures. The third chapter contains TRM protocols for commercial and industrial measures. The TRM also contains appendices to present information that does not easily fit the template of a TRM protocol.

[30] See Table 2-3 for Phase II TRMs.
[31] Note: The effective date is the period of time during which the TRM is actively used to determine energy and demand savings. For retrofit and replacement measures, the applicable TRM (effective date) is based on the in-service date of the energy efficiency measure. For new construction projects, the applicable TRM (effective date) is based on the date the construction permit was issued, or the construction start date if no permit is required.
[32] The 2010 TRM Order was entered on June 8, 2010 and applied retroactively to the start of the Act 129 Program, per direction of the PUC, which effectively overrode all 2009 TRM protocols. This was an exceptional case; normally, such a decision does not occur unless explicitly stated through PUC directives. See 2010 TRM Order, pages 16-17.
[33] See *Energy Efficiency and Conservation Program Implementation Order*, at Docket No M-2012-2289411, (*Phase II Implementation Order*), entered on August 3, 2012, at page 71.
[34] Ibid. at page 75.

2.3.1 Purposes of the TRM

The TRM serves a variety of purposes for Act 129. In addition to providing measure savings protocols, the TRM ultimately seeks to facilitate the implementation and evaluation of Act 129 programs**.** The TRM fulfills the following objectives:

- Serves as a common reference document for energy efficiency measures to be used by EDCs, ICSPs, evaluation contractors, the SWE, the PUC, and other stakeholders
- Establishes standardized, statewide protocols to calculate energy and demand savings for measures. The EDC ICSPs use these protocols to estimate ex ante (reported or claimed) savings achieved for the energy efficiency measures. EDC evaluation contractors use these protocols to estimate ex post (verified) savings achieved for energy efficiency measures
- Increases transparency to all parties by documenting underlying assumptions and tracking references used to develop savings estimates for measures
- Balances the accuracy of savings estimates with costs incurred to measure and verify the savings estimates
- Provides reasonable methods for measurement and verification (M&V) of incremental energy savings associated with EE&C measures without unduly burdening EDC EE&C program implementation and evaluation staff
- Reduces the number of EE&C measures that must be evaluated as custom measures

2.3.2 TRM Update Process

For the TRM to be an effective tool for Act 129, the PUC ordered a regular annual update to the TRM through the 2009 Implementation Order.[35] The PUC intends to update and expand the TRM annually to fulfill the requirements of the EE&C provisions of Act 129.[36] All changes made during the TRM update process will be prospective and thus will not retrospectively affect savings determinations for the program year already underway, unless otherwise determined by the PUC. Updates to the TRM will occur per the typical stakeholder process, which adheres to the Tentative Order, Comment Period, and Final Order procedure (see Figure 2-1).

[35] See Energy Efficiency and Conservation Program Implementation Order at Docket No. M-2008-2069887, (Phase I Implementation Order), entered January 16, 2009.
[36] *Phase II Implementation Order*, pp. 71-72

Figure 2-1: TRM Update Process (Annual)

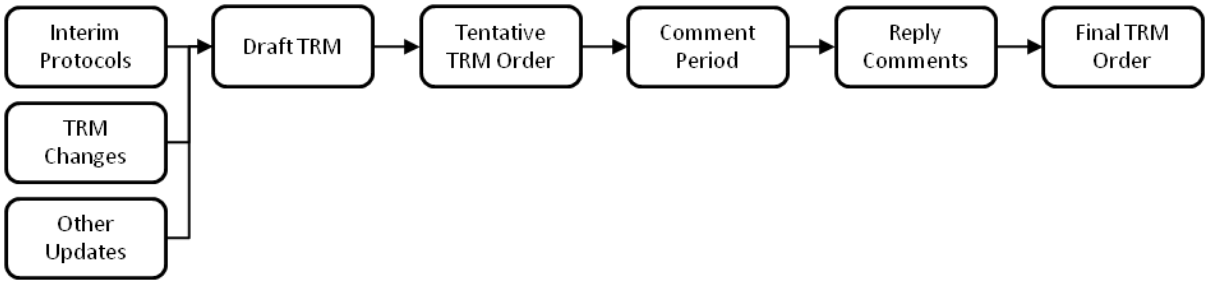


Table 2-3 displays the estimated schedule for approval of the TRM and effective dates. While the actual dates of the deadlines leading to the approval of each TRM may change, the PUC targets an approval date of December 31, which provides a five-month period between the approval date and the effective date, to allow EDCs enough time to update program documents, marketing materials, tracking databases, and other program or system processes. The SWE Team notes that the interval between the approval date and the effective date for previous TRMs has been more than five months.

Table 2-3: Timeline for TRM Updates

TRM Version	Program Year	Approval Date	Effective Date[37]
2013 TRM	PY5	December 20, 2012	June 1, 2013 – May 31, 2014
2014 TRM	PY6	December 19, 2013	June 1, 2014 – May 31, 2015
2015 TRM	PY7	December 31, 2014*	June 1, 2015 – May 31, 2016
2016 TRM	PY8	December 31, 2015*	June 1, 2016 – May 31, 2017
2017 TRM	PY9	December 31, 2016*	June 1, 2017 – May 31, 2018
2018 TRM	PY10	December 31, 2017*	June 1, 2018 – May 31, 2019
2019 TRM	PY11	December 31, 2018*	June 1, 2019 – May 31, 2020
2020 TRM	PY12	December 31, 2019*	June 1, 2020 – May 31, 2021

*****Estimated date

As stated before, the *approval date* of the TRM is the date the TRM Order was entered after approval by the PUC during a public meeting; this differs from the *effective date* of the TRM, which specifies when the TRM shall be used. The effective date typically will align with the next program year, which spans from June 1 to May 31 of the following year.

The PEG, comprising BTUS staff, the SWE, EDCs, and EDC evaluation contractors, has been initiated to review, clarify, improve, and add new savings protocols to the TRM. Generally, the mission of this group is to provide technical guidance to the PUC regarding the quantification of energy and demand savings.

[37] Note: The effective date is the period of time during which the TRM is used. For retrofit and replacement measures, the applicable TRM (effective date) is based on the in-service date of the energy efficiency measure. For new construction projects, the applicable TRM (effective date) is based on the date the construction permit was issued, or the construction start date if no permit is required.

Protocols for any measures that are not already included in the TRM may be proposed through the Interim Measure Process (Section 2.3.6).

As impact evaluation results become available, they will serve as indicators to identify measure protocols that may require updates in the TRM. The PEG review process will explore the applicability of these findings to ensure that the TRM presents the best available estimates of energy and demand savings. Measure attributes will be updated through dedicated measure research studies informed by the impact evaluation findings during the PEG review process.

2.3.3 Alignment of the Evaluation Framework and the TRM

Due to the update schedules of both the Evaluation Framework and the TRM, there is potential for conflicting expectations. When Framework guidance is dependent upon prospective TRM requirements, the effective date of that particular Framework guidance will coincide with the corresponding TRM revision. Concurrent with this update of the Evaluation Framework, there are no such conflicts anticipated between the Evaluation Framework and the 2014 TRM.

2.3.4 TRM Protocols

A TRM protocol is a measure-specific methodology for calculating energy and demand savings. The TRM contains protocols that determine savings for standard measures by either deeming savings or providing an algorithm with variables to calculate savings. Mass market protocols (MMP) have been developed to estimate energy and demand savings associated with residential behavioral modification and low-income weatherization programs. These MMPs are discussed in Section 2.3.7.1 of this Framework.

The Pennsylvania TRM categorizes all measures into three categories: *deemed measures, partially deemed measures,* and *custom measures*.

- *Deemed measures* are well-defined measures that have specified (fully stipulated) energy and demand savings values; no additional measurement or calculations are required to determine deemed savings.
- *Partially deemed measures* are determined using an algorithm with stipulated and open variables, thereby requiring data collection of certain parameters to calculate the energy and demand savings.
- *Custom measures* are considered too complex or unique (because there are highly variable or uncertain savings for the same measure) to be included in the list of standard measures provided in the TRM and so are outside the scope of the TRM (Section 2.3.4.3).

2.3.4.1 Deemed Measures

A deemed measure protocol specifies a pre-determined amount of energy and demand savings per unit. For the PA TRM, deemed measure protocols also may contain an algorithm with stipulated variables to provide transparency into deemed savings values and to facilitate the updating of the deemed savings values. Stipulated variables, which are assumptions that must be used and are established through the TRM update process, cannot be changed mid-cycle without approval from the PUC.

The TRM contains many protocols with deemed savings. This type of protocol typically is used for measures whose parameters are well understood or well documented; it is particularly appropriate for residential measures involving customers with similar electricity usage characteristics, as well as for "give-away" programs.

Recommendations of the SWE to the PUC regarding TRM deemed savings protocols for future years include:

- Maintain an active PEG, chaired by the SWE, including technical experts from the utilities and other independent experts to provide input on evolving technologies and measure assumptions.
- Identify measure protocols to be reviewed annually based on relative savings contributions, evaluation findings, statewide studies, changes to federal and state energy efficiency codes and recent secondary research.
- Conduct a periodic national review of deemed savings databases to determine how others have used this tool and the assumptions they have utilized.
- Examine literature referenced in the TRM during the annual TRM update process that supports the deemed savings assumptions; this would include reviews of the population or tests from which the data were derived and recommendations about the population or technologies to which the generalizations should be applied in Pennsylvania.
- Update the TRM deemed and partially deemed measures on an annual basis for measures identified as needing updates.

2.3.4.2 Partially Deemed Measures

The Pennsylvania EE&C programs include several measures that utilize savings measurement protocols based on partially deemed savings. Customer-specific information is used for each open variable, resulting in a variety of savings values for the same measure. This method is commonly used when well-understood variables affect the savings and can be collected from the applicant. Some open variables may have a default value to use when the open variable cannot be measured.

Open variables include:

- Capacity of an A/C unit
- Square footage of insulation
- Hours of operation of a facility or of a specific electric end use
- Horsepower of a fan or pump motor

Recommendations of the SWE to the PUC regarding TRM partially deemed savings protocols for future years include:

- Identify high-impact measure protocols for review on an annual basis, and provide necessary clarifications or modifications through the PEG, based on evaluation findings, statewide studies, changes to federal and state energy efficiency codes or more recent and reliable secondary research available.

- Analyze algorithms and definitions of terms during the annual TRM update process to verify that the protocols use accepted industry standards and reasonably estimate savings.
- Ensure that the methodologies for implementing protocols are clearly defined and can be implemented practically and effectively.
- For nonresidential measures, establish energy impact thresholds by measure type in the TRM, above which customer-specific data collection is required for open variables. The intent of this change is to reduce the overall uncertainty of portfolio savings estimates by increasing the accuracy of project-level savings estimates for extremely high-impact measure installations.
- Conduct Pennsylvania-specific research studies (for example, residential and commercial and industrial (C&I) lighting metering studies and eQUEST modeling for C&I HVAC, motors, and VFD [variable frequency drive] measures) to update key assumptions for high-impact measures and provide load shapes for each measure variant.[38]
- Examine the literature referenced in the TRM supporting key variables used in partially deemed savings algorithms which warrant further review and discussion by the PEG; this may include reviewing the population from which source data were derived, if available, and providing recommendations regarding the appropriate population or technologies to which the generalizations should be applied.

2.3.4.3 Custom Measures

The TRM presents some information about custom measures that are too complex or unique to be included on the list of standard measures in the TRM. Accordingly, savings for custom measures are determined through a custom-measure-specific process, which is not contained in the TRM (See Section 2.3.7).

2.3.5 Using the TRM

The TRM provides a standardized statewide methodology for calculating energy and demand savings. The TRM also provides a consistent framework for EDC ICSPs to estimate ex ante (claimed) savings and for EDC evaluation contractors to estimate *ex post* (verified) savings.

2.3.5.1 Using the TRM to Determine Ex Ante *Savings*

This section outlines how EDC ICSPs should calculate ex ante savings.[39]

For replacements and retrofits, ICSPs will use the applicable date to determine which TRM version to select to estimate EDC claimed savings is the "in-service date" (ISD) or "commercial date of operation" (CDO), the date at which the measure is "installed and commercially operable"[40] and when savings

[38] The schedule for developing load shapes for measures in the TRM will be determined by the SWE in collaboration with TUS staff, the EDCs, and their evaluation contractors.
[39] In some cases, an EDC may choose to implement a program "in-house" rather than engaging an implementation CSP. In these cases, EDC staff is acting in the capacity of the implementation CSP.
[40] Pennsylvania Public Utility Commission Act 129 Phase II Order, Docket Number: M-2012-2289411 and M-2008-2069887, Adopted August 2, 2012, language in Section K.1.b.

actually start to occur. This scenario is analogous to when a commercial customer's meter "sees" the savings under expected and designed operation.

For projects with commissioning, the CDO is the date commissioning is completed. For incented measures that have been installed, but are not being used (because there is no occupant, or the measure will not be used until an unrelated installation/project is completed), the equipment is not "commercially operable." For these projects, the CDO is the date at which the customer begins using the incented equipment, not the date at which the equipment is energized.

For new construction, selection of the appropriate TRM must be based on the date when the building/construction permit was issued (or the date construction starts if no permit is required) because that aligns with codes and standards that define the baseline. Savings may be claimed toward compliance goals only after the project's ISD.

Methods used by the ICSPs to estimate ex ante savings differ for each of the three measure categories (deemed, partially deemed, and custom measures).

For **deemed measures**, ex ante savings are determined by applying the deemed savings values in the TRM. Assumptions, which may be listed in the TRM for transparency, may not be adjusted by EDC ICSPs using customer-specific or program-specific information.

For **partially deemed measures**, ex ante savings are determined by using the algorithms provided in the TRM; these formulas include both stipulated and open variables. Stipulated variables are defined as any variable in the TRM that does not have an "EDC Data Gathering" option and are fully deemed. These values may not be changed or revised by EDC ICSPs. Open variables[41] in the TRM have an "EDC Data Gathering" option. These values must come from customer-specific information or default values provided in the TRM. EDC ICSPs should attempt to collect customer-specific values for each rebated measure through the application process. Only variables specifically identified as open variables may be adjusted using customer-specific information. If the ICSPs choose to utilize the EDC data gathering option for a particular open variable, the findings of the EDC data gathering should be used for all instances of that variable. ICSPs are not allowed to revert to the default value once the EDC data gathering option is chosen. However, if customers are unable to provide data for the variable, then ICSPs should use the default value found in the TRM for those customers only. For measures where EDC data gathering is utilized, EDCs should report on findings in annual reports.

The SWE will collaborate with the EDCs and their evaluators during the TRM update process to identify any stipulated variable that should be changed to an open variable and vice versa. The criteria for making such changes may include the feasibility of attaining such information, the percent change in savings expected when using open versus stipulated variables, and the uncertainty surrounding default values.

[41] Open variables are listed with a default value and an option for "EDC Data Gathering" in the TRM.

For certain nonresidential end-use categories, Section 1.2 of the TRM defines thresholds where M&V is required if the threshold is exceeded. In other words, if the combined savings for a certain end-use category in a single project is above the corresponding end-use category threshold established in the TRM, the ICSP cannot use default values but is instead required to use customer-specific data collected through M&V activities. If claimed savings for an end-use category (e.g., lighting, motors) within a project falls below the threshold specified in the TRM, the ICSPs may gather customer-specific data or use the default TRM value.

It is helpful for ICSPs to use the same approach as the evaluation contractor for determining when they must use customer-specific data gathering in order to estimate ex ante savings. EDCs or ECs should assist the ICSPs in interpreting the requirements of this Evaluation Framework, including determination of ex ante savings methodologies at the project and/or measure level. The use of similar methodologies to estimate savings between the implementers and evaluators will increase the likelihood of a strong correlation between the ex ante and ex post savings and improve the precision of savings estimates for a given sample size.

If an EDC, ICSP, or evaluation contractor believes the information in the TRM regarding a deemed or partially deemed measure should be revised, they should submit a written request to the PEG for review and consideration in the next TRM update.

For **custom measures**, ex ante savings are determined using the custom measure process described in Section 2.3.7.

Measures that are not included in the TRM but still require a deemed or partially deemed approach may be claimed using the Interim Measure Protocol approach described in Section 2.3.6.

2.3.5.2 Using the TRM to Determine Ex Post *Savings*

Typically, EDC evaluation contractors conduct research studies, site inspections, and documentation reviews based on statistically representative samples to determine ex post savings. The appropriate method used to determine verified savings differs for the three measure categories and may further depend on the magnitude of the project's savings. These measure categories, defined below and summarized in Table 2-4, dictate the methodology to use for estimating ex post savings.

Table 2-4: Measure Categories

Measure Category	Ex Post Calculation Methodology	Example Measures
TRM deemed savings measures	Follow deemed savings per TRM	Furnace whistle
TRM partially deemed measures	Follow TRM savings algorithms, using deemed variables and verified open variables	C&I lighting, residential lighting (CFLs & LEDs), C&I motor
Custom measures	All other, unspecified	Non-TRM VFD, non-TRM chiller, Energy Management System (EMS)

For **deemed measures**, the TRM provides per-unit savings allowances that both the ICSPs and evaluators will use; the energy and demand savings of these measures are deemed with all energy-related variables stipulated. Thus, the evaluation activity for deemed measures will include verification of measure installation, quantity, and correct use of the TRM measure protocol. The evaluator will estimate ex post savings using deemed savings and/or stipulated assumptions in accordance with the TRM.

For **partially deemed measures**, the EDC evaluation contractor will estimate ex post savings using the algorithms provided in the TRM; these formulas include both stipulated and open variables. The open variables typically represent or describe straightforward key measure-specific inputs in the savings algorithms, which improve the reliability of savings estimates (e.g., capacity, efficiency ratings). Evaluation activities for partially deemed measures include: verification of measure installation, quantity, and the correct use of the TRM protocol; verification of open variables, which may entail confirming nameplate data; facility staff interviews; or measurements of the variable(s). Evaluators should attempt to verify as many open[42] values in the TRM algorithm as possible with customer-specific or program-specific information gathered through evaluation efforts. Open variables in the TRM may have a default stipulated value, which should be used if customer-specific or program-specific information is unreliable or the evaluators cannot obtain the information.

Customer-specific data collection and engineering analysis will depend on the type of measure (uncertainty and complexity) and the expected savings (level of impact). The ICSP is primarily responsible for collecting customer-specific data through supporting documentation, phone or in-person interviews with an appropriate site contact, a site visit, pre- and post-installation metering, analysis of consumption histories, analysis of data from building monitoring equipment, and/or energy modeling simulations. For example, estimating savings for commercial lighting projects requires detailed information about pre- and post-installation conditions for lighting retrofits, such as fixture and ballast type, fixture wattage, building and space type, hours of use (HOU), and lighting controls. When allowed by the TRM, using more accurate customer-specific values for a partially deemed measure is encouraged for high-value nonresidential projects above a threshold kWh/yr.[43] Evaluation contractors should verify the customer-specific data for all measures in the evaluation sample. If the evaluation contractor determines that the customer-specific data gathered by the ICSP are not reasonably valid, then the evaluator should conduct independent customer-specific data gathering activities for those measures.

Section 3.3.2.3 provides additional information on nonresidential savings thresholds for project stratification and determination of measure level rigor.

For both deemed measures and partially deemed measures, if an EDC does not wish to use the protocols in the applicable TRM, they may use a custom method to calculate and report savings, as long as they: 1) also calculate the savings using TRM protocols, and 2) include both sets of results in the quarterly and/or annual EDC reports. The EDCs must justify the deviation from the TRM in the quarterly

[42] Open variables are signified by the term "EDC data gathering" in the TRM.
[43] The threshold kWh/yr is stipulated in the TRM and will vary depending on the type of measure.

and/or annual reports, wherein they report the deviations. EDCs should be aware that use of a custom method as an alternative to the approved TRM protocol increases the risk that the PUC may challenge their savings. If an EDC uses a custom method to calculate savings for a TRM measure, the SWE will perform a pre-approval review only if the PUC requires them to do so.

Custom methods to calculate savings differ from using program-specific or customer-specific information for open variables defined in the TRM protocols (See Section 2.3.5.1). EDCs are encouraged to use customer-specific or program-specific information for open variables in the TRM protocol.

For **custom measures**, the savings impacts vary per project. The customer, the customer's representative, or a program administrator typically estimates the project's savings before an EDC pays the incentive. Due to the complexity of custom measures and the information required to reasonably estimate savings for them, EDCs may choose how to estimate reported gross savings. The EDC evaluation contractor must verify reported gross savings to an acceptable degree and level of rigor. In some cases, evaluation activities may require the measurement of energy and/or demand consumption, both before and after the implementation of the custom measure; in other cases, engineering models and regression analysis may be permitted. Therefore, the audit activities for custom measures typically depend on the evaluation process selected for the category of custom projects.

2.3.6 Interim Measure Protocols

Interim Measure Protocols (IMP) are used for measures that do not exist in the TRM, and for additions that expand the applicability of an existing protocol, provided that the additions do not change the existing TRM algorithms and deemed savings values. IMPs serve as a holding ground before a protocol is fully integrated into the TRM via the annual update process.

The TRM serves as a PUC-approved resource to be used to determine, claim, and verify savings that count toward the statutory savings targets for the purposes of Act 129. The PUC recognized that it was necessary to expand the TRM by adding more savings protocols from the early stages of Act 129.[44] The 2009 TRM Order directed that the TRM be updated on an annual basis, which would allow new savings protocols for all measures offered through the EDC programs to be added to the TRM each year.[45] However, EDCs expressed concern that a once-a-year update would limit the EDCs' abilities to offer additional measures that could achieve energy and demand savings cost-effectively. In order to address this concern, interim protocols were introduced.

IMPs exist primarily to allow EDCs to claim ex ante and verify ex post savings using deemed or partially deemed protocols for measures that do not have current TRM protocols. The SWE will maintain a catalog of IMPs, showing their effective dates on the SWE SharePoint site, in order to maintain a database for new/revised measure protocols that should be included in subsequent TRM updates, and for EDCs to use to claim ex ante savings and for evaluators to follow when determining ex post savings.

[44] 2009 TRM Order, Page 10.
[45] Ibid. Page 17-18.

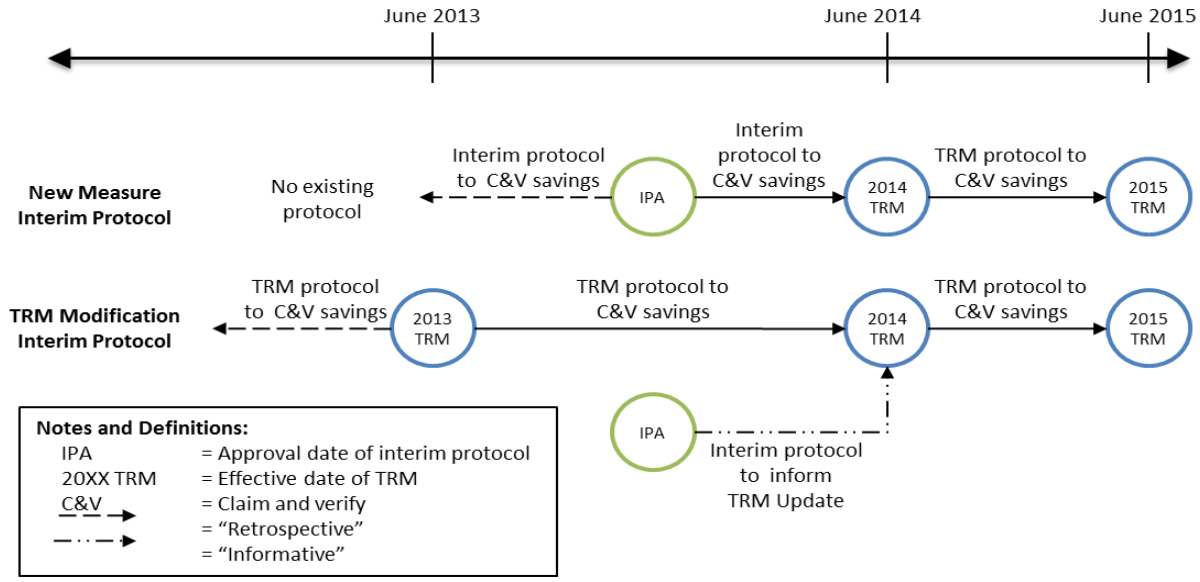
2.3.6.1 Interim Protocol Approval Process

The IMP approval process is informal, and is intended to minimize risk for EDCs planning to offer measures that do not have a TRM protocol by developing savings protocols through a collaborative review process in the PEG. The IMP review and approval process includes the following steps:

1. EDCs submit IMPs to the SWE.
2. The SWE reviews a proposed IMP and returns any suggested revisions to the submitting EDC.
3. After discussion and revision, the SWE sends the IMP to the other EDCs for comment.
4. After an IMP undergoes an iterative review process between the SWE and the PEG, the SWE gives the protocol interim approval as an "interim approved TRM protocol."
5. Interim approval is formalized when the SWE confirms approval via email and posts the final protocol and its effective date on the SWE SharePoint site.
6. The SWE includes all IMPs in the next TRM update for public comment and review, and formal approval.

The effective date of IMPs depends on the nature of the protocol. Two types of protocols have been identified: *new measure interim protocols* and *TRM modification interim protocols*. The SWE determines the appropriate classification of each proposed protocol and announces when the protocol is approved. The effective dates are shown in Figure 2-2.

Figure 2-2: Effective Date of Interim Protocol by Protocol Type



2.3.6.1.1 New Measure Interim Protocols

This category of interim protocols refers to completely new measures or additions that expand the applicability of an existing protocol, provided that the additions do not change the existing TRM algorithms, assumptions, and deemed savings values. For new measures, an approved IMP will apply for

the entire program year in which it was approved. This IMP will be included in the next TRM update for PUC approval. The PUC-approved TRM protocol, whether changed or unchanged, will apply prospectively; a PUC-approved TRM protocol will not apply retrospectively, unless the PUC formally approves a request by an EDC to do so.

2.3.6.1.2 TRM Modification Interim Protocols

This category of interim protocols refers to modifications of existing TRM protocols. This category includes proposed additions to an existing TRM protocol that modify the existing TRM algorithm, assumption, and/or deemed savings values. Because a TRM Modification Interim Protocol is created to amend an existing, PUC-approved TRM protocol, it cannot override that associated PUC-approved TRM protocol. The TRM Modification Interim Protocol will be used to inform the next TRM update. When the PUC approves the proposed Modification Interim Protocol, the protocol will be effective prospectively, although EDCs may submit a formal request to the PUC for retrospective application of the protocol.

2.3.7 Custom Measures

While TRM measures are reviewed and approved by the PUC through the TRM update process, custom measures do not undergo the same approval process. This section describes a process for managing custom measures by establishing a method for documenting energy and demand savings; describing the general requirements for custom measures; and clarifying the roles of the EDCs, ICSP, evaluation contractor, and SWE Team.

EDCs may report ex ante savings for a custom measure according to methodologies used by the customers or contractors and approved by the EDC ICSP. EDCs are not required to submit ex ante savings protocols for custom measures for SWE approval. EDC ICSPs must perform measurements consistent with IPMVP options to collect baseline and/or post-retrofit information for custom measures that have estimated savings above a threshold kWh/yr level.[46] They are not limited from performing measurements for custom measures with estimated savings below the threshold. To reduce the likelihood of significant differences between ex ante and ex post savings, EDC evaluation contractors are encouraged to recommend the IPMVP option and M&V protocols to be used by the ICSP.

[46] The SWE notes that measurement thresholds discussed here apply mainly to installations of energy efficiency measures in commercial or industrial facilities. The threshold kWh/yr is recommended by the SWE to separate large and small custom projects. The EDCs should develop the threshold and propose them to the SWE for approval. Projects below the threshold may include a baseline assessment, depending on the type of measure. The threshold may vary for different project types. The timeframe for EDCs to submit thresholds for SWE approval and final thresholds by EDC must be described in a separate data request and a guidance memo.

The PUC will not determine M&V protocols for custom measures to improve the EDCs' ability to support energy services that meet the EDCs' energy savings goals. EDC evaluation contractors are permitted to determine the appropriate M&V protocols for each project. EDC evaluation contractors must verify impacts for custom measures selected in the verification sample. They must develop an appropriate Site-Specific Measurement and Verification Plan (SSMVP) for each sampled project, per their professional judgment. EDC evaluation contractors must verify the project-specific M&V data (including pre and post metering results) obtained by the ICSPs, as practicable, for projects in the evaluation sample.

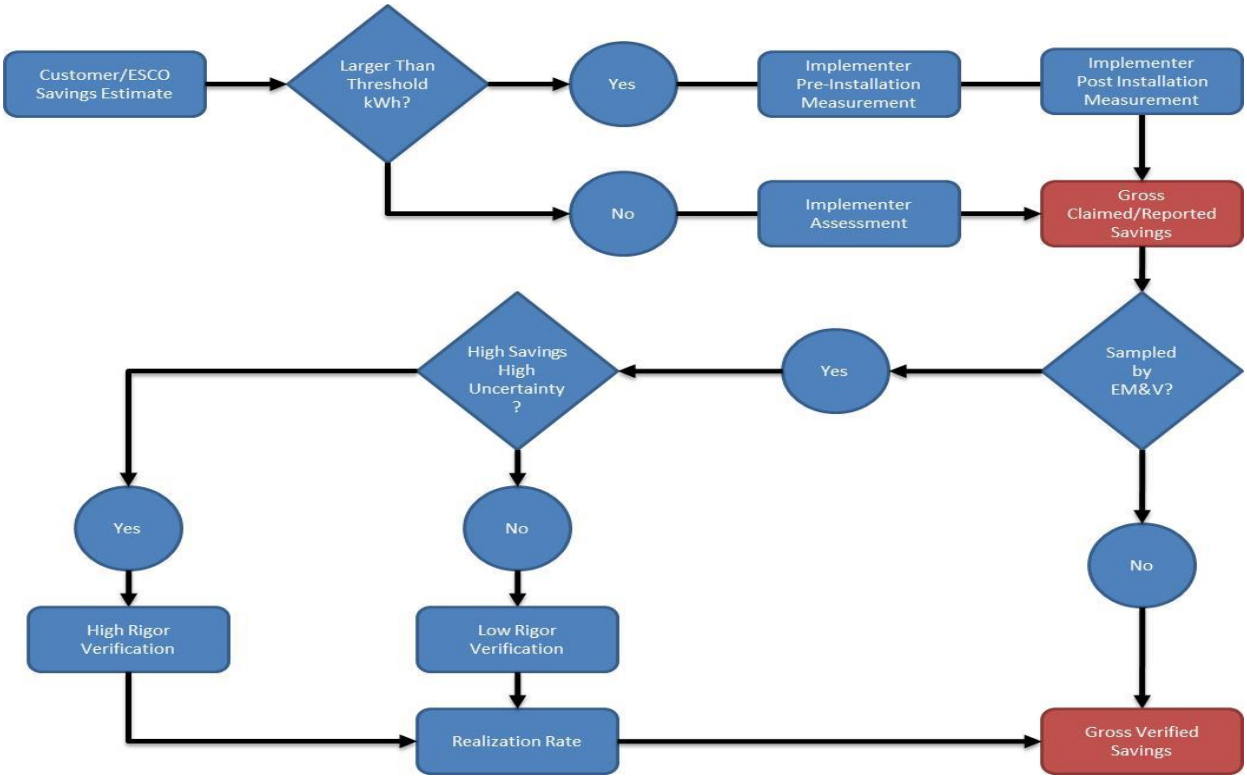
If the evaluation contractor determines that data collected by the ICSPs are not reasonably valid, then the evaluator must perform measurements consistent with IPMVP options to collect post-retrofit information for custom measures that have estimated savings above a threshold kWh/yr level. The evaluation contractor must make baseline assessments in the most efficient and cost-effective manner, without compromising the level of rigor. It is strongly recommended that ICSPs reach out to evaluation contractors to ensure that baseline assessments are being conducted in an acceptable manner and that all necessary data points are being collected for the estimation of savings.

The SWE reserves the right to audit and review claimed and verified impacts of any custom measures or projects. The SWE will randomly choose projects sampled by the EDC evaluation contactors and will audit the evaluators' engineering analysis and realization rates. In addition, the SWE also may select a random sample of projects not sampled by the EDC evaluation contractors and conduct an independent assessment of the ex post savings. The SWE may use these independent samples to augment the sample selected by the EDC evaluation contractors. The results from SWE independent assessments may be included in the program's realization rate calculations at the discretion of the EDC evaluation contractor.

Figure 2-3 presents a flow chart of the generic process to verify savings for custom measures. Deviations from the process are acceptable.[47]

[47] For example, not all projects above the kWh/yr threshold will require baseline measurements. Some may require only post-retrofit measurement.

Figure 2-3: Custom Measure Process Flow Chart



2.3.7.1 Mass Market Protocol

Certain mass market programs in the residential sector are a subset of custom measures. These programs offer measures, or groups of measures, that are not included in the TRM. During Phase I of Act 129, the Program Evaluation Group developed Mass Market Protocols (MMPs) for calculating the energy and demand savings associated with residential behavioral modification and low-income weatherization programs. MMPs approved during Phase I are considered available for use in Phase II by EDC ICSPs and evaluation contractors. Section 3.3.3.3 of this Framework provides additional guidance regarding the gross impact evaluation of mass market programs that do not have TRM savings protocols. Approved MMPs are provided in Appendix D.

2.4 Guidance Memos

This Evaluation Framework is developed to provide an overarching framework for Act 129 programs and therefore may not address all nuances discovered through the actual implementation and evaluation process. For such issues, the SWE will develop guidance memos to clarify and memorialize decisions through an iterative review process with input from EDCs and their evaluation contractors and the BTUS staff. These guidance memos will be the last step in resolving open issues and will formalize high-level decisions that impact all EDCs.

The SWE will post all PUC-approved guidance memos with their effective dates in the Phase II folder on the SWE SharePoint site. All of the guidance memos issued by the SWE in Phase I have been incorporated into this Evaluation Framework. Neither guidance memos nor SWE documents or positions necessarily reflect the opinions, regulations, or rulings of the PUC and, therefore, are not binding on the PUC.

Recently completed memos discussing net-to-gross issues will be reclassified as guidance memos. All memos released by the SWE will be classified as guidance memos in the future.

On an annual basis, the SWE will review and retire any guidance memos that become obsolete.

2.5 Study Memos

It may be necessary to conduct evaluation-related research studies to support the program design or evaluation analysis efforts. Study memos outline a specific research topic for the SWE to investigate. The SWE will work with the EDC teams to identify the need for any near-term and long-term research studies. These collaborative efforts will minimize redundant, independent research and reduce costs. The SWE will collaborate with EDCs primarily through collection of data from previous implementation and evaluation activities. BTUS staff is responsible for approval of study memos. Studies are considered to be an extension of the TRM update process. Results from these studies are intended to inform updates of the TRM.

As the research studies are identified and approved for implementation, all activities will be completed under existing budgets, unless otherwise noted. The SWE will distribute study memos to EDCs for information purposes.

3 Technical Guidance on Evaluation, Measurement, and Verification (EM&V)

This section of the Evaluation Framework is intended to help guide EDC evaluation contractors in the development and execution of successful evaluation plans. Section 3.1 contains the SWE's recommendations and requirements for evaluation plan development. Each efficiency measure that is implemented as part of an EDC's EE&C plan is assigned a reported (ex ante) impact estimate for energy and demand savings. These ex ante savings values are usually generated by an ICSP retained by an EDC to administer a specific EE&C program and associated efficiency measures. Determination of the ex ante savings values is based primarily on TRM protocols; this is discussed in Section 3.2.

The sum of the savings reported (through program tracking databases and systems) by the EDC and/or its ICSP is the gross reported savings for the EE&C program. However, compliance with Act 129 savings targets is based on gross verified savings estimates. In order to develop these estimates for a program, an EDC's evaluation contractor selects a sample of projects from the program population for verification of the ex ante savings estimate, which may include more rigorous measurement and verification activities than those used to prepare the reported savings estimates. These measurement and verification activities are discussed in Section 3.3.

A sample typically is used because it is not feasible or cost-effective to evaluate each of the hundreds or thousands of efficiency measures implemented. Section 3.4 presents the annual evaluation sampling requirements at the portfolio, sector, and program level, and offers technical guidance on sample design, allocation of resources, and presentation of the uncertainty introduced by sampling on gross verified impacts. Section 3.5 describes other sources of uncertainty in an evaluation and how evaluation contractors should address these factors.

3.1 EDC Evaluation Plans

Planning is a critical first step in successful program evaluation. The evaluation plan, or EM&V plan, outlines the approaches the evaluator will use and serves as a guiding document for the evaluation. EDCs must complete an initial evaluation plan for each program and submit it to the SWE SharePoint site for review within 90 days of the program year's start date (by August 31). The evaluation plan should be a single electronic document with a chapter for each program in the portfolio, or a separate document for each program. Within two weeks of this submission, the SWE Team will either approve the plan or suggest modifications to it. If the SWE Team suggests modifications, the EDCs will have two weeks to submit revisions based on the SWE comments and submit a revised evaluation plan. Then the SWE Team will have two weeks to provide final comments or approve the revised plan. Either party may request a time extension if unforeseen circumstances arise.

Changes to program delivery and evaluation approaches can occur from one year to the next within a program phase (such as Phase I, Phase II, or Phase III). The SWE Team recommends that EDCs submit a redline version of the evaluation plan for Program Year 6 and Program Year 7, or whenever intra-year changes are required. Evaluation plan updates will undergo the same review process as the initial evaluation plan for a phase of the Act 129 programs. Evaluation contractors are encouraged to submit evaluation plan modifications to the SWE as early as possible in the program year.

Each EDC and its evaluation contractor will choose the optimal structure and design for their evaluation plans. The evaluation plan should at least reflect a shared understanding of the program delivery mechanisms, research objectives and methodology, data collection techniques, site inspection plans, and intended outcomes. Evaluators should discuss the gross impact evaluation, NTG analysis, process evaluation, and cost-effectiveness evaluation activities and outcomes separately. Evaluation plans also should contain a proposed timeline of activities and a table of key program contacts. Evaluation plans should identify who will conduct site inspections (the EDC, the EDC ICSP, the EDC's evaluation contractor, or some other entity). Evaluations plans should also explain how the EDCs will make site inspections results available to the SWE Team. Sections 3.2 through 3.8 provide technical guidance to the EDC evaluation contractors regarding evaluation plans and activities for Phase II of Act 129.

The PA TRM provides EDCs with open variables for a number of energy conservation measure (ECM savings parameters. Often, a default value is provided as an alternative to customer-specific or program-specific data collection. An EDC evaluation plan should identify open variables for which the ICSP or evaluation contractor intends to utilize the option of "EDC data gathering." The SWE expects the results of these data collection efforts to be used in the calculation of verified gross savings, even if the

resulting savings differ from the impacts calculated from using the default value. However, if the SWE or evaluation contractor determines that the primary data are unreliable or significantly biased, the default value may be used.

3.2 Reported Savings

3.2.1 Tracking Systems

For the EDC evaluation contractors to evaluate programs, it is imperative that EDCs maintain complete and consistent tracking systems for all Act 129 programs. The tracking systems should contain a central repository of transactions recorded by the various program implementers capable of reporting ex ante savings quarterly. The values in the tracking system should be used for reporting ex ante energy and demand savings, customer counts, and rebate amounts in the EDC quarterly and annual reports. Records stored in EDC tracking systems also should be the basis of the evaluation contractor's sample selection processes and contain project parameters relevant to the savings calculation for each installed measure.

The SWE should be able to replicate summations from the tracking systems and match the summed savings value for a program and a program portfolio to the corresponding values in the EDC quarterly and annual reports.[48] EDCs must ensure that the tracking system contains all of the fields that are required to support calculation and reporting of program ex ante savings.[49]

3.2.2 Installed Dates, In-Service Dates, Recorded Dates, Reported Dates, and Rebate Dates

An EDC tracking system must capture several important dates:

- **Installed Date**: The date at which the measure is physically installed and operable; this may or may not coincide with the In-Service Date.
- **In-Service Date** (ISD, also referred to as the "Commercial Date of Operation" or CDO): The date the measure is installed and commercially operating as intended for long term savings. This is the date at which savings begin to be realized by the customer and may be the same as the Installed Date or later.
- **Recorded Date**: The date the measure is entered into the program system of record for future reporting to the PUC. This does not refer to the submission date of a quarterly or annual report.
- **Reported Date**: The date on which savings for a given project are officially submitted to the PUC as part of an annual compliance report. The gross reported and gross verified savings values for a program quarter or program year are the sum of the measures with a Reported Date within the quarter or program year; this does not refer to the submission date of a quarterly or annual report.

[48] Cumulative savings for a time period, especially Cumulative Program Inception to Date (CPITD), may not exactly equal the sum of transactions, quarters, or program years due to expired savings (measure installed during a phase whose measure life expires during that phase) adjustments to transactions, and other factors.
[49] Some worksheets used in the calculation of individual customer impacts will not be embedded in the tracking system.

- **Rebate Date**: The date the program administrator issues a rebate to the participant for implementing an energy efficiency measure; this may be substituted with an "Approval Date," which is the date a rebate is approved for payment within an implementer's system, if there is a time delay between approval of a payment and issuance of the rebate/incentive.
- **Filed Date:** The date an EDC officially submits and files a quarterly or annual report to the PUC as part of a compliance requirement.

In Phase I, an issue was identified related to reporting energy savings and more specifically, *reporting lags*. *Reporting lag* occurs when the savings for a transaction are reported in a later quarter/year than the quarter/year the measure went in-service. For example, a measure may go in-service in PY5 but not be recorded or reported until PY6. There are two types of reporting lags: participant lag and approval lag.

- *Participant lag* describes the time between when a participant buys and installs a measure and submits the associated rebate application to the program administrator; this can be as brief as a few days or as long as six months. This lag largely depends on participant behavior and program policies.[50]
- *Approval lag* describes the time between when a customer submits a rebate application and the program administrator approves the application; this will vary by program and project, and stems from key program processes such as application review, QA/QC procedures, installation verification, and rebate and invoice processing. Approvals of program transactions are guided by EDC communications related to eligibility and deadlines for program application submittal. Similar processes exist for upstream buy-down programs that require time for retailers and manufacturers to compile finalized sales documentation.

The SWE has defined a process for dealing with the two types of reporting lag as related to reporting to the PUC. EDCs are directed to file preliminary annual reports on July 15 and final annual reports on November 15 following the end of the program year[51] using the existing reporting structure, which accounts and works well for all projects with reported dates (and therefore in-service dates) prior to the statutory target date. EDCs opting to account for lagged transactions that have a recorded date after the statutory target date, but an in-service date prior to the statutory target date, must provide a supplemental report with the final verified savings of lagged transactions by the Q2 reporting deadline of the program year following the measure's in-service date.[52] EDCs should include another table representing kW savings.

[50] Act 129 and Orders approving programs recognize savings for measures installed after a specified date. Different programs and program managers may have policies and communications that can impact customer lag.
[51] Secretarial Letter entered on May 25, 2011, at Docket No. M-2008-2069887.
[52] Lagged transactions technically are part of later reporting periods, and therefore should not be portrayed as part of current reporting periods by including them in the actual reports.

Additionally in Phase I, there was uncertainty regarding which date governs the version of the TRM an EDC should use to determine savings. To eliminate such confusion, the SWE and TUS staff agreed that the applicable date for determining which TRM to use (for all measures, excluding new construction) is the in-service date. The TUS staff and the SWE concluded that the in-service date is the correct date to use because it marks the date when the customer starts to realize savings and ensures that savings calculations match the date when they begin to accrue. ICSPs and evaluation contractors should use the TRM in effect at the in-service date when calculating energy and demand savings for Phase II. For new construction, selection of the appropriate TRM must be based on the date when the building/construction permit was issued (or the date construction starts if no permit is required) because that aligns with codes and standards that define the baseline. Savings may be claimed toward compliance goals only after the project's ISD. This requirement is to account for the long lifecycle of new construction projects that are designed to a particular standard prior to construction.

3.2.3 Historic Adjustments

EDCs are required to document any adjustments made to ex ante savings after a quarterly/annual report and quarterly/annual data request response has been submitted. Any change to the reported kWh impact, reported kW impact, or rebate amount for a claimed project is considered a historic adjustment. The SWE understands that such adjustments must be made to correct errors, or reflect better information, but requires that the EDC inform the SWE of these historic adjustments prior to the submission of the EDC Final Annual Report. This process will allow the SWE to update its records and track program progress using the corrected values. Two acceptable methods for submitting these historic adjustments are:

1. **Record replacement** – This technique involves submitting two new records for the measure being revised. The first record will be the inverse of the original tracking record submitted to the SWE (negative kWh, kW, and incentive amounts) and will serve to "zero out" the original values submitted. The second record should contain the corrected project impacts.
2. **Record revision** – This technique involves submitting a single record containing the adjustments to project parameters. For example, if the original measure record contained an impact of 1,300 kWh and it was later discovered that the correct gross reported savings value for that measure is 1,650 kWh, the new tracking record would contain a reported kWh value of 350 kWh.

With either approach, the EDCs should identify historic adjustments using an indicator variable set equal to 1 for adjustment record and equal to 0 for a new tracking record. This indicator variable is needed to produce accurate participation counts by quarter or program year because a project receiving historic adjustments should not be included when determining the participation count for the program (because it was counted previously). If an EDC has an alternate methodology for informing the SWE of historic adjustments to ex ante impacts that is not listed in this section, the approach can be submitted to the SWE Team for consideration and approval.

3.2.4 Key Fields for Evaluation

Because the EDC evaluators use equations to independently calculate verified savings for some partially deemed TRM measures, the SWE requires that the EDCs provide key variables used to calculate savings to the evaluator. The EDC's ICSP should collect these variables so the evaluator will not have to retrieve the variables independently for projects outside of the evaluation sample. For projects in the evaluation sample, it is the evaluation contractor's responsibility to independently verify each parameter in the savings calculation. This requirement also will improve the transparency of the savings calculation process. For example, to calculate savings for residential central air-conditioning equipment using the 2014 Pennsylvania TRM, the ICSP must provide the following fields:

- Cooling capacities (output in Btuh) of the central air conditioner installed
- Seasonal Energy Efficiency Ratio of the baseline unit[53]
- Seasonal Energy Efficiency Ratio of the qualifying unit being installed
- Location of the home so that the default Equivalent Full Load Hours of operation during the heating and cooling seasons can be incorporated into the savings calculation

3.3 Gross Impact Evaluation

3.3.1 Overview

This section establishes guidelines for all evaluation contractors that conduct gross impact evaluations. Impact evaluations determine program-specific benefits, which include reductions in electric energy usage, electric demand, and avoided air emissions[54] that can be attributed directly to an energy efficiency program. As there are many stages to an impact evaluation, decisions must be made at each stage based on the desired accuracy and certainty of the evaluation results and the funds available. Section 3.3 provides evaluators information to support decision-making throughout the gross impact evaluation process.

For C&I programs, impact evaluation contractors use data collected during program implementation and conduct independent data-gathering activities. If the data collected by the ICSP are unreliable, if end-use equipment operating conditions have changed post-installation, or if the ICSP did not conduct or complete project-specific data collection activities for a project with high informational value, the evaluation contractor(s) must collect the appropriate data for sampled projects. The EM&V activities may include surveys or direct observation and measurement of equipment performance and operation at a sample of participant sites to verify that the energy savings reported for the projects are correct and that the equipment is installed and operating. Successful impact evaluations assess the costs incurred with the Value of Information (VOI) received and balance the level of evaluation detail ("rigor" as defined in Section 3.3.2.2) with the level of effort required (cost). How deeply an evaluator goes into the

[53] This assumes that an "early replacement" savings protocol is followed.

[54] While EDCs are not required to report air emissions in EE&C program impact evaluations, estimates of emission reductions can be estimated easily, based on verified gross energy savings and emissions factors from sources, such as PJM, the Energy Information Administration, and the Federal Energy Regulatory Commission.

assessment of key variables at a sampled site or among program participants depends on the value of that information in confirming the claimed savings.

For residential programs, approved impact evaluation methods for the Act 129 residential-sector programs have evolved over the course of Phase I and II of the Pennsylvania Act 129 programs. The Act 129 residential programs are mostly mass-market programs that involve proven and well-tested technologies marketed to most or all households in a service area. As a result, ex ante estimates of gross program savings usually can be calculated using algorithms listed in the applicable Pennsylvania TRM, Interim Measure Protocols (IMP), Custom Measure Protocols (CMP), or Mass Market Protocols (MMP). Basic levels of rigor are typically applied when verifying residential measures. EDC implementation contractors or EDC evaluators then conduct inspections or desk audits of a random sample of installations to determine if measures are installed and operating. Verified gross program savings are then calculated based upon the results of the verification activity. Listed below in Table 3-1 are the general impact evaluation approaches recommended for Pennsylvania Act 129 residential energy efficiency programs.

Table 3-1: Recommended Residential Impact Evaluation Approaches

Act 129 Residential Program by Type	Recommended Impact Evaluation Approach (Verification can occur by site visit or desk audit)
Residential Lighting	Desk audit of invoices for mass-market lighting programs; verification of measure baseline; cross-sector sales surveys; annual kWh and kW savings based on algorithms in the PA TRM
Residential Appliances and Other Efficient Products	Verification of measure purchase for a random sample of program participants to establish installation rate; annual kWh and kW savings based on algorithms in PA TRM
Refrigerator/Freezer Recycling	Verification of appliance removal for a random sample of program participants to establish removal rate; annual kWh and kW savings based on algorithms in PA TRM
Low-Income	Verification of energy efficiency measure installation via onsite QA/QC inspections for a sample of program participants to establish in-service rate; annual kWh and kW savings based on algorithms in PA TRM or based on a statistical billing analysis
Behavior/Education	Annual savings calculated according to latest approved MMP

According to the hierarchy within the process of implementing and evaluating EDC programs, the TRM savings protocols for efficiency measures define how ICSPs generally will calculate the ex ante savings. The impact evaluation protocols are the procedures the EDC evaluators must follow to verify the energy

and demand savings claimed by the ICSPs as defined in this Evaluation Framework. Open communication between ICSPs and evaluation contractors helps reduce or eliminate redundant data collection efforts when appropriate (Section 0). The SWE follows the SWE audit protocols (Table 3-2) to audit and confirm the evaluation contractor's verified impacts.

Table 3-2: Impact Evaluation Hierarchy (General Approaches)

Level	Description	Requirements/ Discussion Points
1. Program Implementation • Deemed savings • Partially deemed savings • Custom/unspecified	M&V protocols and Site-Specific M&V plans (SSMVP) used to calculate ex ante savings • TRM protocols • Interim TRM protocols • SSMVP methodologies	1. TRM protocols are used for measures specified in the TRM. 2. For custom measures, ICSPs should submit documentation that is reviewed by the EDC or the EDC's evaluation contractor. 3. Ex ante estimates of gross program savings usually can be calculated using algorithms listed in the applicable Pennsylvania TRM, Interim Measure Protocols (IMP), Custom Measure Protocols (CMP), or Mass Market Protocols (MMP). Basic levels of rigor are typically applied when verifying residential measures.
2. EDC Impact Evaluation	EDC evaluation contractor samples projects (for those measures where sampling is used) and calculates ex post verified savings with the appropriate EM&V protocol (TRM, IMP, CMP, or MMP).	1. Statistical sample of participants analyzed 2. Field engineering and project-specific analysis 3. Calculation of ex post verified energy savings 4. Calculation of realization rates
3. SWE Audit Activities	SWE works with EDC and evaluation contractor to audit and ensure accuracy of reported savings and verified savings or conducts independent analysis, if needed	1. Coordination with EDC impact evaluation activities (e.g., upstream CFL invoice reviews, joint site-visits) 2. Independent verification activities. (e.g., appliance recycling work order review, site visit with field verification) 3. Recommendations to adjust realization rates and corresponding ex post verified energy savings

3.3.2 Calculating Verified Gross Savings

One of the primary research objectives of an impact evaluation is to calculate gross verified savings, which are the savings achieved by the program as calculated by an independent third-party evaluator. Evaluation contractors should produce an independent estimate of program energy and demand impacts according to the appropriate savings protocols described in the SWE-approved EM&V plan. In most cases, the evaluator and ICSP will use the same savings protocol, so the evaluator's duties may be characterized as *verification*. Evaluators should verify that an appropriate level of measurement rigor was employed by the ICSP, and if needed, conduct independent end-use level measurements for high-impact and high-uncertainty projects. For program evaluations that rely on sampling, these independent estimates should be compared to the claimed savings for a sample of sites within each program to calculate a *realization rate*. This realization rate should then be applied to the population of participants to determine the *verified gross savings*. When appropriate, the collective results of these EDC impact evaluations also will be used to inform updates to the kWh/yr and kW savings in the TRM so that the TRM reflects the latest available information on measure and program savings. The following subsections provide detailed guidance for EDC evaluators for calculating verified gross savings for impact evaluations.

3.3.2.1 Measure Type

Most of the savings anticipated by the Act 129 programs should be estimated and verified through methods described in the TRM. As noted in Section 0, each of the three measure categories (deemed, partially deemed, and custom) dictates use of specific M&V activities. Additionally, the approach to verifying savings should be clear, technically sound, and based on accepted industry standards. The quantification of savings is both an art and a science, as energy savings are the difference between energy that would have been used without the measure and energy that actually was used. In practice, engineering, empirical science, and reasonable assumptions need to be used to estimate what "would have been used" because this value cannot be measured.

A large portion of these savings are either: 1) deemed based on units installed, sold, or given away, or 2) partially deemed and subject to assumptions relative to the performance of the technologies and how the technologies are used. Though metering studies and detailed analysis are encouraged to inform updates of TRM savings protocols, EDC evaluation contractors must verify fully deemed measures with TRM protocols by using TRM protocols and assumptions. Metering or project-specific data collection activities may be required for partially deemed measures with greater variance in end-use operating parameters and custom measures.

3.3.2.2 Level of Engineering Rigor

The level of engineering rigor is defined as the level of detail involved in the verification of the EDC-reported impacts and defines the minimum allowable methods to be used by the EDC evaluation contractors to calculate ex post savings (verified gross savings). This Evaluation Framework establishes a minimum level of detail to ensure that the verified gross savings are at the level of accuracy needed to support the overall reliability of the savings in reference to statutory savings targets. The Framework also provides guidelines on the evaluation methods the evaluation contractors must use for specific

evaluation groups. These groupings consist of multiple programs (program components/measures) having common characteristics that provide evaluation efficiencies in the contracting, supervision, and implementation of evaluation efforts.

The Evaluation Framework defines two levels of rigor: *basic* and *enhanced*. Each level of rigor provides a class of minimum allowable EM&V methods, based on standard evaluation practices, in order to offer flexibility for the evaluation contractors to assess and propose the most accurate and cost-effective methods to verify gross savings while balancing cost and rigor. The choice of basic rigor versus enhanced rigor will depend on the type of measure, relative complexity of savings calculations, level of uncertainty, and most importantly, savings impact. Generally, evaluation contractors are allowed to choose the appropriate level of rigor, as long as they follow the guidelines in this section, including the exceptions listed by impact stratum shown in Table 3-4. Further, the SWE reserves the right to challenge the level of rigor planned by the evaluation contractors and request revision of the verification technique prior to the evaluators' site visit, if necessary. After the site visit, the SWE may recommend revisions to the level of rigor or verification technique to be used on similar future sampled sites.

Table 3-3 provides guidelines regarding the *minimum* allowable methods associated with the two levels of rigor. Evaluators are highly encouraged to collect additional data that may be useful for determining the necessity of future TRM updates that improve the accuracy and reliability of savings protocols.

The EM&V options defined under each level of rigor provide independent evaluators cost-effective methods to verify program impacts without compromising the accuracy of the reviews. In general, the TRM fully deemed measures will follow a basic level of rigor, while custom measures will typically follow an enhanced level of rigor.[55] The TRM partially deemed measures will follow either a basic or enhanced level of rigor, depending on the type of measure, exceptions noted by impact stratum, and level of impact. These paths are depicted in Figure 3-1, which provides guidance on choosing the level of rigor by measure type.

[55] Low-impact and low-uncertainty custom measures may use a basic level of rigor.

Figure 3-1: Expected Protocols for Impact Evaluations

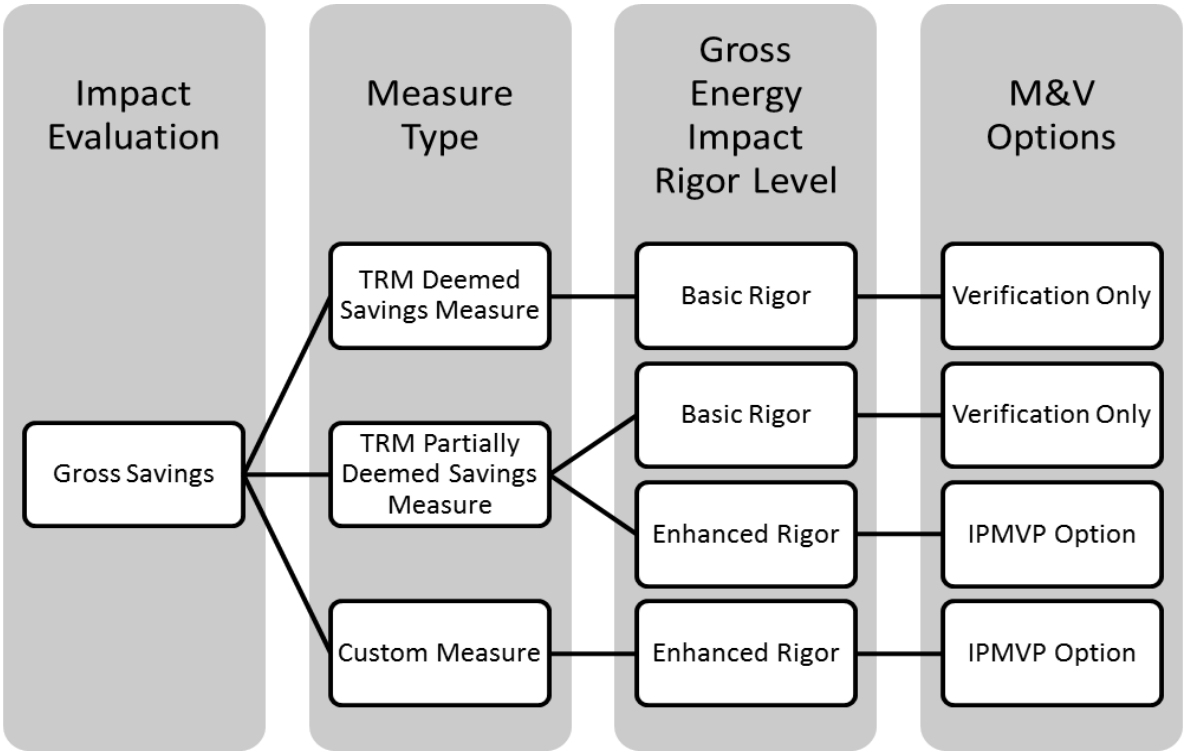


Table 3-3: Required Protocols for Impact Evaluations

Rigor Level	Minimum Allowable Methods for Gross Impact Evaluation
Basic	1. Verification-only analysis for TRM fully or partially deemed measures with impacts below the threshold established in the TRM for requiring customer-specific data collection. Verification of the number of installations and the selection of the proper deemed savings value from the TRM. 2. Verification of appropriate application of the TRM savings algorithms for TRM partially deemed measures using gathered site data that typically is limited to performance specification data and does not need to be measured onsite.
Enhanced	1. Simple engineering model with EM&V equal to IPMVP Option A for TRM partially deemed measures. Required for impacts above the threshold in the TRM. When the TRM specifies an algorithm, this approach includes verification of the appropriate application of TRM savings algorithms and corresponding site-specific stipulations as required and allowed by the TRM. Spot measurement and site-specific information can be obtained by the implementer and verified by the evaluation contractor, or obtained by the evaluation contractor directly. 2. Retrofit Isolation Engineering methods as described in IPMVP Option B. 3. A regression analysis (IPMVP Option C)[56] of consumption information from utility bills with adjustments for weather and overall time period reported. The SWE Team recommends that at least twelve (12) months of pre- and post-retrofit consumption be used when practicable, unless the program design does not allow for pre-retrofit billing data, such as residential new construction. In these cases, well-matched control groups and post-retrofit consumption analysis are allowable. 4. Building energy simulation models as described in IPMVP Option D.

For partially deemed measures that require project-specific data collection and custom measures, it is recommended that the ICSP follow a similar approach to collect this information during application processing or the rebate approval process. The impact assessment methodologies used by the ICSPs and evaluation contractors should be aligned to increase the correlation of ex ante and ex post savings estimates to improve the precision of evaluation results. Evaluation contractors can leverage information collected by the program ICSPs in cases where it would be burdensome to the participant for the evaluation contractor to gather information, such as end-use metering, independently. Evaluators should exercise their professional judgment in testing the credibility and validity of the measurements gathered by ICSPs. The SWE reserves the right to challenge the evaluators' assessment of the ICSP data and may conduct independent measurements for any project in the population.

[56] Further information on statistical billing analysis is available in *Uniform Methods Project: Methods for Determining Energy Efficiency Savings for Specific Measures*, Chapter 8: Whole-Building Retrofit with Consumption Data Analysis Evaluation Protocol.

The following section provides additional detail on the basic and enhanced levels of engineering rigor to assess ex post savings for energy and demand impacts.

3.3.2.2.1 Energy – Basic Rigor Option 1: Verification-Only Analysis

The first class of allowable methods for basic rigor is a verification-only analysis. This analysis applies mainly to the TRM fully deemed measures, but also may be used for TRM partially deemed measures with impacts that have low uncertainty and are below the threshold established in the TRM for requiring customer-specific data collection. The objective is to confirm that measures actually are installed and operational, and the installation meets required standards. Installation verification should be conducted for a random sample of projects claiming energy savings. Verification may occur in person, over the phone, or via a review of project documentation. For each residential program, EDC evaluation plans should specify whether onsite inspections are planned, and if so, whether evaluation contractors or implementation contractors will conduct these inspections. Sampling of measures within a project and sampling at the program level for evaluation purposes should be specified according to the Sampling and Uncertainty Protocols described in Section 3.4.4.

3.3.2.2.2 Energy – Basic Rigor Option 2: Simple Engineering Model Without Measurement

The second class of allowable methods for basic rigor is a verification of the appropriate application of the TRM savings algorithms using documented site data without onsite measurement. If the ICSP collects the project-specific information, evaluation contractors should attempt to confirm the accuracy and appropriateness of the values. This option should be used for partially deemed measures producing savings above the threshold values[57] identified in the TRM as requiring customer-specific data collection, but which have low uncertainty.

3.3.2.2.3 Energy – Enhanced Rigor Option 1: Simple Engineering Model With Measurement

The first class of allowable methods for enhanced rigor is a Simple Engineering Model (SEM) with measurement of key parameters. An SEM is equivalent to IPMVP Option A. The IPMVP provides overall guidelines on M&V methods; however, more program- or technology-specific guidelines are required for the EDC programs. SEMs are straightforward algorithms for calculating energy impacts for measures such as energy-efficient lighting, appliances, motors, and cooking equipment (partially deemed measures). Several algorithms have open variables and require additional site-specific data or measurements. The TRM measure attributes that encourage project-specific data collection will be identified by providing the option of "EDC data gathering" in addition to a default value.

3.3.2.2.4 Energy – Enhanced Rigor Option 2: Retrofit Isolation Engineering Models

The second class of allowable methods for enhanced rigor is the retrofit isolation measurements, as described in Option B of the IPMVP. This method is used in cases where full field measurement of all parameters for the energy use for the system in which the efficiency measure was installed is feasible and can provide the most reliable results in an efficient and cost-effective evaluation. One typical

[57] Thresholds will apply only to nonresidential measures.

example where such a method would be appropriate is a lighting retrofit where both power draw and hours of operation are logged.

3.3.2.2.5 Energy – Enhanced Rigor Option 3: Billing Regression Analysis

The third class of allowable methods for enhanced rigor is a regression analysis of consumption data that statistically adjusts for key variables that change over time and are potentially correlated with consumption. As a way of capturing the influence of weather, evaluators may incorporate weather-normalized consumption as the dependent variable or include heating- and cooling-degree days, or another explanatory variable describing the weather, directly in the model. Other variables that often are correlated with consumption include: the state of the economy (recession, recovery, economic growth), fuel prices, occupancy changes, behavior changes (set-points, schedules, frequency of use), changes in operation, and changes in schedule. The EDC evaluation contractors are free to select the most appropriate additional variables to include.

3.3.2.2.6 Energy – Enhanced Rigor Option 4: Whole Building Simulation

The fourth class of allowable methods for enhanced rigor is building energy simulation programs calibrated as described in the Option D requirements in the IPMVP. The engineering models that meet the Option D requirements are building energy simulation models. This method can be applicable to many types of programs that influence commercial, institutional, residential, and other buildings where the measures affect the heating, ventilation, or air conditioning (HVAC) end use. This method often is used for new construction programs and building HVAC or shell upgrades in commercial and residential programs.

In addition, industrial projects can include changes in process operations where the appropriate type of model could be a process-engineering model. These are specialized engineering models and may require specific software to conduct an engineering analysis for industry-specific industrial processes. Where these types of models are more appropriate, the gross energy impact protocol allows for the use of a process engineering model with calibration as described in the M&V protocols to meet the enhanced rigor level.

3.3.2.2.7 Demand – Basic Rigor

The basic rigor level for the gross demand impact protocol prescribes that, at a minimum, on-peak demand savings be estimated based on the allocation of gross energy savings through the use of allocation factors, coincidence factors, or end-use load shapes during the hours of 2:00 p.m. to 6:00 p.m. on non-holiday weekdays (from June 1-August 31). For TRM deemed measures, TRM deemed coincidence factors are to be used. The use of TRM deemed coincidence factors should be applicable only to the TRM deemed and partially deemed measures that meet the requirements for basic rigor in Table 3-4. Custom measures should follow an enhanced rigor approach.

The SWE encourages EDC evaluation contractors to recommend improved coincidence factors values using a load shape from metered or vetted sources, when applicable, during TRM working group discussions. The SWE will consider the proposed values for prospective TRM updates. The SWE reserves

the right to request additional documentation to investigate the applicability of the load shapes submitted.

3.3.2.2.8 Demand – Enhanced Rigor

The enhanced rigor level for the gross demand impact protocol requires primary data from the program participants. These data could be interval-metered data, either from TOU consumption billing data (if appropriate), an EMS system, or field measurement. If the methodology and data used can readily provide an 8,760 savings profile, one should be calculated for the project.

For energy efficiency measures that produce savings during peak periods, end-use interval meter data, if available, should be used to construct pre- and post-retrofit peak-hour load shapes. The data should be adjusted for weather, day type, and other pertinent variables. If end-use interval meter data are not available, spot metering/measurement at peak pre- and post-retrofit should be conducted to assess impacts during non-holiday weekday afternoons from 2:00 p.m. to 6:00 p.m. during summer months (June 1-August 31). These data will be used with one of two engineering modeling approaches: 1) full measurement IPMVP Option B or 2) calibrated engineering model Option D, where the modeling approach must meet all requirements in the IPMVP protocol.

3.3.2.3 Level of Engineering Rigor Mapped to Program Stratification

The impact evaluation sample for nonresidential programs should be stratified based on the constituent projects' level of impact. The stratification method in this Evaluation Framework assumes three strata in programs with a large variety of rebated measures and associated variability of savings and potential impact. However, the stratification plan and level of rigor to be used in an evaluation will be determined and documented by the evaluation contractor. The actual number of strata used will be at the evaluation contractor's discretion and thus this section should be interpreted accordingly. Typically, Stratum 1 will include the projects with the highest impact and/or uncertainty measures, the lowest sampling weight, and the most rigorous evaluation expectations. Conversely, Stratum 3 includes the projects with the lowest impact and/or uncertainty measures, the highest sampling weight, and the least-rigorous evaluation expectations. Measures that fall into Stratum 2 require either basic or enhanced levels of rigor. If a specific measure meets one of the exceptions listed in Stratum 2 (shown in Table 3-3, below), an enhanced level of rigor is required. However, sound engineering judgment is necessary to determine the applicability of the exceptions to individual measures. Generally, flexibility is allowed in determining if these conditions are met; however, the SWE reserves the right to challenge the level of rigor used by the evaluation contractors and request revision of the verification technique for future evaluation plans.

Table 3-4: Definitions of Program Strata and Their Associated Levels of Rigor for Impact Evaluation of Nonresidential Programs

Stratum Level	Minimum Allowable Methods for Gross Impact Evaluation
Stratum 1 – High-Impact and/or High-Uncertainty Measures [58]	Enhanced rigor
Stratum 2 – Medium-Impact and/or High-Uncertainty Measures	Either an enhanced or basic level of rigor may be used, depending on the applicability of the exceptions listed in this table cell and the Value of Information. As a guide, enhanced rigor should be used if the measure meets one or more of the following criteria: 1. Irregularity of loads: a pattern does not exist sufficient enough to predict loads with ease and accuracy 2. Irregularity of operating periods: a pattern does not exist sufficient enough to predict operating periods with ease and accuracy 3. Savings consistency: a one-time "snapshot" assessment likely does not capture the savings over time (e.g., measures heavily dependent upon human interaction/control) 4. High probability of substantial variance in savings calculated from a default value in the TRM 5. Significant interactive effects, which are not already taken into account in the TRM, exist between measures. An interactive effect is considered significant if the EDC evaluation contractor suspects that inclusion of interactive effects in the impact estimates for the project has the potential to increase or decrease the energy or demand savings by more than 15%.
Stratum 3 – Low-Impact Measures	Basic rigor

3.3.3 EM&V Activities

This section provides a list of EM&V methods that are acceptable for verified savings estimation, separated per the level of engineering rigor discussed in Section 3.3.2.2. Appendix B provides detailed guidance by measure type for common nonresidential measures, and Appendix C offers detailed guidance by measure type for common residential program types.

[58] The EDC and evaluation contractor may determine the appropriate level of impact and uncertainty when stratifying measures. The EDC and evaluation contractor's discretion also includes determining the relative impact of programs within the portfolio when determining level of rigor to be used. For example, the "high-impact/uncertainty" stratum of a program with relatively lower savings may not require as rigorous evaluation activities as the "high-impact/uncertainty" stratum of a program with relatively much larger savings.

3.3.3.1 Basic Rigor EM&V Activities

3.3.3.1.1 Baseline Assessment

At a basic level of rigor, both early replacement and replace-on-burnout scenarios leverage TRM assumptions regarding the baseline equipment case. The EDC evaluator should verify that TRM assumptions are appropriate for the measure delivery option being evaluated.

3.3.3.1.2 Measure Installation Verification

The objectives of measure installation verification are to confirm that the measures actually were installed, the installation meets reasonable quality standards, and the measures are operating correctly and have the potential to generate the predicted savings during compliance years. At a basic level of rigor, phone interviews, combined with appropriate invoices and manufacturer specification sheets, may be used to verify the measure type. Additional guidance on key data points that should be collected for selected measures can be found in Appendix B and Appendix C.

During Phase II of Act 129, measure installation verification will follow the methodology set forth in the Market Potential Study conducted for Phase II. According to that methodology, if the evaluation contractor finds that a measure was uninstalled or not currently operating, but the ICSP reported that the measure was installed and correctly operating, appropriate savings shall still be allotted to the measure. In future Phases of Act 129, measure installation verification will continue to follow the methodology used in the corresponding Market Potential Study.

If the evaluation contractor finds that a measure is operating, but in a manner that renders the TRM values not directly applicable, TRM deemed values should not be directly applied and the evaluation contractor must incorporate the noted differences in savings calculations. When possible, measure design intent (i.e., the designed measure function and use and its corresponding savings) should be established from program records and/or construction documents. If the TRM values were applied incorrectly, the evaluator should recalculate savings using the correct TRM values applicable to the measure.

3.3.3.2 Enhanced Rigor EM&V Activities

3.3.3.2.1 Baseline Assessment

Where applicable and appropriate, it will be recommended to conduct pre-installation inspections to verify the existing equipment and gather the equipment baseline data in order to compute the partially deemed or custom savings estimates. The first objective is to verify that the existing equipment is applicable to the program under which it is being replaced. Additionally, the baseline equipment energy consumption and run-time patterns may be established to complete the engineering calculations used to estimate savings. At an enhanced level of rigor, early replacement existing equipment values should be verified by onsite inspection when possible, and replace-on-burnout existing equipment values should be based on local or federal minimum codes and standards.

3.3.3.2.2 Measure Installation Verification

As discussed in the basic rigor EM&V section, the objectives of measure installation verification are to confirm that the measures actually were installed, are operating correctly, and have the potential to generate the predicted savings during compliance years. Similarly, measure installation verification will follow the methodology set forth in the Market Potential Study conducted for Phase II. According to that methodology, if the evaluation contractor finds that a measure was uninstalled or not operating, but the ICSP reported that the measure was installed and correctly operating, appropriate savings shall still be allotted to the measure. In future Phases of Act 129, measure installation verification will continue to follow the methodology used in the corresponding Market Potential Study.

Evaluation plans should describe site inspections planned for residential and nonresidential programs. At an enhanced level of rigor, measure installation should be verified through onsite inspections of facilities. Equipment nameplate information should be collected and compared to participant program records as applicable. Sampling may be employed at large facilities with numerous measure installations. As-built construction documents may be used to verify measures, such as wall insulation, where access is difficult or impossible. Spot measurements may be used to supplement visual inspections, such as solar transmission measurements and low-e coating detection instruments, to verify the optical properties of windows and glazing systems. Appendix B and Appendix C contain additional guidance on key data points that should be collected for selected measures.

Correct measure application and measure operation should be observed and compared to project design intent. For example, for C&I, evaluation contractors should note CFL applications in seldom-used areas or occupancy sensors in spaces with frequent occupancy during measure verification activities, then modify hours-of-use categories appropriately. Further, if the evaluation contractor finds that a measure is not operating in the manner specified in the TRM, they should not apply the TRM deemed values directly, and they must incorporate the noted differences in savings calculations. For example, if the evaluation contractor discovers that a chiller is being used in an application other than comfort cooling, they should not use the TRM algorithm based on comfort cooling operating characteristics. In addition, they should obtain and review commissioning reports (as applicable) to verify proper operation of installed systems. If measures have not been commissioned, measure design intent should be established from program records and/or construction documents. Functional performance testing should be conducted, when applicable, to verify equipment operation in accordance with design intent.

3.3.3.2.3 Onsite Sampling of Installations

This section provides guidance in determining the number of installations to verify during the onsite inspection of a large project such as a lighting retrofit with several thousand fixtures within a facility. The methods explained below are not exhaustive, and evaluation contractors are encouraged to propose other options in their program evaluation plans.

The first method is to verify a census of all of the installations onsite. This activity is to be done in cases where a limited number of installations were made, or when the variance in operating parameters is large and impacts are high and need to be documented in combination with the verification activity of

the evaluation contractor. For projects where a visual inspection of each installed measure would require excessive time or facility access, a statistically valid sample can be used. Samples of measures selected for verification at a particular site should be representative of all measures at the site and should be selected at random. Measures within a building should be grouped according to similar usage patterns, thus reducing the expected variability in the measured quantity within each usage group. Within each usage group, the sampling unit should be the individual measure, with the goal being to verify the measure quantity recorded in the program tracking data.

When verifying installation quantities, the recommended relative precision for sampling onsite installations is ± 20% at the 90% confidence level at the facility level. The sampling unit (line item on the Commercial Lighting Retrofit Savings Calculator [TRM Appendix C form[59]], condensing unit, appliance, etc.) should be identified in the program evaluation plan for prescriptive measures or the Site-Specific Measurement and Verification Plan (SSMVP) for custom measures. The initial verification proportion (p) assumption for determining the minimum sample size for binary (fully deemed) outcomes should be set at 50% as this will maximize p*(1 – p) and guarantee that precision targets are met. For continuous outcomes, such as the number of fixtures within a space on the Commercial Lighting Retrofit Savings Calculator [TRM Appendix C form], a C_v of 0.5 is appropriate.

The sample, in general, should be representative of the population; this is where stratification will be of great use. Measures with similar operating characteristics and end-use patterns should be grouped into homogeneous strata and the sampling algorithm should be designed to achieve 90/20 confidence/precision for each facility. For example, lighting retrofits in common areas should be separated from those in individual suites in an office building, or air handler unit (AHU, such as a fan) motor retrofits should be grouped separately from chilled water pump replacements for C&I applications.

Since a certain degree of uncertainty is expected with any onsite counting exercise, an error band[60] should be specified within which the claimed installations or savings will be accepted. The SWE recommends using a maximum 5% error band. The error band should be calculated based on the sampling unit. If the verification counts for each usage group in the sample are within +/- 5% of the reported counts, the installed quantity should be accepted at the claimed value. For example, if the program tracking record for a project claims that 240 fixtures were retrofitted in the hallways of an office building but the evaluation contractor only counts 238 fixtures, it is not necessary to adjust the claimed fixture count in the ex post savings calculation (because the error is within +/- 5%). However, if the evaluation contractor verifies only 210 fixtures in the facility hallways, ex post savings values should be calculated based on the evaluator's observations.

[59] http://www.puc.pa.gov/filing_resources/issues_laws_regulations/act_129_information/technical_reference_manual.aspx

[60] This error band is applied solely when verifying the ex ante savings (that is, when calculating the ex post savings and determining the realization rate).

3.3.3.2.4 Site-Specific Measurement and Verification Plan

A Site-Specific Measurement and Verification Plan (SSMVP) is designed to specify the data collection techniques for physical evidence or survey responses from field installations of energy-efficient technologies. SSMVPs for projects within a prescriptive program will be very similar. A common plan is typically updated with the specifics of each project prior to the site visit. For custom measures, SSMVPs are individually created for each project in the evaluation sample. The evaluation contractors must design and document SSMVPs for each measure and define the quantitative data that must be collected from the field or other primary sources. The SSMVP should cover all field activities dedicated to collecting site-specific information necessary to calculate savings according to the engineering equations specified at the project level and to prepare for an evaluation audit of gross savings impacts. This procedure includes specifying data to be gathered and stored for field measurements that document the project processes and rationale. For non-custom measures, general measure-specific data collection workbooks may be used for preparing and completing onsite visits. For custom measures, the SSMVP should include a full narrative describing all of the associated evaluation activities and ensuing calculations. These activities typically include:

- Measure counts
- Observations of field conditions
- Building occupant or operator interviews
- Measurements of parameters
- Metering and monitoring

Appendix B and Appendix C provide detailed guidance on data collection and documentation for common high-impact measures in the TRM that should also be used in the development of SSMVPs. Note that Appendix B and Appendix C prescribe the minimum level of information required for the TRM measure SSMVPs. EDC evaluation contractors are encouraged to supplement the information in Appendix B and Appendix C with project-specific considerations to develop SSMVPs for projects in the evaluation sample.

For custom measures, special considerations should be taken into account for developing SSMVPs. Field measurements are an important component of determining savings for complex projects. The SSMVPs should follow the requirements of the IPMVP. Note that the IPMVP is written to allow for flexibility, but its application requires a thorough knowledge of measure performance characteristics and data acquisition techniques. Energy use varies widely based on the facility type and the electrical and mechanical infrastructure in the facility or system. A measurement strategy that is simple and inexpensive in one building (such as measuring lighting energy at a main panel) may be much more expensive in a similar building that is wired differently. For this reason, evaluation resources, costs, and benefits must be considered and allocated given the type of measure and its impact.

EDC evaluation contractors should assess the expected uncertainty in the end-use energy consumption variables and develop an SSMVP for a sampled custom measure that manages the uncertainty in the most cost-effective manner. The contribution of specific engineering parameters to the overall

uncertainty in the savings calculations should be identified and used to guide the development of the SSMVP.

The SSMVP for sampled measures should include the following sections:

1. Goals and Objectives
2. Building Characteristics and Measure Description
3. EM&V Method
4. Data Analysis Procedures and Algorithms
5. Field Monitoring Data Points
6. Data Product Accuracy
7. Verification and Quality Assurance Procedures
8. Recording and Data Exchange Format

The content of each of these sections is described below.

Goals and Objectives: The SSMVP should state explicit goals and objectives of the EM&V.

Site Characteristics: Site characteristics should be documented in the plan to help future users of the data understand the context of the monitored data. The site parameters to be documented will vary by program and measure. The site characteristics description should include:

- Relevant building configuration and envelope characteristics, such as building floor area, conditioned floor area, number of building floors, opaque wall area and U-value, window area, and solar heat gain coefficient;
- Relevant building occupant information, such as number of occupants, occupancy schedule, and building activities;
- Relevant internal loads, such as lighting power density, appliances, and plug and process loads;
- Type, quantity, and nominal efficiency of relevant heating and cooling systems;
- Relevant HVAC system control set points;
- Relevant changes in building occupancy or operation during the monitoring period that may affect results; and
- Description of the energy conservation measures at the site and their respective projected savings.

The SWE recognizes that not all of these site descriptions are attainable before the site visit occurs and while drafting the SSMVP. However, evaluators should include as many attainable descriptions as feasible in the SSMVP and include any remaining descriptions in the final onsite report.

EM&V Method: The EM&V method chosen for the project should be specified. EM&V methods generally adhere to the applicable IPMVP protocol for the defined level of rigor. The evaluation contractors have considerable latitude regarding the development of an SSMVP, which may be a combination of the IPMVP options.

Data Analysis Procedures and Algorithms: Engineering equations and data points for collection should be identified in advance and referenced within the SSMVP. Engineering calculations should be based on the TRM for partially deemed measures. The equations and documentation supporting baseline assumptions as part of the SSMVP may be presented in the most convenient format (spreadsheet or written report), but should always be clearly stated and explained. This aspect is a key component of an SSMVP, in addition to the application documents. Fully specifying the data analysis procedures will help ensure presentation of an efficient and comprehensive SSMVP.

Field Monitoring Data Points: If any actual field measurements are planned, they should be specified, including the sensor type, location, and engineering units.

Verification and Quality Assurance Procedures: Data analysis procedures to identify invalid data and treatment of missing data and/or outliers must be provided. This should include quality assurance procedures to verify data acquisition system accuracy and sensor placement issues.

Recording and Data Exchange Formats*:* Data formats compliant with the data reporting guidelines described in Section 4.1 of this Evaluation Framework should be specified.

3.3.3.3 Additional Guidance for Ancillary Programs

In addition to the common guidelines provided in Section 3.3.2.3, certain ancillary programs may require special treatment. The Mass Market Protocols (MMPs) for each of these programs is discussed in detail in this section. The full versions of these MMPs are included in Appendix D of this Framework.

3.3.3.3.1 Targeted Low-Income Programs

The Phase II Implementation Order for Act 129 requires each EDC to achieve at least 4.5% of its consumption reduction requirement from the low-income sector. In order to reach this goal, many EDCs have chosen to offer a targeted low-income program to assist low-income residential customers in making their homes more energy-efficient. Savings calculations for these programs may utilize an Option C Billing Analysis if the measures offered include weatherization or other weather-dependent improvements. If TRM fully or partially deemed measures (such as CFLs, smart strips, or faucet aerators) are provided by the ICSP, the EDC may choose to use the deemed savings value for the measure. However, a billing analysis approach captures all improvements to a premise, including the measures with deemed savings. Therefore, the EDC should ensure that the use of the deemed savings values does not result in double-counting (i.e., the savings from deemed measures are also included in the billing analysis) or should use only the billing analysis for the savings calculation to avoid double-counting impacts. In summary, EDCs can utilize the Mass Market Protocols included in Appendix DAppendix D of this Framework to determine kWh savings from Act 129 low-income programs or they may utilize fully or partially deemed measure protocols in the applicable TRM.

During Phase I of Act 129, the EDC low-income programs were very similar in scope to the Low Income Usage Reduction Program (LIURP) efforts the EDCs also administer. In many cases during Phase I, the two programs used the same ICSPs. In Phase II, the FirstEnergy companies, for example, are moving

away from this model. The FirstEnergy Act 129 programs will be different from LIURP programs and will mostly provide TRM partially deemed or fully deemed measures.

The following guidelines are established in an effort to standardize savings calculations for Act 129 low-income energy efficiency programs between EDCs:

1. **Weather Normalization** – Savings calculations based on a comparison of pre- and post-retrofit consumption must be weather-normalized to ensure that savings estimates are not unfairly influenced by differences in the observed weather patterns between the two periods. The relationship between home energy consumption and weather should be determined via a regression analysis that uses cooling and heating degree days (CDD) with a default base 65 degree F temperature).[61] Instead of using the default base temperature values of 65 degrees F, EDCs may also determine the appropriate temperature to use for the base temperature using multiple regression analysis. These relationships should be applied to the normal values for the appropriate region. EDCs may use either a 15- or 20-year period to calculate normal temperatures.
2. **Exclusions** – All homes receiving improvements (non-TRM measures) via an Act 129 low-income program should be included in the impact evaluation sample for the program, provided the following conditions are met:
    - If a comparison group is not available, then 12 months[62] of pre-retrofit billing history and 12 months of post-retrofit billing history need to be available.
    - Occupancy has been continuous at the premise/meter. If multiple account numbers have been associated with the premise/meter during either 12-month period, only billing history data that belong to the original tenant should be included in the analysis.
    - There are no apparent vacancies in the billing history. If the consumption reading for a 1-month interval is less than 100 kWh, the month should be excluded from analysis.
3. **Retrofit Month** – The billing month during which the retrofit is completed should be excluded from the analysis. If retrofits are staged across multiple visits and billing periods, all billing periods from the first visit to the final visit should be excluded from the analysis.

Aggregation of participating homes from multiple years stabilizes the load reduction estimates from billing analysis and increases the likelihood of similar estimates from year to year. Evaluators may aggregate participants from multiple program years into a single billing analysis if there are fewer than

[61] The SWE notes that the base of 65 degrees F for heating and cooling degree days may not always be the appropriate base temperature. EDCs may use a different base temperature as long as they provide an explanation to the PUC of why a different base temperature was selected. The SWE notes that EDC evaluation contractors sometimes use multiple regression analyses to determine the appropriate base temperature for the pre and post conditions. If an energy efficiency program is primarily including envelope and load reduction measures, this can fundamentally change the base temperature pre and post.

[62] Utility billing periods do not always follow calendar months containing 30/31 days. Evaluators should use their discretion with regard to billing intervals when applying this guidance.

500 participating homes that meet the exclusion requirements described above for a given program year. If an EDC is able to achieve the 500-home minimum using participants from a single program year, aggregation should be avoided to ensure that an independent estimate of program impacts is calculated for each program year.[63]

If an EDC changes the measure mix offered, the EDC has two choices: (1) do not include homes prior to the change in the savings analysis for homes evaluated after the change occurred, or (2) conduct separate analyses for homes before and after the program change occurred. Under Option 1, for example, if an EDC provides additional attic insulation to participants in PY5, but discontinues this practice in PY6, participating homes from PY5 could be excluded in the savings analysis for PY6 because the PY6 homes without insulation improvements can be expected to show a lower per-unit kWh/yr savings than the PY5 homes that received additional insulation.

3.3.3.3.2 Behavior and Education Programs

Currently, the TRM does not address programs that focus on occupant behavior and education. However, these programs must be evaluated. The evaluation approach should follow an MMP. In 2013, evaluators and SWE agreed that evaluation would follow SEE Action guidelines.[64] As a guide, the SEE Action protocol recommends using a randomized controlled trial (RCT) for behavior-based efficiency programs, which would result in robust, unbiased program savings impact estimates. The SWE will coordinate with the EDCs to determine whether evaluators should adopt the UMP protocol for Behavior and Education Programs when it becomes available.

There is potential for double-counting of savings if the program causes participants to engage in other programs within an EDC's portfolio. One example is a participant in a behavioral program who also receives a rebate for replacing a heat pump. Savings from the heat pump replacement will be claimed in the program that supplied the rebate, but also will be observed in the billing analysis. The SWE recommends subtracting the savings associated with other rebate offerings from the impact estimate for the behavior program.

Upstream programs (such as discounted CFLs) complicate the situation, because customer participation is not tracked and evaluators will be unable to identify and link upstream participants to the utility's efficiency database. To estimate CFL savings counted in the behavioral program and upstream lighting program, evaluation contractors should assume that the test group's program participation is proportional to the ratio of upstream savings to downstream program savings in the residential portfolio as a whole. Consider an EDC where 60% of the residential energy savings come from upstream CFLs and

[63] The Mass Market Protocol for residential low-income program savings determination will be updated to reflect that calculating impacts for a single program year is preferable when there are 500 or more participants in a given program year. The TUS staff and the SWE Team find that using a threshold of 500 or more participants for a given program year will allow for the development of robust program savings estimates for that year.

[64] State and Local Energy Efficiency Action Network, 2012. *Evaluation, Measurement, and Verification (EM&V) of Residential Behavior-Based Energy Efficiency Programs: Issues and Recommendations*. Prepared by A. Todd, E. Stuart, S. Schiller, and C. Goldman, Lawrence Berkeley National Laboratory.

40% from rebated (tracked) measures.[65] If an analysis of the behavioral program participants determines that 2% of that program's savings actually came from measures rebated (tracked) in other programs, the following calculation will determine the percentage of behavioral program savings that should be attributed to upstream CFLs.

$$CFL\ Share = \frac{60\%}{40\%} * 2\%$$

$$CFL\ Share = 3\%$$

In this example, the evaluation contractor should assume that 3% of the behavioral program's savings were achieved by upstream CFLs and that these savings already were counted in the impacts of the upstream CFL program. As a result, the evaluator should subtract both the 2% savings share attributable to residential rebate programs and the 3% savings share attributable to upstream CFLs from the gross impact estimate for the behavioral program.

3.4 Sampling Statistics and Presentation of Uncertainty

Gross verified energy and demand savings estimates for EE&C programs usually are determined through the observation of key measure parameters among a sample of program participants. A census evaluation would involve surveying, measuring, or otherwise evaluating the entire population of projects within a population. Although a census approach would eliminate the sampling uncertainty for an entire program, the reality is that M&V takes a lot of resources. When a sample of projects is surveyed and analyzed, the sample statistics can be extrapolated to provide a reasonable estimate of the population parameters. Therefore, when used effectively, sampling can improve the overall quality of an evaluation study. By limiting resource-intensive data collection and analysis to a random sample of all projects, more attention can be devoted to each project surveyed.

There is an inherent risk, or uncertainty, that accompanies sampling, because the projects selected in the evaluation sample may not be representative of the program population as a whole with respect to the parameters of interest. As the proportion of projects in the program population that are sampled increases, the amount of sampling uncertainty in the findings decreases. The amount of variability in the sample also affects the amount of uncertainty introduced by sampling. A small sample drawn from a homogeneous population will provide a more reliable estimate of the true population characteristics than a small sample drawn from a heterogeneous population. Variability is expressed using the coefficient of variation (C_v) for programs that use simple random sampling, and an error ratio for programs that use ratio estimation. The C_v of a population is equal to the standard deviation (σ) divided by the mean (μ) as shown in Equation 1.

[65] For this analysis, researchers must remove savings from programs such as residential new construction for which the behavioral program cannot affect participation.

Equation 1: Coefficient of Variation

$$C_v = \frac{\sigma}{\mu}$$

When ratio estimation is utilized, standard deviations will vary for each project in the population. The error ratio is an expression of this variability and is analogous to the C_v for simple random sampling.

Equation 2 provides the formula for estimating error ratio.[66]

Equation 2: Error Ratio

$$Error\ Ratio = \frac{\sum_{i=1}^{N} \sigma_i}{\sum_{i=1}^{N} \mu_i}$$

Equation 3 shows the formula used to calculate the required sample size for an evaluation sample[67], based on the desired level of confidence and precision. Notice that the C_v term is in the numerator, so required sample size will increase as the level of variability increases.

Equation 3: Required Sample Size

$$n_0 = (\frac{z * C_v}{D})^2$$

Where:

n_0 = The required sample size before adjusting for the size of the population
Z = A constant based on the desired level of confidence (equal to 1.645 for 90% confidence two-tailed test)
C_v = Coefficient of variation (standard deviation/mean)
D = Desired relative precision

Unfortunately, the evaluation contractor does not know the C_v and error ratio values until after the verified savings analysis is complete, so they must make assumptions about the level of variability in the savings values, based on previous program years or evaluations of similar programs in other jurisdictions. In the absence of prior information regarding the C_V for the targeted population, EDC evaluators can assume a default C_V equal to 0.5 for each sample population to determine target sample sizes. Once the C_V has been demonstrated, evaluators may use that C_V in developing their sampling

[66] **Error! Reference source not found.** is based on the methodology set forth in the California Evaluation ramework. The National Renewable Energy Laboratory's Uniform Methods Project (NREL UMP) provides a slightly different formula for the calculation of error ratio that is an acceptable alternative if evaluation contractors wish to use it.
[67] If ratio estimation is used, evaluators may replace C_v with error ratio in Equation 3.

plans. Evaluators should trend the actual C_V values for each sample pool and report the values in their annual reports and subsequent evaluation plans.

The sample size formula shown in Equation 3 assumes that the population of the program is infinite and that the sample being drawn is reasonably large. In practice, this assumption is not always met.

For sampling purposes, any population greater than approximately 7,000 may be considered infinite for the purposes of sampling. For smaller, or finite, populations, the use of a finite population correction factor (FPC) is warranted. This adjustment accounts for the extra precision that is gained when the sampled projects make up more than about 5% of the program savings. Multiplying the results of Equation 3 by the FPC formula shown in Equation 4 will produce the required sample size for a finite population.

Equation 4: Finite Population Correction Factor

$$\boldsymbol{fpc} = \sqrt{\frac{\boldsymbol{N} - \boldsymbol{n_0}}{\boldsymbol{N} - \boldsymbol{1}}}$$

Where:

N = Size of the population

n_0 = The required sample size before adjusting for the size of the population

The required sample size (*n*) after adjusting for the size of the population is given by Equation 5.

Equation 5: Application of the Finite Population Correction Factor

$$\boldsymbol{n} = \boldsymbol{n_0} * \boldsymbol{fpc}$$

3.4.1 Annual Evaluation Precision Requirements for Verified Savings Estimates

Table 3-5 provides levels of sampling uncertainty prescribed for the Act 129 gross impact evaluations in order to balance the need for accurate savings estimates while limiting the costs of evaluation. The values in Table 3-5 assume a two-tailed design must be met or exceeded annually. A gross verified energy savings estimate with 10% relative precision at the 90% confidence indicates that if evaluators resampled the same population repeatedly, 90% of the time the resulting intervals would include the true value of the measured parameter,[68] assuming an unbiased sample. In reality, there are a number of other sources of uncertainty that are less straightforward to quantify and reduce the precision of savings estimates. These factors are discussed in Section 3.5, but should not be addressed by evaluators when calculating the achieved precision of a verified savings estimate.

[68] (Lohr, 2010)

Table 3-5: Minimum Annual Confidence and Precision Levels

Portfolio Segment	Confidence and Precision Level
Residential Portfolio	90/10
Nonresidential Portfolio	90/10
Individual Programs Within Each Portfolio	85/15

Special consideration should be given to the following situations:

1. Cross-cutting programs that span both the residential and nonresidential sectors must[69] be evaluated as independent programs, one for the residential sector and one for the nonresidential sector.
2. The government, non-profit and institutional populations, and the low-income population should be evaluated as independent programs if their contribution to their respective sectors [the residential sector for the low-income population, and nonresidential sector for the government, non-profit, and institutional (GNI) population] is greater than 20%. The annual contribution should be assessed at the end of Q4. This will provide evaluation contractors sufficient time to increase verification rates prior to the final annual report if the low-income or GNI populations contribute a greater share of savings than was anticipated.
3. An EDC evaluator should not choose to aggregate all non-GNI, commercial, and industrial projects into a single "umbrella" program for evaluation. An "umbrella" program is defined as a program that contributes more than 75% of the nonresidential-sector portfolio non-GNI annual energy savings, and more than 60% of the overall nonresidential-sector portfolio annual energy savings. The evaluation of the nonresidential-sector portfolio should have at least two distinct, non-umbrella, non-GNI programs, and each program evaluation should meet the requirements established in Table 3-5.
4. The list below provides suggestions for possible program boundaries within the C&I customer segment.
    - Small C&I and Large C&I
    - Commercial and Industrial
    - C&I Lighting and C&I Non-Lighting
    - Custom and Prescriptive

    If an evaluation contractor chooses to employ an umbrella program, a C_v of 1.0 should be assumed when determining the required sample size for the impact evaluation.

[69] The SWE may approve exceptions during the review of EDC EM&V plans. For example, small businesses may be eligible to participate in an appliance recycling program, but 99% of the program savings will come from the residential sector. The 1% of program savings from the nonresidential sector does not need to be evaluated as a standalone program.

5. It often is more challenging to obtain accurate peak demand savings estimates than annual energy savings estimates, and peak demand savings estimates will exhibit a greater degree of variability between ex ante and ex post. The levels of precision established in Table 3-5 are required for energy savings estimates. If achieved precision values for peak demand impacts are significantly greater than the precision of energy estimates, evaluators should examine the source of the variation to determine if revisions to ex ante demand savings assumptions or ex post analysis techniques are warranted.

Evaluation contractors may use their professional judgment in the design of the sample as long as they meet the minimum precision requirements. Evaluation contractors should design evaluation samples to exceed the minimum requirements so they will not miss the precision requirements established in this Evaluation Framework if program characteristics (population size, variability) are slightly greater than anticipated. If the annual confidence and precision targets are not met, corrective actions will be required in the current or subsequent evaluation year within the compliance period.

Evaluators may propose alternative minimum confidence and precision requirements for programs with special circumstances on an individual basis to the SWE for review and approval. Programs with smaller savings contributions may be sampled at lower precision levels, and evaluators should propose the approach with adequate justification in the evaluation plan.

Programs should use stratification to ensure that the sample is efficiently designed and representative of the population of measures within the program by creating homogeneous population groups to the greatest extent possible where multiple measures are rebated within one program, and there is a variance in the savings across program measures. Where programs provide incentives for projects (that can include multiple measures) the sample should represent the population, and stratification may be used, if needed. Evaluators should use their professional judgment to develop size thresholds and definitions for the project strata, subject to review and approval by the SWE. The SWE audit of evaluator sample designs is discussed in more detail in Section 4.3.4.1. For high-impact/high-uncertainty project strata, evaluators should ensure that they evaluate savings using an enhanced level of rigor. Section 3.3.2.3 of this Framework discussed the expected level of rigor by stratum in more detail. Strata boundaries are specific to the population studied and can be changed for the sampling batches. Section 0 provides references to documents that contain additional guidance on the efficient determination of quantitative (size) stratum boundaries.

Programs such as low-income weatherization, behavior modification, or customer education often rely on a billing regression analysis on a census or near census of program participants to determine verified savings. These programs require special consideration because a census, rather than a sample, of program participants is evaluated, so theoretically there is no sampling uncertainty. Instead, the precision of savings estimates is determined using the standard error of the regression coefficient(s) that determine savings. Evaluators should attempt to meet the requirements established in Table 3-5 for programs that use a census regression approach. The resulting precision of a regression model is difficult to predict or control. Additional sample points cannot be added, and the evaluator has little control over

the variability of the results. If evaluators do not meet the 15% relative precision target, they should reconsider the model specification to determine if the inclusion of additional independent variables or transformations of existing variables in the model can help explain the behavior of the dependent variable(s) and increase the precision of savings estimates.

For a program that uses a billing regression analysis, if the 85% confidence interval around the savings estimates includes 0 kWh, an EDC should not claim savings for the program until a more precise estimate of program impacts is developed. For example, if the per-home savings estimate for a program is equal to 200 kWh/yr ± 400 kWh/yr, no verified savings for the program should be claimed because the evaluator cannot ensure that the program impact is not equal to zero at the 85% confidence level.

3.4.2 Overview of Estimation Techniques

Evaluators may choose to employ two broad classes of probability estimation techniques in the impact evaluation of EE&C programs.

1. **Estimation in the absence of auxiliary information** (also referred to as *mean-per-unit estimation*): This technique is useful if the projects within a population are similar in size and scope. Simple random sampling is recommended for residential programs that include a large number of rebates for similar equipment types.

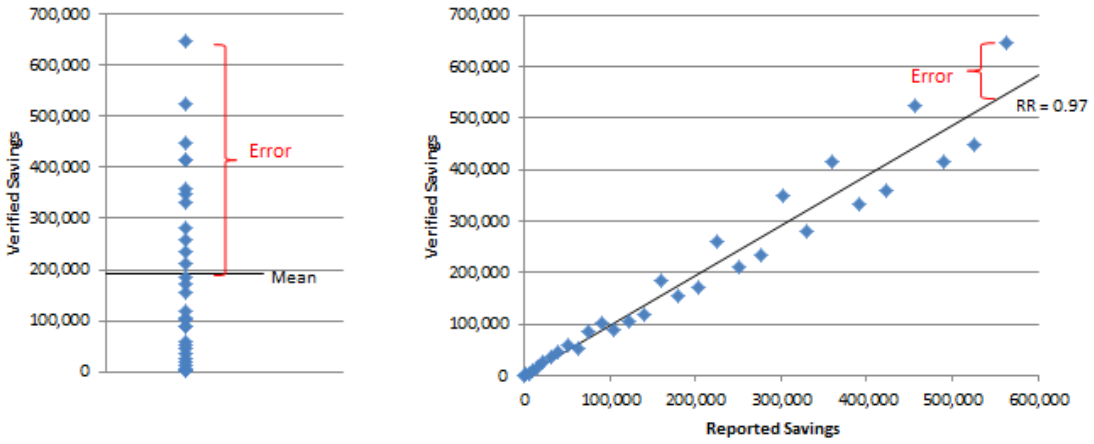
2. **Estimation using auxiliary information** (also referred to as *ratio estimation*): This is recommended for nonresidential programs, or residential programs offering a variety of measures with varying savings, because the sizes of the savings estimates of the projects within a program vary considerably within the program population. Ratio estimation can be used with or without stratification. This technique relies on auxiliary information reported in the program tracking system – usually the ex ante kWh/yr savings of the projects. This technique assumes that the ratio of the sum of the verified savings estimates to the sum of the reported savings estimates within the sample is representative of the program as a whole. This ratio is referred to as the *realization rate*, or *ratio estimator*, and is calculated as follows:

$$Realization\ Rate = \frac{\sum_i^n Verified\ Savings}{\sum_i^n Reported\ Savings}$$

Where *n* is the number of projects in the evaluation sample.

Figure 3-2 shows the reduction in error that can be achieved through ratio estimation when the sizes of projects within a program population vary considerably. The ratio estimator provides a better estimate of individual project savings than a mean savings value by leveraging the reported savings estimate.

Figure 3-2: Comparison of Mean-Per-Unit and Ratio Estimation



Sample stratification can be used with either of the two classes of estimation techniques presented previously. *Stratified random sampling* refers to the designation of two or more sub-groups (strata) from within the program population prior to the selection process. It's imperative that each sampling unit (customer/project/measure) within the population belongs to one (and only one) stratum. Typically, the probability of selection is different between strata; this is a fundamental difference from *simple random sampling*, where each sampling unit has an identical likelihood of being selected in the sample. The inverse of the selection probability is referred to as the *case weight* and is used in estimation of impacts when stratified random samples are utilized. Stratification is advantageous for the following reasons:

- Increased precision if the within-stratum variability is small compared to the variability of the population as a whole. Stratification potentially allows for smaller total sample sizes, which can lower evaluation costs.
- A stratified sample design allows evaluation contractors to ensure that a minimum number of units within a particular stratum will be verified. For example, a C&I program with 1,000 projects in the population, may have only 10 that are CHP projects. If the sample size is 40 and simple random sampling is used, each project has a 4% chance of being included in the sample, and the probability that the resulting sample contains one or more CHP projects is only 33.6%. On the other hand, if stratified random sampling is used and one stratum is defined as including only CHP projects, then as long as the sample size within each stratum is one or more projects, the sample will include a CHP project with certainty and each CHP project will have a 10% probability of being selected.
- Additional sample designs can be considered within each stratum. It is easy to implement a value-of-information approach through which the largest projects are sampled at a much higher rate than smaller projects.
- Sampling independently within each stratum allows for comparisons among groups. Although this Framework only requires that a single relative precision be met at the program level annually, EDCs and their evaluation contractors may find value in comparing results between strata; e.g., comparing the verification rates between measures within a program.

Evaluation contractors are encouraged to limit the use of simple random sampling to programs with homogenous measure populations, such as Appliance Recycling, and to employ stratification for programs which offer a diverse mix of measures. However, the choice of using stratified random sampling or simple random sampling is ultimately left up to the discretion of the EDC evaluation contractor.

3.4.3 Additional Resources

The 2009 and 2011 versions of the *Audit Plan and Evaluation Framework for Pennsylvania Energy Efficiency and Conservation Programs* include detailed information regarding sample design, sample size calculations, definitions and formulas for error ratio, coefficient of variation, and relative precision. This information has been excluded from the 2013 Evaluation Framework. If EDCs, their evaluation contractors, or stakeholders require additional information regarding sampling, the following resources will be helpful:

- *The Uniform Methods Project: Methods for Determining Energy Efficiency Savings for Specific Measures*. Prepared for the National Renewable Energy Laboratory by The Cadmus Group, January 2013.
- *The California Evaluation Framework*. Prepared for the California Public Utilities Commission and Project Advisory Group by TecMarket Works, June 2004.
- *Audit Plan and Evaluation Framework for Pennsylvania Act 129 Energy Efficiency and Conservation Programs*. Prepared for the PUC by GDS Associates, November 2011.

3.4.4 Presentation of Uncertainty

There are no minimum precision requirements for EDC evaluations of Phase II savings as a whole. However, if the annual minimums established in Table 3-5 are met, the relative precision values of the total Phase II savings will meet or exceed the annual requirements at the same levels of confidence. In the annual report for each program year, each EDC should report the verified energy and demand savings achieved by each program in its portfolio and estimates for the entire portfolio. Verified savings estimates should always represent the point estimate of total savings, or the midpoint of the confidence interval around the verified savings estimate for the program. In addition to the verified savings estimates for energy and demand, EDCs should report the error bound, or margin of error, and the relative precision of the savings estimate such that:

Equation 6: Error Bound of the Parameter Estimate

$$Error\ Bound = se * (z - statistic)$$

Where:

se = The standard error of the population parameter of interest (proportion of customers installing a measure, realization rate, total energy savings, etc.) This formula will differ according to the sampling technique utilized.

$z - statistic$ = Calculated based on the desired confidence level and the standard normal distribution.

Table 3-6 provides the appropriate z-statistic to use for several commonly used confidence levels. Each value assumes a two-tailed design.

Table 3-6: Z-statistics Associated with Common Confidence Levels

Confidence Level	Z-statistic
80%	1.282
85%	1.440
90%	1.645
95%	1.960

Use of a z-statistic implies normality. The Central Limit Theorem shows that the means of sufficiently large random samples drawn from a population will follow a normal distribution, even if the population that is the source of the sample is not normally distributed. However, for sample sizes smaller than 30, the Central Limit Theorem begins to break down and the normality assumption no longer is valid. A t-distribution is the appropriate distribution for evaluators to consider when drawing samples of fewer than 30 projects/measures. In this case, a t-statistic will be used in estimation once the sample has been collected. The t-statistic replaces the z-statistic in Equation 6 and is calculated using the *degrees of freedom* (sample size minus the number of estimates). As the sample size becomes larger, the t-statistic gets closer to the z-statistic.

In cases where the parameter of interest is a proportion or realization rate, the estimate is applied to the reported savings values in order to calculate the gross verified savings for the program. The error bound of the *verified savings estimate* (in kWh/yr or kW) should be reported for each program and is calculated as follows:

Equation 7: Error Bound of the Savings Estimate

$$Error\ Bound_{(kWh\ or\ kW)} = Error\ Bound_{Parameter} * Gross\ Reported_{(kWh\ or\ kW)}$$

The *relative precision value* of the verified savings estimate[70] for each program should be reported, as well as the confidence level at which it was calculated. This formula is shown in Equation 8:

Equation 8: Relative Precision of the Savings Estimate

$$Relative\ Precision_{Verified\ Savings} = \frac{Error\ Bound_{(kWh\ or\ kW)}}{Gross\ Verified_{(kWh\ or\ kW)}}$$

Evaluations of programs that use stratified ratio estimation require an additional step because each stratum will have its own realization rate and error bound that should be reported.

[70] The relative precision of the verified savings estimate should equal the relative precision of the population parameter; it can be determined prior to calculating the error bound of the energy or demand savings estimate.

At the conclusion of Phase II of Act 129, each EDC will have three annual verified savings estimates for energy and three annual verified savings estimates for demand across each program in its portfolio, (one for each program year). The Phase II verified savings estimate is the sum of these values, minus any savings that expire prior to May 31, 2016.[71] Although the annual error bound estimates are expressed in the unit of interest (kWh/yr or kW), they cannot be summed to produce the error bound for Phase II impacts. Equation 9 shows the formula for calculating the error bound of the Phase II impacts. The same methodology should be used to calculate the error bound and relative precision of the annual sector- and portfolio-level verified savings estimates. Phase II error bounds and relative precisions should be calculated and reported at the 90% confidence level. This will require a recalculation of the annual error bounds if the 85% confidence level were used for a program. To convert the annual error bound to the 90% confidence interval, evaluators should perform the calculations shown in Equation 6 and Equation 7 using the standard error of the parameter estimate and the z-statistic associated with the 90% confidence interval (1.645).

Equation 9: Phase II Error Bound

$$Error\ Bound_{Phase\ II} = \sqrt{Error\ Bound_{PY5}^2 + Error\ Bound_{PY6}^2 + Error\ Bound_{PY7}^2}$$

Using this methodology, evaluators will have a verified savings estimate for the program and an error bound for that estimate. The relative precision of the Phase II verified savings for the program is then calculated using these two values.

Equation 10: Relative Precision of Phase II Savings Estimate

$$Relative\ Precision_{Phase\ II} = \frac{Error\ Bound_{Phase\ II}}{Gross\ Verified\ Savings\ Estimate_{Phase\ II}}$$

Equation 9 also should be used to combine the Phase II error bounds from programs to the sector level and from the sector level to the portfolio level. Note that Equation 9 assumes that estimated savings in PY5, PY6, and PY7 are independent. The independence assumption must hold for this formula to be applied to the combination of program-level savings to the sector level within a portfolio and/or program year.

3.5 Systematic Uncertainty

Section 3.4 of the Evaluation Framework discussed the uncertainty that is introduced into evaluation findings when a sample, rather than a census, of projects is used to determine program impacts. *Sampling uncertainty*, or error, largely is random and can be estimated using established statistical procedures. On the other hand, *systematic uncertainty* represents the amount of error that is introduced into evaluation results consistently (not randomly) through the manner in which parameters are measured, collected, or described. Systematic uncertainty is more challenging to quantify and

[71] The Phase II Implementation Order requires that measures with an effective useful life shorter than the length of a program phase be replaced in order to count towards compliance with savings targets.

mitigate than sampling uncertainty because sources of systematic uncertainty often are specific to the program, measure, or site being evaluated. However, to present evaluation results as though sampling error were the only source of uncertainty in an evaluation misrepresents the accuracy with which an EDC can estimate the impacts achieved by its EE&C Plan. EDC annual reports should discuss major sources of systematic uncertainty and the efforts the evaluation contractor made to mitigate them.

Common sources of systematic uncertainty, which should be considered in an EDC's evaluation plan include:

1. **Deemed or Stipulated Values** – TRM values are based on vetted engineering principles and provide reasonable estimates of measure energy and demand impacts while expending relatively few evaluation resources. Using these values in evaluation results can introduce considerable bias if the values are not adequately prescribed or do not fully capture the complexity of a measure. Dated values or adjusted values from secondary research are likely to introduce systematic error in the evaluation findings.

2. **Data Collection and Measurement** – According to sampling theory, when a project is selected in the impact evaluation sample and energy and demand savings values are calculated, those savings values are discrete. In reality, the reliability of these estimates is subject to a host of uncertainties that must be considered. Survey design can introduce a variety of biases into evaluation findings. Consider a lighting survey that includes questions to a facility contact about the typical hours of operation in their building. If the survey does not include questions about business closings for holidays, the survey responses will systematically overestimate the *equivalent full load hours* (EFLH) of fixtures in the facility. Evaluators also must consider another source of systematic uncertainty: human error. If the engineer visiting a site in the evaluation sample forgets to complete a key field on the data collection instrument, an assumption must be made by the analyst calculating savings for the project regarding the parameter in question. Onsite metering is considered a high-rigor evaluation approach and is reserved for high-impact/high-uncertainty projects, but these results can be biased by equipment placement, poor calibration, or differences in the pre/post metering period not addressed in the analysis.

3. **Sample Design** – Evaluation samples are constrained by evaluation budgets and the practicality of collecting information. Non-coverage errors can arise if the sampled population does not accurately represent the population of interest. For instance, an evaluation survey that is conducted via email with a random sample of EDC customers necessarily excludes all customers who do not have an email address, or have chosen not to provide their EDC with this information. If this population of customers somehow differs from the population of customers with known email addresses (the sample pool) with respect to the parameter in question, the value calculated from the sample will not accurately reflect the population of interest as a whole.

   Non-response and self-selection errors occur when some portion of the population is less likely (non-response) or more likely (self-selection) to participate in the evaluation than other

portions. Retired customers frequently are over-represented in residential evaluation findings because daytime recruiting calls to a home phone number are far more likely to reach retired program participants. Values calculated from samples that over-represent certain segments and under-represent others are subject to systematic uncertainty if the customer segments differ with respect to the parameter of interest.

3.5.1 Calculating and Reporting Systematic Uncertainty

If the standard error of a parameter estimate is available in the TRM,[72] EDC evaluation contractors should attempt to quantify the systematic uncertainty associated with deemed or stipulated values. This calculation follows statistical procedures similar to those for estimating sampling uncertainty. One example is the CFL_{hours} term in the ENERGY STAR® lighting section of the Pennsylvania TRM, which deems the average daily hours-of-use (HOU) estimate for a residential CFL. In the 2014 Pennsylvania TRM, the value for CFL HOU is 2.8 hours per day. This value was taken from a study conducted in another jurisdiction, and was based on a sample of homes and fixtures. Because not all fixtures metered in the study were turned on precisely 2.8 hours per day, there is a non-zero standard error term associated with the mean value. In an upstream lighting program, the evaluation method is usually a census review of the tracking data. However, to just report that there is no sampling error in the evaluation implies a false precision that can be misleading to stakeholders.

If the standard error of this stipulated value is equal to 0.2 hours per day and an EDC's upstream lighting program rebates 400,000 13-watt CFL bulbs during a program year, the following calculations will estimate the amount of systematic uncertainty associated with the use of the deemed CFL_{hours} value.

The per-bulb savings estimate for the program is equal to:

$$\Delta\text{kWh/yr} = (\text{Watts}_{base} - \text{Watts}_{CFL}) * \text{CFL}_{hours} * 365 / 1000 * \text{ISR}^{73}{}_{CFL}$$

$$\Delta\text{kWh/yr} = (60 - 13) * 2.8 * 365 / 1000 * 0.84$$

$$\Delta\text{kWh/yr} = 40.349$$

The program savings estimate is equal to:

$$Total\ Savings = Number\ of\ Units * Savings\ Per\ Unit$$

$$Total\ Savings = 400{,}000 * 40.349$$

$$Total\ Savings = 16{,}139\ MWh/yr$$

[72] Once the SWE lighting metering studies are complete, standard error values will be associated with lighting parameters in the TRM protocols.

[73] ISR is the *in-service rate* of the measure, or the proportion of bulbs estimated to be installed. This value is equal to 84% in the 2013 TRM.

These upper and lower bounds of the deemed CFL_{hours} term are calculated as follows:

$$2.8 \pm (standard\ error) * (z - statistic)$$

Where the standard error of the mean is equal to 0.2 hours per day, and the z-statistic at the 90% confidence interval is equal to 1.645.

$$2.8 \pm (0.2) * (1.645) = (2.471, 3.129)$$

Applying these upper and lower bounds to the per-bulb and program savings estimates formulas above return program savings that could be as low as 14,243 MWh/yr or as high as 18,036 MWh/yr. When possible, EDC evaluation contractors should attempt to estimate systematic uncertainty in this manner, and present the margin of error at the 90% confidence level and separate systematic uncertainty from sampling uncertainty to avoid increasing the level of rigor required to meet the precision requirements established in Section 3.4.

Standard error values historically have not been included in the TRM for deemed values, but will be a point of emphasis for high-impact measures in future TRM updates. Once complete, the results of the SWE's residential and nonresidential lighting metering studies will significantly improve the ability of EDC evaluation contractors to quantify the systematic uncertainty introduced through the use of TRM-specified *equivalent full load hours* (EFLH) and *coincidence factors* (CF) values. For measures where a SWE metering study has not been completed and standard error values are not given in the TRM, it is unlikely that evaluation contractors will have sufficient data to produce meaningful estimates of standard error and systematic uncertainty.

Given the number of deemed or stipulated values in the TRM, including the systematic uncertainty in the reporting of impacts could be cumbersome to EDC evaluation contractors if it is not limited to a small number of high-impact measures with specified standard error values. For other measures, EDC evaluators should use their professional judgment to determine other sources of systematic uncertainty from stipulated values that merit discussion in evaluation reports. If an EDC evaluator believes that use of a stipulated value in the TRM may systematically cause over- or under-estimating program impacts, the analysis supporting this finding and an estimate of the relative impact should be included in an appendix to the annual report. These observations will provide critical feedback to the SWE and help prioritize measure reviews during the annual TRM update.

Unlike the systematic uncertainty introduced by stipulated values, the systematic uncertainty resulting from data collection and measurement or sample design cannot be easily quantified with a formula. EDC evaluators should discuss the steps taken to mitigate systematic error from these sources and any analysis undertaken to understand where significant sources may exist. The Uniform Methods Project Sampling Protocols[74] (UMPSP) identifies six areas, which may be examined to determine how rigorously

[74] The protocols can be found at http://energy.gov/eere/downloads/uniform-methods-project-methods-determining-energy-efficiency-savings-specific .

and effectively an evaluator has attempted to mitigate sources of systematic error. A summary of the six areas is as follows:

1) Were measurement procedures (such as the use of observational forms or surveys) pretested to determine if sources of measurement error could be corrected before the full-scale fielding?
2) Were validation measures (such as repeated measurements, inter-rater reliability, or additional subsample metering) used to validate measurements?
3) Was the sample frame carefully evaluated to determine which portions of the population, if any, were excluded in the sample? If so, what steps were taken to estimate the impact of excluding this portion of the population from the final results?
4) Were steps taken to minimize the effect of non-response or self-selection in surveys or other data collection efforts? If non-response appears to be an issue, what steps were taken to evaluate the magnitude and direction of potential non-response bias? Were study results adjusted to account for non-response bias via weighting or other techniques?[75]
5) Has the selection of formulas, models, and adjustments been conceptually justified? Has the evaluator tested the sensitivity of estimates to key assumptions required by the models?
6) Did trained, experienced professionals conduct the work? Was the work checked and verified by a professional other than the one conducting the initial work?

EDC evaluation plans and annual reports should discuss the steps evaluation contractors took to answer as many of the questions above as possible in the affirmative. SWE audit activities will consider the appropriateness of evaluators' techniques to mitigate systematic uncertainty and identify areas where changes or additional research are warranted.

3.6 Net Impact Evaluation

The PUC stated in the Phase II Implementation Order that NTG adjustments be treated the same way for Phase II as they were during Phase I. "*Specifically, the Commission [PUC] proposed that NTG research be used to direct Act 129 program design and implementation, but not for compliance purposes*."[76] The reasons for not using NTG for compliance are:

> *"One, if a NTG ratio of less than 1.0 is used, this will raise the acquisition cost per annual kilowatt-hour (KWh) saved to the EDC, which will result in a lower target, due to the 2% budget cap. The current targets include an assumed NTG ratio of 1.0. Two, the Commission recognizes that the calculation of NTG ratios is inexact at best. 'Free riders' are difficult and expensive to calculate, but even more difficult and costly to calculate is 'spillover'."*[77]

[75] Some common methods to deal with non-response by incorporating response rates into the sampling weights are presented in *Applied Survey Data Analysis* by Heeringa, West, and Berglund (2010).
[76] Pennsylvania Public Utility Commission, Energy Efficiency and Conservation Program Implementation Order, at page 82, at Docket No M-2012-2289411, (Phase II Implementation Order), entered August 3, 2012.
[77] Ibid., page 83.

Further, some published studies conclude these two effects, if properly measured, will likely come close to offsetting each other and result in a NTG ratio close to 1.0.[78] Due to the substantial additional costs to calculate "free-riders" and "spillover," the PUC questions whether it is cost-effective to use ratepayer funds for these analyses, only to find that the NTG ratio is close to 1.0. No stakeholders have provided evidence to the contrary, so the PUC will continue to mandate that the EDCs continue to use net verified savings in their TRC for program planning purposes, and compliance in Phase II will be determined using gross verified savings.[79]

However, the PUC does recognize that net savings are valuable to inform program design and program planning, as well as future phases of Act 129. Therefore, EDCs' evaluation contractors should conduct NTG research and consider conducting additional research to assess market conditions and market effects to determine net savings. Market effects research is discussed in Section 3.6.1.3.

3.6.1 Acceptable Approaches to Conducting NTG Research

NTG research traditionally has two primary purposes: 1) attribution: adjusting gross savings to reflect actual program influence on savings, and 2) explicating customer decision-making and the contribution the program made to the customer's decision to install an energy-efficient solution. This research helps to determine whether a program should be modified, expanded, or eliminated based on its NTGR.

Program evaluators traditionally use one of several methods to assess a program's net savings, including self-report surveys, econometric methods, and market effect approaches. Much has been written about the various methods and their relative strengths and weaknesses.[80] In light of increasing program activity, as well as activity external to the program that contributes to customers' engagement with energy efficiency, net savings estimation is increasingly difficult to compute. The most cost-effective measurement technique for net savings is self-report surveys, however, social science research shows that measurement of the counterfactual using self-reports is problematic. In addition, while increased participant and nonparticipant spillover installations may be making a greater contribution to savings than the amount that free-ridership detracts from savings, measuring spillover using self-reports suffers from similar problems to free-ridership, and when onsite confirmation is included, it becomes very costly. [81]

On the other hand, econometric and market effect approaches may be even more costly. Moreover, with these approaches, it is not possible to disaggregate the effects of free-ridership and spillover, and

[78] Haeri, H. and M. Sami Khawaja "The Trouble with Freeriders." *Public Utilities Fortnightly*. March 2012 (http://www.fortnightly.com/fortnightly/2012/03/trouble-freeriders).
[79] Pennsylvania Public Utility Commission, Energy Efficiency and Conservation Program Implementation Order, at page 83, at Docket No M-2012-2289411, (Phase II Implementation Order), entered August 3, 2012.
[80] A general review of issues and recent bibliography is provided in Haeri, H. and M. Sami Khawaja, "The Trouble with Freeriders," *op cit*.
[81] Peters, J. S. and M. McRae. "Free-Ridership Measurement is Out of Sync with Program Logic…or, We've Got the Structure Built, but What's Its Foundation?" In *Proceedings of the 2008 ACEEE Summer Study on Energy Efficiency in Buildings*. American Council for an Energy-Efficient Economy.

they do not directly address customer decision-making or the program's influences on decision-making. For this reason, the SWE has determined that EDCs should use survey methods for assessing free-ridership and spillover for downstream programs and has provided descriptions of common methods for doing those assessments; these approaches must be used for the specific programs they apply to, though they may be used in combination with other methods. The SWE has established a procedure whereby EDCs may identify downstream programs for which the common methods are not suitable; in such cases, EDCs may propose a method, which the SWE will review. The EDCs may use other methods for upstream programs or to specifically provide information on market effects.

The primary concern of the SWE is whether the EDCs' NTG evaluations are helping the EDCs fully understand the effects/attribution of their programs on the markets in their service territory. Further, the SWE must ensure that NTGRs are reasonable and ratepayer funds appropriately support customers who need that support in order to invest in energy-efficient solutions.

3.6.1.1 Using Self-Reports for Estimating Free-Ridership and Spillover

Using self-reports to measure free-riders and spillover is subject to bias and therefore may not yield an accurate estimate of free-ridership or spillover; this concern supports the PUC's decision that self-report-based NTG should not be used to calculate net savings estimates for compliance purposes.[82] However, careful application of social science methods may help mitigate biases.[83] Years of research have shown that various NTG self-report assessments tend to produce consistent results. Thus, even if they do not necessarily produce accurate estimates of net savings at any given time, they may be useful in assessing trends over time. Thus, the SWE believes that self-report assessments of free-ridership and spillover may be useful in assessing changes over time or differences across programs.

- **Free-ridership** – The purpose of measuring free-riders is to ensure that the program is primarily serving those who need the program in order to invest in energy efficiency. Thus, over the course of many years of DSM program evaluation, evaluators have developed methods to estimate the number of free-riders and then to estimate the net savings resulting only from those who required the program services in order to install the energy-efficient solutions.
- **Spillover** – The purpose of measuring spillover is to ensure that the program is credited with energy savings that come from participants and nonparticipants who install energy-efficient solutions without using program resources, and do so because of the program, either as participants who take additional efficient actions (inside or participant spillover) or as nonparticipants who take actions the program recommends but do not request program services (outside or nonparticipant spillover).

The NTGR removes free-riders from the savings calculation and adds program spillover. The NTG formula is defined in Equation 11:

[82] Ibid.
[83] Haeri, H. and M. Sami Khawaja "The Trouble with Freeriders." *Public Utilities Fortnightly*. March 2012 (http://www.fortnightly.com/fortnightly/2012/03/trouble-freeriders).

Equation 11: NTG Formula

$$NTG = 1 - FR + SO$$

Where:

FR = *Free-ridership* quantifies the percentage of savings (reduction in energy consumption or demand) from participants who would have implemented the measure in the absence of the EDC program.

SO = *Spillover* quantifies the percentage reduction in energy consumption or demand (that is, additional savings) caused by the presence of the EDC program. Spillover savings happen when customers invest in additional energy-efficient measures or activities without receiving a financial incentive from the program.

3.6.1.1.1 Free-Rider Measurement

The SWE has determined that, where possible, EDCs should use standard sampling techniques, data collection approaches, survey questions, survey instruments, and analysis methodology for free-ridership assessment. Standardization can provide consistency in explications of the programs' effects. EDCs may implement other methods concurrently.

Care must be taken when developing the questions used to measure free-ridership. The SWE considers the research approaches used in Massachusetts and Wisconsin (and by Energy Trust of Oregon [Energy Trust]) to describe some of the best practices for free-ridership and spillover calculation. In early Phase II, the SWE developed common methodologies for estimating free-ridership in downstream programs that EDCs should use or adapt to their purposes. One common approach applies to a broad range of incentive-based programs; the other is specific to appliance recycling programs. The SWE common approach is similar to that chosen by Energy Trust, which uses a short battery of questions but has been found to produce results that are comparable to those produced by much longer batteries.[84] The approach for appliance recycling programs is based on the approach described by the U.S. Department of Energy's Uniform Methods Project. Both approaches have undergone detailed review by the PEG.

The common method uses responses to a sequence of free-ridership questions to compute an overall free-ridership score for each measure or program. It is very important that more than one question be used to determine the level of free-ridership. Free-ridership questions in the common method include two additive and equally weighted components:

- Participant intention
- Program influence

[84] http://energytrust.org/library/reports/101231_Fast_Feedback_Rollout.pdf

Each component provides a possible score of 0 to 50. When added, the resulting score, which has a range of possible values of 0 to 100, is interpreted as a *free-ridership percentage*; this is also how *partial free-riders* emerge. A score of more than 0% and less than 100% indicates a partial free-rider.

Net savings for the appliance retirement program is based on the participants' self-report of what they would have done absent the program. Savings are attributed based on four scenarios: 1) they would have kept the unit but instead replaced it with a more efficient one (savings equals delta energy usage from old to new unit); 2) they would have kept the unit but instead recycled it and did not replace it (savings equals energy usage of old unit); 3) they would have put the unit back into usage elsewhere, sold or given the unit away to another user, or sold or given away a unit was less than 10 years old to a retailer (savings equals a mix of full savings, delta old to new, and no savings); or 4) they would have taken the unit out of usage, sold or given a unit at least 10 years old to a retailer, hauled it to the dump, or hired someone to discard it (free-rider – no savings).

The SWE produced memos describing the common approaches, which are included as Appendix E and Appendix F of this Framework. The memos describe both the general form of questions to use and rules for calculating free-ridership scores from responses to questions. As described in the memos, EDCs may adapt the questions to fit each program, pending SWE review. EDCs may also add questions and/or use alternative formulas for calculating free-ridership scores *in parallel with* the calculations resulting from the methods described in the memos.

The confidence and precision for free-ridership estimates should be consistent with those for gross savings estimate requirements – that is, 85% confidence with ±15% in precision at the program level, and 90% confidence with $\pm$10% precision at the sector level. Note that this does not mean that the estimated net savings (obtained by applying the NTGR, developed from both free-ridership and spillover estimates, to gross savings) must be at the 85/15 or 90/10 level of confidence/precision. Since net savings are not relevant to compliance, there is no specific precision requirement for net savings. The purpose in specifying confidence and precision levels for free-ridership estimates is to ensure results that will be valuable for program planning purposes.

3.6.1.1.2 Spillover Measurement

Net savings claims that include spillover studies are more robust than those that just include free-ridership estimates. The SWE also has determined that, where possible, EDCs should use standard techniques, instruments, and methods for spillover assessment. However, the SWE has determined that, while estimation of nonparticipant spillover is desirable, it is not required. If assessed, nonparticipant spillover may be assessed through either a general population (nonparticipant) survey or a survey of trade allies.

In early Phase II, the SWE developed a common methodology for estimating participant and (if EDCs choose to assess it) nonparticipant spillover in downstream programs. The SWE produced a memo describing the common approaches, which is included as Appendix G. The memo describes both the general form of questions to use and rules for calculating spillover scores from responses to questions. The memo describes the degree of latitude the EDCs have in adapting the methods. EDCs may also add

questions and/or use alternative formulas for calculating spillover scores *in parallel with* the calculations resulting from the methods described in the memo.

The spillover approach is based on self-report. The SWE recognizes that self-reported spillover without verification may be inaccurate, and therefore the EDCs should interpret findings with caution. However, verifying spillover reports through on-site assessment is costly and therefore not required.

The common approach for participant spillover assesses, for each participant:

- The number and description of non-incented energy efficiency measures taken since program participation
- An estimate of energy savings associated with those energy efficiency measures
- The program's influence on the participant's decision to take the identified measures.

Details of assessment and calculation of participant spillover totals and rates are provided in Appendix G.

For EDCs that choose to assess it, nonparticipant spillover may be assessed through either a general population (nonparticipant) survey or through a survey of trade allies. If a general population survey is selected, it should assess, for each survey respondent:

- The number and description of non-incented energy-efficiency measures taken since program participation
- An estimate of energy savings associated with those energy-efficiency measures
- The program's influence on the participant's decision to take the identified measures.

Evaluators should submit draft survey questions to the SWE.

If an evaluator chooses to assess nonparticipant spillover through trade ally surveys, separate surveys should be conducted for the residential and nonresidential sectors. Each survey should assess, for each sampled respondent:

- The number of program-qualified measures sold or installed within the specified sector, the specified utility's service territory, and the specified program year
- The percentage of such installations that received rebates from the specified program
- The trade ally's estimate of the proportion of their sales or installations of non-rebated measures that went to prior program participants
- The trade ally's judgment of the specified program's influence on sales of the common program-qualified but not rebated measures.

Details of assessment and calculation of nonparticipant spillover totals and rates are provided in Appendix G.

The SWE recommends – but does not require – that the evaluation strive to achieve confidence and precision levels sufficient to provide meaningful feedback to EDCs. If nonparticipant spillover is assessed, the sampling approach should produce a sample that is representative of the target population (nonparticipants or trade allies) or capable of producing results that can be made representative through appropriate weighting of data. In the case of trade ally surveys, the sampling plan should take trade ally size (e.g., total sales, total program savings) and type of equipment sold and installed (e.g., lighting or non-lighting) into consideration. Again, the SWE does not specify a minimum level of confidence and precision, but the evaluations should strive to achieve confidence and precision levels sufficient to provide meaningful feedback to EDCs.

3.6.1.2 Econometric Approaches

Econometric approaches may be used to estimate net savings. When used for buildings, these use historical billing data and require a nonparticipant group of similar buildings for which the owner has invested in end-use improvements without program support. When used for estimating changes in sales such as market lift or market share, sales data would be used.

The ideal application for econometric analysis is when customers are randomly assigned to treatment (participant) and non-treatment (nonparticipant) groups, such as with large-scale opt-out programs.[85] The analysis of customer billing data between the two groups distinguishes program effects and net savings. Survey data may be added to this approach to enhance the analysis and interpretation of program effects.

For opt-in or voluntary commercial-sector programs, the evaluator may conduct onsite verification of the energy efficiency level of the equipment and a survey of both participants and nonparticipants. A discrete choice model estimates the "probability" of participation, given certain characteristics and this "probability" is used to calculate net savings.

For opt-in or voluntary residential programs, the evaluator may use a quasi-experimental design with participants and nonparticipants with similar buildings. A second-stage model using survey data can facilitate inclusion of other factors, such as structural and end-user characteristics to explicate the differences between the nonparticipant and participant groups. Often for low-income programs, an econometric model uses rolling-enrollment to capture participation effects.

The primary disadvantages of these two approaches are: 1) the difficulty in identifying comparison groups of similar buildings, or those in which new end-use equipment has been installed, and 2) the additional cost. Further, it is not possible to disaggregate free-riders or to identify spillover, so approaches using econometric modeling provide total net savings estimates.

[85] The term *opt-out* refers to a program design in which customers automatically are enrolled by the EDCs. This is common in some behavior intervention program designs where a randomly selected group of customers is provided information that other customers do not receive.

3.6.1.3 Market Share and Market Effects Studies

Studies of market share and market effects help estimate program effects and provide information on the market needs and changes for energy efficiency. The purpose of measuring market effects is to make appropriate strategic decisions about program offerings and timing so the market for energy-efficient products and services grows more readily than it would have without the program. Like the econometric models just discussed, market effect studies provide an estimate of overall market effects, from which free-ridership and spillover are not disaggregated.

The definition of a *market effect* in the *California Protocols* is "a change in the structure or functioning of a market or the behavior of participants in a market that result from one or more program efforts. Typically, these efforts are designed to increase the adoption of energy-efficient products, services or practices and are causally related to market interventions."[86]

Market share and market effects studies also examine changes in the market and determine the source of those changes. There are four factors to consider in conducting market share and market effects studies.[87]

1. Baseline measurements are very important; these form the basis of comparison and may be measure-specific or program-specific. They should be broad enough to cover possible interactions with other external influences.
2. There needs to be a "theory of change" against which progress is assessed; development of metrics of expected market changes as part of "theory of change" can be useful.
3. Researchers must assess progress toward the metrics of expected change, with particular attention to: changes in advertising, pricing, product stocking, product availability, market share.
4. Program cost-effectiveness, which includes market assumptions, lends itself better to estimating program effects, as the assumptions can be tested and refined throughout program implementation.

In summary, NTGRs will not be applied when determining whether the EDCs met their energy and demand reduction targets in Phase II of Act 129. Net savings studies such as NTG, econometric, or market share and market effects research should be conducted to: 1) monitor the effects the program is having on the market, 2) gain a more complete understanding of attribution of savings, and 3) identify when specific program measures no longer need ratepayer support. Net savings studies can also be

[86] TecMarket Works Team. *California Energy Efficiency Evaluation Protocols: Technical, Methodological, and Reporting Requirements for Evaluation Professionals*. Prepared for the California Public Utilities Commission. San Francisco, CA. April, 2006.

[87] Hoefgen, L., A. Li, and S. Feldman. *Asking the Tough Questions: Assessing the Transformation of Appliance Markets*. *Proceedings of the American Council for an Energy-Efficient Economy Summer Study on Buildings*. In Volume 10, pp. 14-25. August 2006. Herman, P., S. Feldman, S. Samiullah, and K. S. Mounsih. *Measuring Market Transformation: First You Need A Story… Proceedings of the International Energy Program Evaluation Conference.* pp. 3.19-326. August 1997.

useful to determine if program design changes are needed relating to features, such as measure incentive levels, measure eligibility, or program eligibility requirements.

3.7 Process Evaluation

The purpose of process evaluation is to determine if there are ways to alter the program to improve program cost-effectiveness or the program's efficiency in acquiring resources. Process evaluations are a significant undertaking, which must be designed and executed systematically to ensure unbiased and useful results.

The process evaluation consists of in-depth examinations of the design, administration, delivery/implementation, and market response to energy efficiency programs. As with all evaluations, a process evaluation should address the specific program goals. Process evaluations, while they primarily serve the EDC's program staff and management, also provide a vehicle for sharing program design and operational improvements with other professionals in the field. Below are examples of how decision-makers can use the results of process evaluations:

- Improve program performance with respect to internal administration and communications, promotional practices, program delivery, incentive levels, and data management
- Provide a means of improving customer satisfaction and identifying market threats and opportunities
- Provide information to regulators and other interested parties that programs are being implemented effectively and modified or refined as necessary
- Provide a means of contributing to industry-wide knowledge and best practices so that other EDCs can improve their programs

This section provides a minimum set of standards for process evaluations across the EDCs' portfolios that ensure the necessary flexibility and control for program administration and management so the PUC can be confident that the EDCs manage their programs as cost-efficiently as possible.

3.7.1 Process Evaluation Approaches and Timing

Process evaluations use program data, secondary data, document review, and different types of one-on-one or group interviews and surveys to gather information to describe and assess programs. The design for each process evaluation should begin with the program's original design intent and should provide evidence of program progress in achieving its goals and objectives from the perspective of the program's various target audiences. Process evaluations:

- Highlight areas of program success and challenges
- Make recommendations for program modification and improvement
- Identify best practices that can be implemented in the future

Each process evaluation should have a detailed plan that describes the objectives, sampling plan (for surveys, interviews, or focus groups), research activities, and specific issues to be addressed, along with a schedule of milestones and deliverables.[88]

Every program should have at least one process evaluation in every funding cycle or phase. The process evaluation may be either an in-depth, comprehensive process evaluation or one of several types of focused process evaluations. Process evaluations should be timed to coincide with decision points for the program design and implementation process. The primary types of process evaluations are described below:

1. *Standard Comprehensive Process Evaluation* – This includes data collection activities with each of the program's target audiences, including participants, nonparticipants, end users, and trade allies. These are complex projects that require resources and time to implement. The New York State Process Evaluation Protocols[89] provide excellent guidance on the best practices for all process evaluations, and in-depth, comprehensive process evaluations will adhere to the majority of those protocols.
2. *Market Characterization and Assessment Evaluation* – Market characterization and market assessment activities are important to help program staff understand how the market is structured, operating (characterization), and responding to the program offerings (and to activities external to the program [assessment]). Such studies usually focus on specific technologies or product and service types. They are conducted in order to inform program design and redesign, and may be integrated into a comprehensive process evaluation.
3. *Topic-Specific Focused Evaluation* – Not every process or market evaluation must be comprehensive. In cases where a comprehensive evaluation has been conducted recently, it may be appropriate to conduct an abbreviated process evaluation that focuses on specific items, such as program features or ideas program staff want to explore to see if changes to the program are warranted; data collection for this type of evaluation will involve targeted questions to carefully selected audiences.
4. *Early Feedback Evaluations* – New programs, recently updated/modified programs, and pilot programs benefit from early program evaluation feedback. Such evaluations can help program designers and managers refine the program design before full-scale rollout or during the current program cycle. These early feedback evaluations should be short and focus on as few as three to six months of program operation in order to give program staff rapid and specific feedback.
5. *Real-Time Evaluation* – In many cases, process and market evaluation can help programs be more effective if the information on program progress and performance can be conducted and

[88] The SWE reserves the right to review the process evaluation plans.

[89] Johnson Consulting Group. New York State Process Evaluation Protocols. Prepared for the New York State Research and Development Authority, the New York State Evaluation Advisory Group, and the New York Public Service Commission. January 2012. Accessed 4/10/13.
http://www3.dps.ny.gov/W/PSCWeb.nsf/96f0fec0b45a3c6485257688006a701a/766a83dce56eca35852576da006d79a7/$FILE/Proc%20Eval%20Protocols-final-1-06-2012%20revised%204-5-2013.pdf

reported in real time. When evaluators work with program designers and managers during program development and embed the evaluation into the program, data can be collected throughout the implementation period that informs the program staff about opportunities for improvement. Real-time evaluations typically last for one to two years, with ongoing data collection and quarterly to bi-annual reporting that targets the type of information program staff need to gauge their program's progress and effectiveness.

3.7.2 Data Collection and Evaluation Activities

Process evaluation efforts can include a wide range of data collection and assessment efforts, such as:

- Interviews and surveys with an EDC's program designers, managers, and implementation staff (including contractors, sub-contractors and field staff)
- Interviews and surveys with trade allies, contractors, suppliers, manufacturers, and other market actors and stakeholders
- Interviews and surveys with participants and nonparticipants
- Interviews and surveys with people using the technologies (e.g., usability studies of websites)
- Interviews and surveys with key policy-makers
- Observations of operations and field efforts, including field tests and investigative efforts
- Operational observations and field-testing, including process-related measurement and verification efforts
- Workflow, production, and productivity measurements
- Reviews, assessments, and testing of records, databases, program-related materials, and tools
- Collection and analysis of relevant data or databases from third-party sources (e.g., equipment vendors, trade allies and stakeholders, and market data suppliers)
- Focus groups with participants, nonparticipants, trade allies, and other key market actors associated with the program or the market in which the program operates.

The following sections describe in more detail considerations to be taken in data collection.

3.7.2.1 Review of Program Information and Data

Process evaluators glean a wealth of information about the program from data that the program maintains, including: the tracking system; program communications documents (usually electronic); and the materials used for marketing, outreach, and publicity. There also may be process flow diagrams, program theory and logic documents, planning documents, and regulatory documents that set forth the purpose and intention of the program. The process evaluator should be familiar with these documents, using them to understand the context for the program and to provide data in addition to those obtained in interviews.

3.7.2.2 Interviews with Program Managers, Administrators, and Implementers

Program managers and staff are a valuable source of information, as they typically know the program better than anyone. Interviews with lead program planners and managers, their supervisors, and a sampling of program staff, including both central staff and field staff, is the first step in a process

evaluation. Data from these interviews help the evaluator assess the program design and operations in order to recommend any changes to improve the program's ability to obtain cost-effective energy savings.

Subjects important to discuss with these individuals include: communication within the program, communication with customers and stakeholders, and overall understanding of program goals and objectives. In addition, through the interviews, evaluators can get a sense of the program's strengths and weaknesses, its successes, and the quality of work; they then compare and contrast with information stakeholders and participants express during interviews and surveys.

3.7.2.3 Interviews, Surveys, and/or Focus Groups with Key Stakeholders and Market Actors

In addition to program staff, many other individuals are involved in a program, including policy-makers (such as PUC staff); utility managers; key stakeholders (including trade associations and tenant groups); and other market actors, such as product manufacturers, distributors, installation contractors, and service personnel. It is useful to interview a sample from a variety of key market actor groups in order to obtain their insights into what the program is doing well, and what can be improved.

3.7.2.4 Interviews, Surveys, and/or Focus Groups with Participants and Nonparticipants

One purpose of virtually all process evaluations is to understand the customer's experience in order to inform program improvements. Program participants have valuable perspectives on aspects of the program that work well and others that represent barriers to participation or satisfaction. Detailed feedback from participants also is important for determining whether the customer's perceptions of specific program attributes and delivery procedures conflict or mesh with those of program designers and managers. Beneficial detailed feedback can include levels of satisfaction with various elements of the program, such as the: product(s), organization, scheduling, educational services, quality of work performed, attitude of site staff, responsiveness to questions/concerns, and saving levels achieved.

3.7.2.5 Other Types of Data Collection Efforts

There are many other types of data collection methods to consider, including: ride-along observations with auditors or contractors; intercept surveys; mystery shopping; shelf-stocking counts; and electronic, in-person, or mail data collection instead of phone surveys. Similar data to those mentioned above, if collected for programs in other jurisdictions, can be used to draw comparisons or develop best practices. It is essential to select the optimal data collection approach and the appropriate sample, and to draw conclusions consistent with the limits of the data and sample.

3.7.3 Process Evaluation Analysis Activities

The process or market evaluation analysis is considered triangulation. Because much of the data are qualitative, the evaluation team's analysts must be systematic and careful in order to draw accurate conclusions across the different sources.

Evaluators must construct the data collection instruments carefully to ensure that similar questions are posed across groups; it is also essential to select samples that accurately represent the target audiences so that the evaluator's conclusions are justified.

3.7.4 Process and Market Evaluation Reports

Each process evaluation should include the findings from the research tasks, and draw conclusions and recommendations that address the research objectives. The EDC, SWE, and the PUC cannot implement long lists of recommendations. Instead, targeted, actionable recommendations are expected.

Once the EDC conducts a process evaluation, the following will occur:

- The evaluation contractor's process evaluation methodology, findings, and recommendations for all process and market evaluations conducted during the year will be presented in the EDC final annual report (November 15). To provide for some flexibility to EDCs, the required process evaluation information can be provided directly in the EDC final annual report, or alternatively in a separate report that is attached to the EDC final report. If an EDC decides to provide the process evaluation information in a separate report attached to the EDC's final annual report, the final annual report should include an executive summary of each process evaluation.
- The SWE will follow up with the EDC staff to determine how the EDC plans to address each of the process evaluation recommendations made by the EDC's evaluation contractor.
- Through these conversations, the EDC will tell the SWE what, if any, action (accept, reject, still under consideration, etc.) they are planning to take based on the recommendations.
- The SWE will summarize the reports, recommendations, and the EDC's response to the recommendations in its annual report to the PUC (February).

3.8 Cost-Effectiveness

Results from the EDCs' surveys and M&V activities, evaluation reports, audits, and the statewide impact evaluations will be input into a benefit/cost model and other models, as appropriate, to assess the cost-effectiveness of the EDCs' efforts at the measure, program, sector, and portfolio levels. In accordance with the PUC's requirements for determining cost-effectiveness, the EDC's EE&C programs will be evaluated based on the Total Resource Cost (TRC) Test. The guidelines for the TRC are stipulated in 2009 and 2011 Phase I TRC orders, and refined in the Phase II 2013 TRC Order. All cost-effectiveness evaluations and assessments will be conducted in accordance with the PUC's latest TRC Order.

3.8.1 TRC Method

The PUC has adopted the *California Standard Practice Manual: Economic Analysis of Demand-Side Programs and Projects* TRC Test definition, formula, and components with a few slight modifications. Act 129 defines the TRC Test as "a standard test that is met if, over the effective life of each plan not to exceed 15 years, the net present value of the avoided monetary cost of supplying electricity is greater than the net present value of the monetary cost of energy efficiency conservation measures."[90]

According to the California manual:

> *The Total Resource Cost Test measures the net costs of a demand-side management*

[90] *California Standard Practice Manual: Economic Analysis of Demand-Side Program and Projects*: October 2001

> *program as a resource option based on the total costs of the program, including both the participants' and the utility's costs.*
>
> *The test is applicable to conservation, load management, and fuel substitution programs. For fuel substitution programs, the test measures the net effect of the impacts from the fuel not chosen versus the impacts from the fuel that is chosen as a result of the program. TRC Test results for fuel substitution programs should be viewed as a measure of the economic efficiency implications of the total energy supply system (gas and electric).*
>
> *Benefits and Costs: This test represents the combination of the effects of a program on both the customers participating and those not participating in a program.*

EDC evaluation contractors should refer to the 2013 TRC Order for Phase II, and the *California Standard Practice Manual*, for detailed formulae and definitions related to the proper calculation of the PA TRC Test.[91,92]

3.8.2 Application of 15-Year Avoided Cost Streams

The TRC Order limits the effective useful life of any energy efficiency measure to 15 years for the purposes of the benefit/cost calculations but does not specifically address which 15 years of avoided costs should be used. EDCs should follow the guidelines below while developing their TRC models for Phase II of Act 129.

- The 15-year avoided cost stream for each program year should begin with the calendar year at the close of the program year using avoided costs that are calculated by calendar year. For example, for a measure installed in PY5 (June 1, 2013-May 31, 2014) with a 15-year measure life, the avoided cost stream used would be from January 2014 through December 2028.
    - All EDCs should consider using short-term avoided capacity costs forecasts from the PJM Base Residual Auction for TRC calculations, since the PJM delivery year is aligned to Act 129 program years (June 1-May 31).

3.8.3 Aligning Measure Savings with Incremental Measure Costs

To determine energy efficiency cost-effectiveness using the TRC Test, the energy efficiency measure/program savings and costs must be determined and aligned properly. For the TRC Test, the appropriate cost to use is the cost of the energy efficiency device in excess of what the customer otherwise would have spent, regardless of what portion of that incremental cost is paid by the participant or paid by an EDC. Thus, the incremental measure cost should be evaluated with respect to a baseline. For instance, a program that provides an incentive to a customer to upgrade to a high-efficiency central air conditioner would use the cost difference between the efficient air conditioner and

[91] Ibid: October 2001; P. 18.
[92] Pennsylvania Public Utility Commission, *2013 Total Resource Cost Test Order*, Docket No. M-2012-2300653, August 30, 2012.

the base model that otherwise would have been purchased. Similarly, the savings are calculated as the reduced energy consumption of the efficient unit compared to the base model.

Five basic measure decision types are referenced in Table 3-7, along with a summary of the definition of incremental measure costs and savings for each of the decision types.

Table 3-7: Measure Decision Types

Type of Measure	Incremental Measure Cost ($/Unit)	Impact Measurement (kWh/yr/Unit)
New Construction	Cost of efficient device minus cost of baseline device	Consumption of baseline device minus consumption of efficient device
Replace on Burnout (ROB)	Cost of efficient device minus cost of baseline device	Consumption of baseline device minus consumption of efficient device
Retrofit: An additional piece of equipment or process is "retrofit" to an existing system. (e.g., additional insulation or duct sealing)	Cost of efficient device plus installation costs	Consumption of old device minus consumption of efficient device
Early Replacement[93]: Replacement of existing functional equipment with new efficient equipment	Present value of efficient device (plus installation costs) minus present value of baseline device (plus installation costs)	*During remaining life of old device:* Consumption of old device minus consumption of efficient device *After remaining life of old device:* Consumption of baseline device minus consumption of efficient device
Early Retirement (No Replacement)	Cost of removing old device	Consumption of old device

The 2013 TRC Order defines incremental measure cost as either the cost of an efficient device minus the cost of the standard device (ROB), or the full cost of the efficient device plus installation costs (simple retrofit). However, the Order also permits EDCs to utilize the Early Retirement calculation methodology, provided the EDC documents which method they used and why.

[93] The early replacement case is essentially a combination of the simple retrofit treatment (for the time period during which the existing measure would have otherwise remained in service) and the failure replacement treatment for the years after the existing device would have been replaced. Proper treatment of early replacement measure costs is discussed earlier in this Framework (Section 2.2.3).

3.8.4 Data Requirements

To quantify the benefits of energy efficiency and evaluate the cost-effectiveness of individual measures, programs, and EE&C portfolios, evaluators must develop significant general modeling and measure/program-specific data assumptions. A full discussion of these data requirements can be found in the 2013 PA TRC Order[94] or the National Action Plan for Energy Efficiency's "Understanding Cost-Effectiveness of Energy Efficiency Programs" report.[95] Below is a brief list of these data requirements:

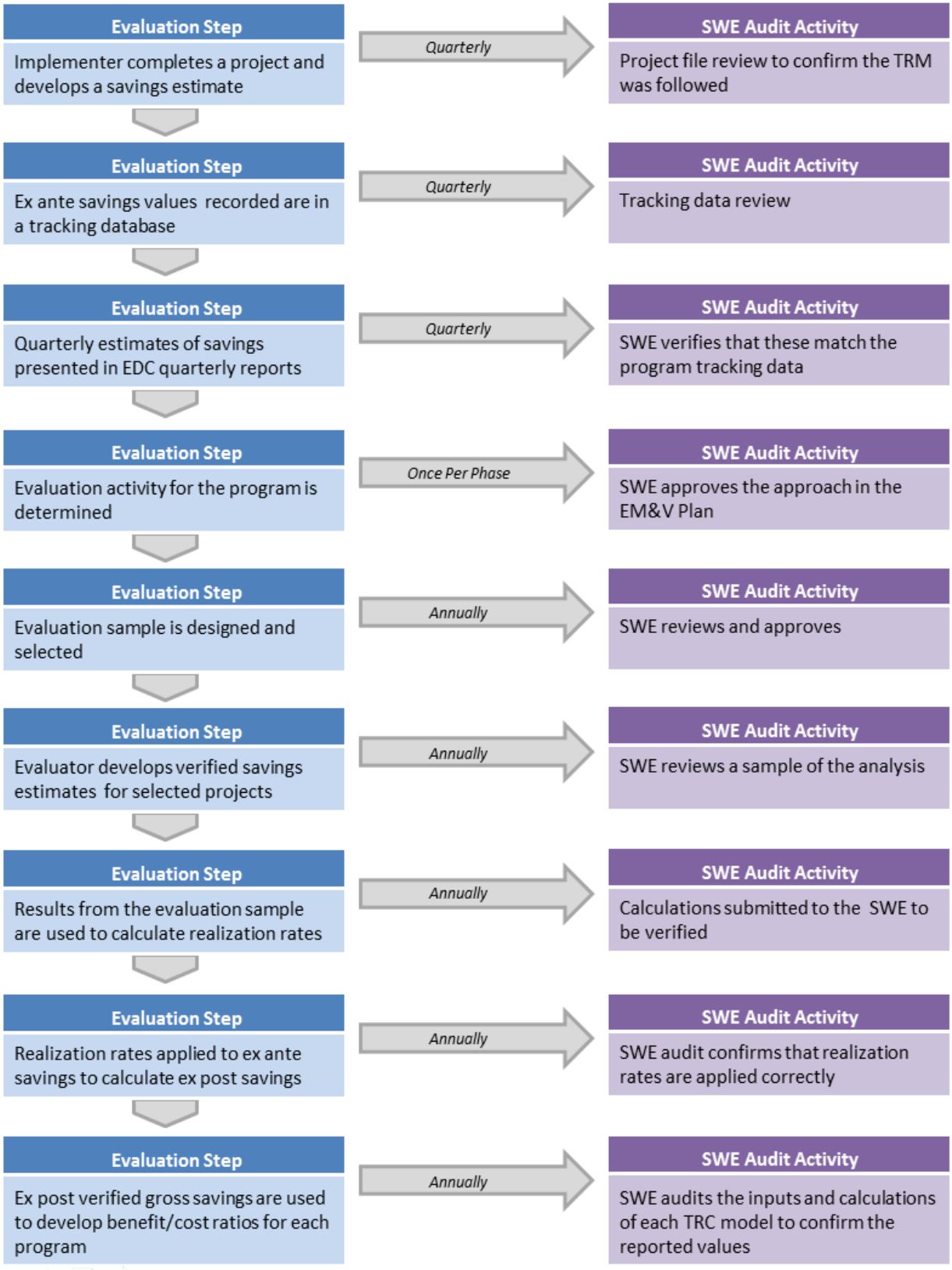
- General Modeling Assumptions
  - Avoided generation energy costs
  - Avoided generation capacity costs
  - Avoided transmission and distribution costs
  - Energy and peak demand line losses
  - Utility Discount Rate
  - General rate of inflation
- Program-/Measure-Specific Assumptions
  - Number of participants
  - Annual energy (kWh) and demand savings (kW)
  - Measure Useful Life
  - Incremental measure cost
  - Avoided O&M benefits (optional)
  - Outside rebates/tax credits (if quantifiable)
  - Additional direct/marketing costs[96] (non-incentive costs)
  - Program/measure load shapes
  - Measure-specific peak coincidence factor

4 Statewide Evaluator Audit Activities

This section describes the actions and activities conducted by the SWE to audit the implementation and the evaluation of each EDC's EE&C plan. This includes review/audit of EDC program delivery mechanisms and all evaluation processes and results submitted by each EDC's evaluation contractor. The overall SWE audit findings should be used to inform the EDC evaluation teams when conducting the actual program evaluations. The SWE will use the audit activity findings, which will parallel the EDC evaluation activities, to assess the quality and validity of the EDC gross-verified savings estimates, net-verified savings estimates, process evaluation findings and recommendations, and benefit/cost ratios. Figure 4-1 shows the specific SWE audit activities and their correspondence to the evaluation steps.

[94] Pennsylvania Public Utility Commission, *2013 Total Resource Cost Test Order*, Docket No. M-2012-2300653, August 31, 2012.
[95] http://www.epa.gov/cleanenergy/documents/suca/cost-effectiveness
[96] Direct or marketing costs include program administration, EDC Implementation CSP, EDC Evaluation contractor etc.

Figure 4-1: SWE Audit Activities



To the extent possible, the SWE will provide the EDCs with "early feedback" on the results of its audit activities – particularly if discrepancies are identified. The intent of early feedback is to allow the EDCs to work with ICSPs and evaluation contractors to implement corrective actions within the program year.

4.1 EDC Report and SWE Report Schedule

The quarterly and annual reports defined by the PUC are one of the ways by which stakeholders are informed about the spending and savings impacts of Act 129 EE&C plans. These quarterly and annual EDC and SWE reports are public documents. This section of the Framework provides an overview of the EDC and SWE reporting requirements for Phase II.

4.1.1 EDC Report Schedule

The EDCs are required to submit quarterly and annual reports to the SWE Team and the BTUS. In the *Phase II Implementation Order* entered August 2, 2012, the PUC noted that Act 129 requires EDCs to submit an annual report documenting the effectiveness of their EE&C plans, measurement and verification of energy savings, evaluation of the cost-effectiveness of their expenditures, and any other information the PUC requires.[97]

The SWE Team has provided the EDCs with quarterly and annual report templates, which are available on the PA Act 129 SharePoint Site (https://sp.gdsassociates.com/sites/39701/SitePages/Home.aspx). The deadlines for the EDC reports are provided in Table 4-1.

Table 4-1: EDC Reporting Schedule

Report	Due	Savings Reported
Program Year X, Quarter 1	October 15	• Quarter 1 Report • Implementation and Evaluation Updates • Gross Reported Savings as of August: IQ, PYTD, and Ph II-Q • Preliminary Verified Gross Savings (if available)
Program Year X, Quarter 2	January 15	• Quarter 2 Report • Implementation and Evaluation Updates • Gross Reported Savings as of November: IQ, PYTD, and Ph II-Q • Preliminary Verified Gross Savings (if available)
Program Year X, Quarter 3	April 15	• Quarter 3 Report • Implementation and Evaluation Updates • Gross Reported Savings as of February: IQ, PYTD, and Ph II-Q • Preliminary Verified Gross Savings (if available)
Program Year X – Preliminary	July 15	• Quarter 4/Preliminary Annual Report • Gross Reported Savings as of May: IQ, PYTD, and Ph II-Q • Preliminary Verified Gross Savings (if available)
Program Year X – Final	November 15	• Final Annual Report • Gross Reported Savings as of May 31 (IQ, PYTD, and Ph II-Q) • Gross Verified Savings as of May 31 (PYTD, Ph II-Q, Ph II-Q+CO)

[97] Implementation Order issued August 2, 2012, at Docket No. M-2012-2289411 and M-2008-2069887.

The preliminary annual reports, final annual reports, and quarterly reports shall be filed with the PUC's Secretary and the SWE Team via the PA Act 129 SharePoint Site. The PUC will post these reports on its website for public access.

Beginning in Program Year 5, the EDC Final Annual Report template will include a section requesting a comparison of actual program performance to the planning estimates filed in their EE&C plans. Requested items will include:

- How did expenditures in the program year compare to the budget estimates set forth in the EE&C plan?
- How did program savings compare to the energy and demand savings estimates filed in the EE&C plan? Discuss programs that exceeded and fell short of projections and what factors may have contributed.
- Are there measures that exceeded or fell short of projected adoption levels? Discuss those measures, if any.
- How did the program year TRC ratios compare to the projected values in the EE&C plan?
- Are any changes to the EE&C plan being considered based on observations from the previous program year?

4.1.2 Statewide Evaluator Report Schedule

In Phase II, the SWE Team will submit semi-annual and annual reports to the PUC with updates on impact evaluations, cost-effectiveness, and process evaluations. These reports will:

- Summarize program and portfolio progress to date for each EDC
- Summarize energy (MWh/yr) savings and peak demand (MW) reductions
- Identify each EDC's savings achievement levels to date
- Identify best practices exhibited to date
- Identify areas for improvements
- Identify any recommendations for adjusting EDCs' reported and verified savings based on the current findings; and
- Provide a summary of audit activities and findings based on the audit work completed during the period.

The reports also will include a summary of general activities corresponding to the responsibilities of the SWE Team. This could include the status of TRM updates, resolutions from PEG meetings, or a summary of recently issued guidance memos.

The deadlines for the SWE reports to the PUC are presented in Table 4-2.

Table 4-2: SWE Reporting Schedule

Report	Due	Savings Reported
DRAFT Program Year X, Mid-Year Report	February 15	• Quarter 1 and 2 reports • Summary of EDC progress • Summary of audit activities and findings • Summary of SWE Team activities
FINAL Program Year X, Mid-Year Report	March 15	• Final mid-year report; comments from BTUS staff and EDCs addressed
DRAFT Program Year X	January 16	• Annual report • Summary of gross savings as of May 31 • Summary of verified savings as of May 31 • Summary of audit activities and findings[98] • Summary of SWE Team activities
FINAL Program Year X	February 27	• Final annual report; comments from BTUS staff and EDCs addressed

4.2 Reported Savings Audit

The SWE will conduct quarterly audits of the ex ante savings values claimed by EDCs and stored in EDC tracking systems. These audit activities are intended to give the PUC confidence in the gross reported savings values presented in EDC quarterly and annual reports. Gross reported savings estimates are the basis upon which the ex post evaluation is conducted.

4.2.1 Quarterly Data Request – Ex Ante

In a standing quarterly data request memo, the SWE Team requested information and data from the EDCs pertaining to the following:

- General implementation and evaluation information
- Residential program data
- Commercial and industrial program data

The SWE Team requested that, going forward, the EDCs provide the same type of information with each quarterly report.[99] All information provided in response to the SWE data request should correspond to activities occurring during the quarter for which the report was submitted. Additionally, the memo included instructions for uploading the data requested to the EDC-specific directory of the Act 129

[98] The SWE will complete its audit activities for a program year in conjunction with the completion of the SWE's annual report to the PUC for that program year.
[99] In this context, the EDC preliminary annual report due July 15 serves as the Q4 report.

SharePoint site page. The most recent version of the SWE Quarterly Data Request can be found on the SharePoint site[100] and in Appendix H.

The SWE requires program-specific information for each program audit.

For the *residential programs*, such as high-efficiency lighting, refrigerator recycling, efficient products, whole building/weatherization programs, and new construction, the SWE requests the entire EDC program database, listing participants and details about the measures installed and savings values for those measures. For high-efficiency lighting, the SWE also requests participant invoices on a quarterly basis.

For *low-income programs*, the SWE's key audit activity is to monitor the site inspections conducted by the EDCs to determine if measures reported as being installed by contractors actually were installed. In addition, the SWE requests on a quarterly basis data relating to program participation and measures installed. Further information on the SWE's data request relating to low-income programs can be found in the Quarterly Data Request document in Appendix H.

For the *commercial programs*, the SWE requests five separate pieces of data. As with the residential programs, the SWE first requests the program database for all participants, detailing the measures installed and reported savings and rebates for those measures. Second, the SWE requests sample documentation (application forms, approval forms, installation confirmation, savings calculation sheets) from a specified number of program participants, depending on the size of the program. For programs with fewer than 50 participants during the quarter, the SWE requires information for 5 randomly selected projects. For all other programs, the SWE requires information for 10 randomly selected projects. These projects should be randomly selected by the EDC or evaluation contractor separately from the sample drawn by the EDC evaluation contractor for the gross impact evaluation. Although the SWE independent sample does not add to the evaluator's sample points, nor are their results used to adjust RR, the SWE independent samples provide additional information for consideration. Thus, having the SWE and EDC evaluation contractor review different projects prevents redundancy and increases the total number of projects reviewed. Because random selection is used, it is possible that a project could be selected for both the SWE and evaluation contractor samples, but this is unlikely. The rest of the items in the SWE quarterly request for nonresidential programs are related to ex post savings analysis; these are discussed in Section 4.3.3.

4.2.2 Desk Audits

As part of its contract with the Pennsylvania PUC, the SWE will complete desk audits for the nonresidential, low-income, residential lighting, residential appliance rebate, residential appliance recycling, and residential new construction programs. These audits will seek to verify the ex ante savings of EDCs' programs by collecting, recording, maintaining, and parsing EDC program data obtained via the

[100] Please note that the data request is subject to change and an updated version of the Evaluation Framework will not be issued if a change is made to the quarterly data request.

SWE data requests described above. The SWE's desk audits will consist of the following three primary elements:

1. A **database review** through which the SWE will verify that EDCs are using the correct values and algorithms from the Pennsylvania TRM in their savings calculations. For deemed measures, the SWE will verify that the EDC used the correct deemed savings value unless otherwise approved by SWE and BTUS. For partially deemed measures, the SWE will use the values from the EDC database to independently calculate savings and verify them against the savings reported by the EDC.
2. **Quarterly and annual report reviews** through which the SWE will verify that the values presented in EDC quarterly and annual reports match the values calculated by the SWE from the EDC database.
3. A **sample check** through which the SWE will cross-check actual program files, receipts, invoices, and work orders against their corresponding database entries to verify that the EDCs have reported program data correctly and consistently. For commercial programs, the SWE will check a sample of EDC project files in a separate "project file review." The project file review is designed to audit the accuracy of the savings values stored in the EDC tracking system and to confirm that the EDCs' calculations were performed in accordance with the current TRM. The uploaded project files include project savings calculation workbooks, specification sheets for equipment installed, invoices, customer incentive agreements, and post-inspection forms. Through these reviews, the SWE will verify that savings values recorded in project files and the program tracking database are consistent.

4.3 Verified Savings Audit

The SWE will conduct an annual audit of the gross impact evaluation methodology and results for each program in an EDC portfolio, and will summarize the findings and recommendations in the final annual report for the program year. The intent of the audit is to provide confidence in the gross verified program savings documented in the EDC annual reports, and transparency in the evaluation process.

4.3.1 Survey Instrument Review

Participant surveys are the most common form of data gathering used by EDC evaluation contractors to collect information about program populations because it is possible to generate a representative and large sample size at relatively low cost. Surveys can be conducted online, in person, via mail, or over the telephone. During Phase II, the evaluation contractors must submit draft survey instruments (in advance of survey implementation) that include process or impact evaluation questions to the SWE for review prior to implementation. A question whose responses will be used as a parameter in a deemed or partially deemed algorithm is considered to be an impact evaluation question. Impact questions for a deemed measure typically involve a straightforward verification that the measure was installed as recorded in the program tracking system. Impact questions for a partially deemed measure could include the size, efficiency, fuel type, replacement protocol, or any other input that affects the savings estimate for the installed measure.

The SWE Team should be alerted via email by EDC evaluation contractors once survey instruments have been uploaded to the SWE SharePoint site for review. The SWE will provide comments and suggest any possible revisions within five business days. Evaluators are not required to change the survey instruments based on the SWE's feedback, but they should consider the guidance carefully. If the evaluators do not receive comments from the SWE within five business days, they can begin implementing the survey. The intent of the SWE review is to confirm that the survey instrument is designed according to industry best practices, that the impact questions will produce accurate and unbiased estimates of program impacts, and that the process questions are clear and will provide useful information for the process evaluation. The following list includes some of the issues the SWE will consider as it reviews survey instruments:

- Are the skip patterns adequately delineated? Are there any combinations of responses that will lead to key questions being omitted from the battery?
- Are any of the survey questions leading or ambiguous? (Improperly worded questions can compromise the reliability of survey results.)
- Are there any missed opportunities? Are there important questions that are not included in the battery, or are follow-up questions needed to answer the research questions?

4.3.2 SWE Annual Data Request

The SWE must submit its annual data request 15 days after the submittal of an EDC's final annual report for a program year. This request includes only the ex post savings analysis the EDC evaluation contractor used to calculate gross verified savings. Responses should be uploaded to the EDC-specific directory of the SWE SharePoint site in a folder titled "PY_ Annual Data Request Responses." The three components of the SWE annual data request are presented below.

4.3.2.1 Evaluation Sample Population

For each program or evaluation group,[101] the evaluation contractor should provide a table that contains the following information for each project in the completed evaluation sample. The number of evaluation groups will vary by EDC according to the design of the portfolio. The underlined terms below may be used as column headers in the table.

- Unique Identifier: This field should correspond to an identifier variable provided to the SWE for the project in the quarterly tracking data for the program; this may be a rebate number, project number, or enrollment ID.
- Stratum: If a stratified sample design is used, in which stratum was this project located?
- Selection Type: When the sample was designed, was this project a primary sample or an alternate?

[101] The term *evaluation group* refers to several similar programs that are grouped for evaluation purposes. For example, an EDC may group projects from its lighting, motor, and HVAC programs within the Small Commercial sector into a single "Small Commercial" evaluation group.

- Evaluation Activity: What type of evaluation activity was performed in order to develop verified savings estimates for this project (e.g., phone interview, online survey, desk review, site inspection, building simulation, or multiple methods)?
- M&V Approach: Which approach was used to calculate the verified savings for this project (e.g., simple verification, IPMVP Option A-D, or other appropriate methodology)?
- Meters Deployed: Was any type of logging equipment deployed at this site to collect information on key parameters in the savings calculations? (Yes/No)
- Verified kWh/yr: What are the verified annual kWh/yr savings for the project?
- Verified kW: What are the verified peak kW savings for the project?

Evaluators should provide the following, if available: supporting documentation showing the sample selection for each evaluation group, and any error roll-up sheets that show the calculation of error ratio/C_v and achieved precision for the evaluation group. For programs that utilize a regression-based analysis of monthly utility bills for an attempted census of participants, it is not necessary to identify each participant in this section of the response. Instead, evaluators should include relevant regression output such as:

- Number of observations used, number of missing values
- ANOVA table with degrees of freedom, F-value, and p-value
- R-square and adjusted R-square values
- Parameter estimates for each of the independent variables, including the associated standard error, t-statistic, p-value, and confidence limits
- Residual plots or other model validation graphics
- Variance Inflation Factors (VIFs) or other tests for multicollinearity

4.3.2.2 Evaluation Sample Audit

The SWE will select a sample of projects from each evaluation group provided in response to Section 4.3.2.1 and provide the EDC evaluation contractor with a list of the Unique Identifiers (UI) for those projects. Within 15 days of receiving the UIs, EDC evaluators must provide the evaluation documentation and findings for each project. The SWE will conduct a desk audit of these projects to confirm the reliability of the savings estimates. There is additional detail regarding these SWE desk audits in Section 4.3.4.2.

The documentation and findings to be supplied by the EDC evaluation contractor will vary per the evaluation approach they used. These items should include:

- Site-specific M&V plans (SSMVPs)
- Completed site inspection reports
- Savings calculations worksheets
- Photos taken during the site inspection
- Building simulation model input and output files, or spreadsheet models used to calculate verified savings

- Monthly billing data used for an Option C analysis
- Data files from end-use metering
- Survey responses

4.3.2.3 TRC Model Audit

The evaluation contractor should submit an electronic version of or provide the SWE access to the model(s) used to calculate the TRC ratios for each EDC program in the EDC final annual report. The TRC model(s) should contain all inputs and outputs to the benefit/cost ratio. Key inputs the SWE will examine include:

- Discount rate
- Line loss factors
- Avoided costs of generation energy and capacity as well as T&D avoided costs
- Incremental measure costs
- Program administration costs
- Verified savings
- Effective useful life of measures or measure groups
- End-use load shapes or on-peak/off-peak ratios used in benefit calculations

4.3.3 SWE Quarterly Data Request

The majority of the information requested in the SWE's quarterly data requests is intended to support the SWE audit of ex ante savings values. This includes program tracking data, equipment invoices, project files and results of onsite inspections conducted during the quarter on each program, including residential programs. In addition, this request will pertain to the ex post savings analysis, which the EDC evaluation contractor will use to calculate gross verified savings. The three components of the SWE quarterly data request that pertain to the ex post savings analyses are presented below.

4.3.3.1 Evaluation Sample Audit

For each program, or evaluation group,[102] the evaluation contractor should provide information regarding the projects in the sample drawn by the EDC evaluation contractor. Depending on the evaluation approach the evaluator used, the evaluator should supply the following documentation and findings:

- **Description of the process they used to select the sample** – If they used a batch process to combine quarters, they must explain their rationale for doing so.
- **List of samples and alternates selected by the EDC evaluator** –This list must include key fields such as the unique identifier, reported energy savings, and reported demand savings.
- **Participant-submitted documentation for sampled projects, containing:**
    - Application forms, approval forms, installation confirmation
    - Equipment specifications, invoices, certifications

[102] Ibid

- Savings calculations sheets
- Other pertinent information

The SWE will review the evaluator's sample design to check the target confidence and precision levels, assumed coefficient of variation, and stratification techniques for each program in the EDC's portfolio. The SWE will ensure that the evaluator-sampled projects are in the EDC program population and note any discrepancies.

4.3.3.2 Evaluation Audit for SWE Ride-Along Site Inspections

For each nonresidential program, the evaluation contractor should provide verification data for projects in the sample drawn by the EDC evaluation contractor. The evaluator may provide this documentation as it becomes available; it is not required when the SWE quarterly data request is due. The evaluator must submit documentation and findings that are appropriate for the evaluation approach they used. At a minimum, these should include evaluation reports containing:

- Site-specific M&V plans (SSMVPs)
- Site findings
- Savings calculations
- Other pertinent information

The SWE will review the evaluator's sample list and inspection schedule to select priority sites for ride-along inspections. The SWE will request project files, evaluator M&V plans, and other supporting information prior to the visit. The SWE will closely coordinate with the EDC evaluation contractors to conduct inspections. After the inspection is completed, the SWE will review the site inspection reports submitted by the EDC evaluators. The SWE will note findings and recommendations in its ride-along site inspection reports. These reports will describe any discrepancies between the EDC evaluators' and SWE's findings. Additional detail regarding these SWE evaluation audits for ride-along inspections is provided in Section 4.3.4.3.

4.3.3.3 Ad-Hoc Audit for SWE Independent Site Inspections

EDCs or their evaluation contractor should provide the following information for nonresidential projects as requested by the SWE Team on an as-needed basis. This information is required when the SWE: identifies specific issues regarding a particular project, reviews project information prior to ride-along site inspections, or prepares for independent site inspections.

- Complete project documentation (electronic forms or scans) containing:
  - Application forms, approval forms, installation confirmation
  - Equipment specifications, invoices, certifications
  - Savings calculations sheets
  - Other pertinent information
- Evaluation reports containing:
  - Site-specific M&V plans (SSMVPs)

- Site findings
- Savings calculations
- Other pertinent information

The SWE will select a sample of projects for independent site inspections, excluding any sites already contacted by EDC evaluators. The SWE will coordinate with each EDC to obtain project files, schedule inspections, and conduct inspections. The SWE will note findings and recommendations in the independent site inspection reports. Additional detail regarding these SWE evaluation audits for independent site inspections is provided in Section 4.3.4.3.

4.3.4 SWE Verified Savings Audit Activities

The SWE will present the findings and recommendations from its annual audit activities in its annual report for each program year. Unless errors are discovered, or the SWE has significant concerns about the methodology used to calculate verified savings for an EDC program, the SWE will recommend that the PUC accept the verified savings provided in the EDC's annual report. If an EDC reports program savings using more than one calculation methodology, the SWE will offer its professional opinion regarding which method produces the most accurate representation of the program impacts in the SWE annual report. This situation typically arises when an EDC believes that a TRM algorithm or value does not accurately reflect the impact of a measure or the conditions in its service territory. In such cases, the EDC evaluation contractor will present the savings impacts using both the TRM savings protocol and the protocol deemed more appropriate for the measure. The SWE will review the savings protocol proposed by the EDC evaluator and provide a recommendation to the PUC to approve or reject the protocol. The SWE's recommendation should not be construed as PUC approval, as the PUC has the ultimate authority to approve or reject the savings calculated using the proposed protocol.

The majority of the SWE's findings and recommendations will be addressed prospectively in annual TRM updates, evaluation plans, and other M&V protocols used by the EDC evaluation contractors. Data gathered during the audit of an EDC program may be supplemented with best practice recommendations and techniques from other EDCs or national sources. The focus of the SWE's prospective recommendations will be to enhance program delivery and cost-effectiveness and improve the accuracy of savings protocols used by the ICSPs and EDC evaluation contractors.

4.3.4.1 Sample Design Review

The precision requirements for the gross impact evaluation of Act 129 programs were described in Section 3.4.1. The SWE will review the EDC evaluation contractors' sampling approaches at three stages during program evaluation.

1. **Evaluation, Measurement, and Verification (EM&V) Plan** – A thorough evaluation plan is an essential component of a successful evaluation. Sample design is one of many issues addressed in the EM&V plan for a program. The plan should outline who will be contacted, how many will be contacted, what type of evaluation activity will occur, and when the evaluation activity is expected to occur. During its review of EDC EM&V plans, the SWE will consider the proposed sampling plan and request revisions, if needed. It is important to note that the EM&V plan is

assembled in advance of the program year, so the sample design must be flexible enough to adapt if program participation patterns differ from expectations.

2. **Quarter 3 of the Program Year** – Within a month of filing its Q3 report (i.e., by May 15) for each program year, evaluation contractors should submit an updated sampling plan for each EDC program. At that point in the program year, it is possible to estimate the final disposition of the program population for the year more precisely. The SWE will approve the EDC evaluation contractor's sampling plan for the program year via telephone or email exchanges. If the SWE has concerns about the sample size, sample disposition, or level of rigor used within the sample, the SWE will suggest modifications.
3. **SWE Final Annual Report** – Following the close of each program year, the SWE will review the evaluated results of each EDC program and provide recommendations for future program years. If the SWE feels a particular technology was under-represented in the evaluation sample, the annual report will contain a recommendation to focus more heavily on that technology the following year. If the evaluator's variability estimates (C_v or error ratio) proved to be too high or too low, the SWE will recommend changes to the sample design for the following year. For programs that rely on participant surveys, the SWE will examine the sample frame carefully to assess whether there is any appearance of non-response bias or self-selection. If the SWE identifies any concerns, it will discuss the issue and suggest possible corrective actions.

4.3.4.2 Desk Audits

The SWE audit of the EDC evaluations will include all review activities required to assess the quality control, accuracy, and uncertainty of verified savings estimates. Annually, the SWE Team will request verification data for projects in the sample drawn by the EDC evaluation contractor for each EDC program as described in Section **Error! Reference source not found.**. Typically, projects for the SWE valuation Sample Audit will be selected after the EDC annual report has been filed from the evaluation sample population submitted as part of the SWE Annual Data Request. If an evaluation contractor completes a significant share of the verified savings analyses for a program year in advance of the reporting deadline (November 30), the SWE will consider a two-stage sampling process to allow increased discussion prior to the inclusion of audit findings in the SWE Annual Report. The SWE will audit the M&V methods used by the evaluator to ensure the verified savings are calculated using approved protocols.

The SWE will review the evaluation processes and compare them with the approved evaluation plans. In addition, for quality assurance, the audit activities will include some ex ante savings checks such as: a review of randomly selected incentive applications, verification of the proper application of TRM assumptions, and assessment of the consistency of data between incentive applications and the EDC data tracking system. The evaluation reports requested from the EDC evaluation contractor should include the following information:

- Site-specific M&V plans (applicable only to commercial and industrial programs), clearly showing the data collection process and how it is utilized in savings analysis
- Site inspection findings (applicable to all programs)

- Description of metering methods, including measurement equipment type, location of metering equipment, equipment set-up process, photographs of meter installation, metering duration for which data were collected, metered results, and accuracy of the results
- Savings calculations, with all supporting information
- Incentive applications, and
- Other pertinent information

In its annual reports, the SWE will document findings and recommendations resulting from these desk audits, as well as actions taken by EDCs to address them. If an EDC evaluation contractor submits verified savings analyses for audit before the November 30 due date, the SWE will work to provide audit findings and recommendations to the EDCs for review and discussion prior to documenting them in the SWE's annual report.

4.3.4.3 Site Inspections

Site inspections are essential for the accurate evaluation of programs and will represent a significant portion of the EDCs' evaluation efforts for residential and nonresidential programs.[103] Because of the importance of this task, the SWE Team will work closely with the EDCs to ensure that site inspections are planned and executed carefully and that site inspectors have the appropriate experience and training. The SWE Team will audit the following steps in each EDC's site inspection process:

- Training of site inspectors to collect site-specific information
- Random sampling of projects
- Development of the evaluation tracking database and site inspection forms
- Grouping of inspections by geographic location (as appropriate) to minimize time allocation, labor, and direct costs associated with conducting inspections
- Contacting sites prior to any visit to ensure availability and to ensure the resident or facility staff is not "surprised" by the visit
- Performing site inspections and entering all required data into the program evaluation database.

In general, the SWE audit activities will fall into two categories:

1. **Ride-Along Site Inspections (Audits):** The SWE may perform "ride-along audits," in which the SWE accompanies the EDC evaluator on a site inspection to validate and confirm that EDC evaluators are using approved protocols when performing evaluation activities. This includes checking for adherence with the TRM, where applicable, and compliance with the SWE Evaluation Framework. The ride-along audits are a sub-set of the EDC evaluation sample, focusing on high-impact and high-uncertainty projects. The site-specific savings may be adjusted based on the SWE's findings and recommendations.

[103] SWE site inspections are typically focused on large nonresidential projects, but may include a small number of site visits for low-income or residential new construction in Phase II.

2. **Independent Site Inspections (Audits):** Although less frequent than ride-along audits, the SWE may perform an independent audit of any project in the program population with either high impact or high uncertainty, as determined by the SWE at any point in the program year. This may include sub-samples of the EDC evaluation sample or projects outside the EDC evaluation sample. The SWE will conduct relatively fewer independent site inspections than ride-along inspections. The SWE expects to conduct more independent inspections at the beginning of each Phase and then fewer such inspections as it becomes more confident that the ICSPs' reported savings estimates and evaluation contractors' verification activities are accurate. Independent site inspections will include a detailed assessment of the measures beyond what would be performed by the SWE during ride-along inspections, to ensure that the measures are being operated to yield the energy and demand savings claimed in the rebate application. As appropriate, independent site inspections will include spot measurements or trending of important performance parameters and independent verified estimates for energy and peak demand savings.

The SWE is committed to working collaboratively with the EDCs and the EDC evaluators to conduct audit activities and ensure the accuracy of ex ante savings and realization rates that support unbiased estimations of verified gross energy and demand impacts for the Act 129 programs.

The SWE will produce and distribute its ride-along site inspection reports (RA-SIRs) and independent site inspection reports (I-SIRs) to EDC evaluators within 15 business days of completing a ride-along to document its site inspection findings and verified savings calculations. In the case of ride-along inspections, the EDC evaluation contractors will calculate verified savings and SWE inspectors will verify them. Findings and recommendations resulting from RA-SIRs and I-SIRs, as well as actions taken by EDCs to address the findings and recommendations, will be documented in the SWE quarterly and annual reports.

1. **Ride-Along Site Inspection Reports:** RA-SIRs will focus on process findings that also may affect the gross impacts verified by the evaluation contractors. The SWE also will review evaluators' site inspection reports to ensure that all savings calculations and critical site findings have been identified. The RA-SIRs will be completed after the EDC evaluators have shared their site inspection reports and engineering calculations with the SWE. EDC evaluators will have the opportunity to review RA-SIRs and discuss key issues and/or discrepancies with the SWE. Resolutions will be reached collaboratively by the SWE and the EDC evaluators.
2. **Independent Site Inspection Reports:** If an independent site inspection is completed by the SWE, I-SIRs will include process findings related to program delivery and an independent SWE assessment of ex ante project impacts. The SWE will calculate verified savings for all independent inspection samples. Because independent site inspections are conducted on sites not selected by the EDC evaluation contractors, I-SIRs will be issued shortly after SWE evaluation activities have been completed.

If the SWE Team elects to conduct an independent site inspection, the EDC and evaluation contractor will be notified well in advance of the visit. Verified savings estimates from projects receiving a SWE I-SIR can be included in the gross impact evaluation sample and subsequent realization rate calculation at the discretion of the EDC evaluation contractor. EDC evaluators will not be required to incorporate the results from I-SIRs in the final realization rate calculations. As appropriate and with substantial justification, the SWE will request further quarterly and annual information on specific observations made during independent site inspections. The EDC evaluators will be responsible to address the SWE's independent observations in a timely manner.

4.4 Net Impact Evaluation Audit

Any Act 129 net impact research will be audited by the SWE. Further, EDCs are expected to conduct net impact research to inform program planning.

4.4.1 Research Design

The SWE will audit the research design as part of the review of the EM&V plan, and again as part of the review of the reported results. The audit will assess whether the approach used is consistent with common methods recommended for downstream programs and for appliance retirement programs (Appendix E and Appendix F).

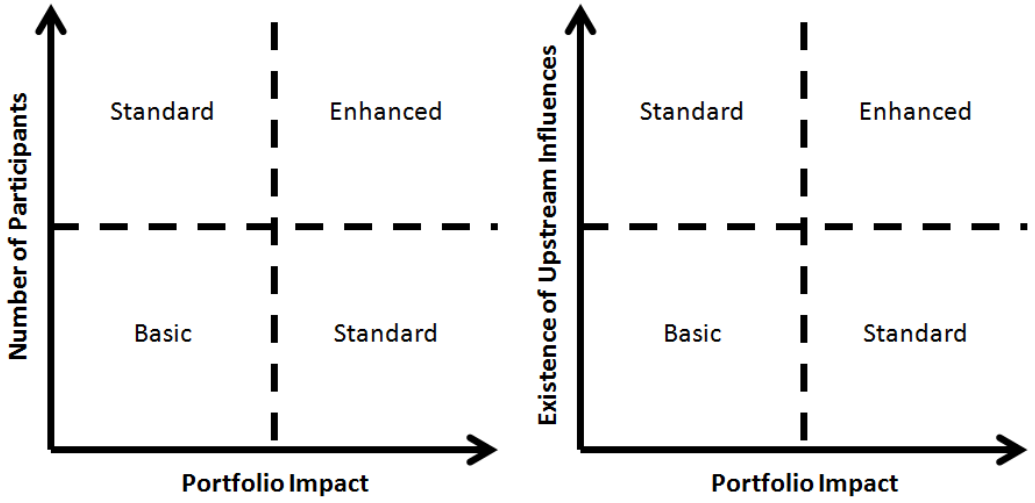
For programs that cannot use the common method, the audit review will be based on the SWE-defined levels of rigor of analysis in the SWE Net-to-Gross Study Methods guidance document distributed to the EDCs and their evaluation contractors on February 27, 2012, which remains the document in effect for programs not addressed by the guidance memos. The levels of rigor (basic, standard, and enhanced) and the methods involved in each are outlined in Table 4-3.

Table 4-3: Rigor Levels Adapted from the California Energy Efficiency Evaluation Protocols

Rigor Level	Methods of Net Impact Evaluation (Free-Ridership and Spillover)
Basic	• Deemed/stipulated NTG ratio • Participant self-reporting surveys • Expert judgment
Standard	• Billing analysis of participants and nonparticipants • Enhanced self-report method using other data sources relevant to the decision to install or adopt a measure. These could include record/business policy and paper review; examination of other, similar decisions; interviews with multiple actors and end users; interviews with midstream and upstream market actors; and interviews with program delivery staff. • Market sales data analysis • Other econometric or market based studies
Enhanced	• Triangulation. This typically involves using multiple methods from the standard and basic levels, including an analysis and justification of how the results were combined.

Method selection should follow the recommended threshold guideline based on a program's contribution to total portfolio savings. If the energy savings of an EDC's program is less than or equal to 5% of the EDC's total portfolio energy savings, a basic level of rigor analysis (e.g., stipulated/deemed or simple survey) is acceptable to estimate NTGRs. If the energy savings of an EDC's program is greater than 5%, the SWE recommends a more complex approach to determine whether the basic, standard, or enhanced level of rigor were appropriate. These recommendations are based on benefit/cost considerations, as the added costs of a greater level of rigor generally are unwarranted for programs with low savings contributions. Figure 4-2 outlines the approach EDCs should use for programs with portfolio contribution percentages greater than 5%. NTG sections in EDC annual reports will be expected to explicate how these factors were used to guide the decision regarding the level of rigor to apply to estimate NTGRs.

Figure 4-2: Program Attributes Quadrant Maps



4.4.2 Sample Design

The audit will determine whether the sampling was appropriate. Probability sampling (described in sections 3.4.2and 4.5.2) should be used for net savings or market share/market effects studies. The sample design will be audited as part of the review of the EM&V plan, and again as part of the review of the reported results.

4.4.3 Transparency in Reporting

The audit requires that the EDC and their evaluation contractors describe the reasons the approach was selected, the sample, the questions used, and the methods used in the analysis and application of the NTGR. Such information should include the methodology, data collection, sampling, survey design, algorithm design, and analysis. Free-ridership or NTG ratios should include explanation or description regarding how they were derived. A transparent approach to net savings is necessary for an effective and useful audit.

4.4.4 Use of Results

The audit also will examine how the EDC and its evaluation contractors are using the results for the purposes of modifying and improving program design and implementation while operating within Act 129 budget, cost-effectiveness, and compliance constraints.

4.5 Process Evaluation Audit

The SWE will audit process and market evaluation research plans, data collection instruments, and final reports to ensure that the:

- Research objectives are complete relative to the type of process or market evaluation planned.
- Sample design is sufficient and appropriate to address the objectives.
- Data collection approaches are appropriate and executed per plan.
- Data collection instruments address the objectives and do not introduce bias.
- Analysis and report writing convey the findings clearly and draw reasonable conclusions.
- Recommendations are actionable and clearly identify which parties should address the recommendation.
- EDCs follow up on process evaluation recommendations and report to the SWE the action the EDC has taken on each recommendation.

4.5.1 Guidance on Research Objectives

The SWE audit will review the process evaluation with expectations that the process evaluation will address objectives as appropriate to the program. Examples of objectives that may be relevant to a program are noted below.

4.5.1.1 Program Design

- Program design, design characteristics, and design process
- Program mission, vision, and goal-setting and process
- Assessment or development of program and market operations theories and supportive logic models, theory assumptions, and key theory relationships - especially their causal relationships
- Use of new practices or best practices

4.5.1.2 Program Administration

- Program oversight and improvement process
- Program staffing allocation and requirements
- Management and staff skill and training needs
- Program information and information support systems
- Reporting and the relationship between effective tracking and management, including both operational and financial management

4.5.1.3 Program Implementation and Delivery

- Description and assessment of the program implementation and delivery process
- Clarity and effectiveness of internal staff communications
- Quality control methods and operational issues
- Program management and management's operational practices
- Program delivery systems, components, and implementation practices
- Program targeting, marketing, and outreach efforts
- The level of financial incentives for program participants
- Program goal attainment and goal-associated implementation processes and results
- Program timing, timelines, and time-sensitive accomplishments
- Quality-control procedures and processes

4.5.1.4 End-User and Market Response

- Customer interaction and satisfaction (both overall satisfaction and satisfaction with key program components, including satisfaction with key customer-product-provider relationships and support services)
- Customer or participant energy efficiency or load reduction needs and the ability of the program to provide for those needs
- Trade allies' interaction and satisfaction
- Low participation rates or associated energy savings
- Trade allies' needs and the ability of the program to provide for those needs
- Reasons for overly high free riders or too low a level of market effects, free drivers, or spillover
- Intended or unanticipated market effects

4.5.2 Sample design

Sampling for process and market evaluations should follow sampling approaches similar to those used for impact evaluations whenever it is important to generalize to the population. (Note, this does not mean that the sampling should be the same for impact and process and market evaluation, just that the approaches when generalization is important are similar.) Table 4-4 outlines the three primary options for sampling; all may be used with process and market evaluations when appropriate. Section 3.4.2 provides additional guidance on probability sampling.

Table 4-4: Sampling Options

Option	What Is Measured	Applicability of Precision Estimates	Rank Order of Defensibility
Census	Measures the entire population, so results represent the entire population	Statistical precision is not applicable because it counts every outcome and, therefore, provides a full rather than partial enumeration.	Highest
Probability Sample: Simple random and stratified random	Measures a randomly selected subset of the population, therefore the probability selection to the sample is known and results can be generalized to the population	Sampling precision depends on the number of items; e.g., participants measured. The more measured, the better the precision.	Varies
Systematic Sample: Any non-random method of sampling	Measures a non-randomly selected subset of the population, so the probability of selection to the sample is unknown, and generalization to the population is not possible	Statistical precision is not applicable. Carefully selected representative samples sometimes are claimed to have properties "similar to" probability samples.	Lowest

Non-probability samples sometimes are acceptable for process and market evaluations. When sampling from small groups in which a census or near-census is possible, precision and confidence do not apply, and a census or near-census should be pursued. Non-probability samples also are acceptable when the purpose is to gain a greater sense of knowledge of the topic and not to generalize. In such cases, systematic sampling is acceptable. Evaluators must ensure that they have used robust, systematic sampling approaches and have articulated the justification for using a non-probability sample clearly in the process evaluation section of the EDC final annual report.

The process and market evaluators must identify the population, prepare an appropriate sampling frame, draw the sample consistent with the frame, and ensure that inference is consistent with the sampling approach.

4.5.3 Data Collection Instruments

The SWE must review all data collection instruments (in advance of survey implementation) and complete the review within five business days per the guidelines below.

4.5.3.1 General Instrument Characteristics

The SWE reviewers will audit the instruments looking for various elements as described below:

- Title: including contact type (e.g., program staff, participants, nonparticipants, trade allies, industry experts)
- Statement of purpose (brief summary for interviewer, client, and survey house)
- Listing and explanation of variables to be piped into the survey and the source of these values (if applicable)
- Instructions to the interviewer/survey house/programmer regarding how to handle multiple response questions (e.g., process as binary)
- Scheduling script: collect time and date for re-contact, verification of best and alternative phone numbers
- Brief introduction: mentions client and requests client feedback for appropriate purposes
- Statement as to whether responses will be treated as confidential or will not be reported
- Screening questions: if needed, and if interviewer instructions include directions regarding when to terminate the survey
- General flow: from general questions directed to all contacts through specific topics (with headings), including skip patterns where needed
- Insertion of intermittent text, or prompts, to be read by the interviewer, informing the contact of new topics that also serve to improve the flow of the interview
- Use of a SWE standard set of demographic /firmographic questions (e.g., comparable to Census or industry data)
- If needed, request for permission to call back or email with follow-up questions (especially useful when conducting in-depth interviews); collection of appropriate call-back information, best phone, email address, etc.
- Request for any additional comments from respondent
- Conclusion, with a thank-you message

4.5.3.2 Question Review

The SWE will check for and comment on questions that are:

- Double-barreled (this *and* that)
- Leading and or biased (questions that encourage participants to respond to the question in a certain way)
- Confusing or wordy (editing for clarity)
- Appear not to be related to research issues or analysis plan
- Are related to research issues or analysis plan but do not appear to achieve the research objectives
- Clearly indicate whether to read or not read responses and when multiple responses are accepted
- Missing a timeframe anchor (e.g., in the past year)

- Driven by a skip pattern (Survey developers and reviewers must check that the skip is necessary, and is asked of all contacts, if at all applicable. It is best to avoid skips within skips that reduce the size of the sample.)
- General readability

4.5.4 Analysis Methods

The EDCs must use the appropriate levels of analysis for process evaluation data. Inference from the data should be consistent with the sampling strategy, and claims should not overreach the data. Data will be either qualitative or quantitative.

4.5.4.1 Qualitative Analysis

The EDC evaluators should respect the respondents' rights and not report names or affiliations except at a general level (e.g., program staff, implementers, customers, contractors, and trade allies). Reports should clearly document the program activities and lessons learned from the research. Findings should permit the reviewer to understand the data source for the finding and to understand how different audiences responded to the research objectives. The population always should be clearly defined, and all tables and reported data should clearly articulate the portion of the sample responding for the finding [e.g., 7 of 10 people, or seven said (n=10)] and that tables are clearly labeled.

4.5.4.2 Quantitative Analysis

The EDC evaluators should ensure that response dispositions are tracked and reported consistent with the guidance of the Council of American Survey Research Organizations (CASRO).[104] The population always should be clearly defined, and all tables and reported data should clearly articulate the portion of the sample responding for the finding [e.g., 70% (n=349)] and ensure that tables are clearly labeled.

Further, the EDC evaluation contractor should use appropriate quantitative methods. For instance, if data are ordinal – means should not be used – the top two boxes are acceptable. If data are not normally distributed, non-parametric tests should be used. Similarly, evaluators should choose statistical tests and analysis methods carefully to ensure that they are appropriate for the data collection process.

4.5.5 Assessment and Reporting by the SWE

The SWE process evaluation assessment will include a review of findings and recommendations relative to program design, program delivery, administrative activities, and market response. These findings will be reported in the SWE Annual Report.

- The SWE review of process findings for these various programs by EDC will help to identify best practices across the state.
- The SWE also will compare process evaluation findings to process and delivery strategies of similar best programs throughout the United States.

[104] See CASRO Research Guidelines at http://www.casro.org/?page=ResearchGuidelines

- The SWE will present the findings in a manner that highlights areas of success within the portfolio of EDC projects and that identifies areas of improvement.
- The SWE also will report on selected EDC responses to the recommendations.

4.6 Cost-Effectiveness Evaluation Audit

The SWE cost-effectiveness assessment will include a review of the benefit/cost (B/C) ratio formulas, benefits, costs, and TRC ratios at the EDC project level, EDC program level, and EDC plan level. The SWE will determine whether TRC calculations have been performed according to the PUC's latest TRC Order and whether EDCs are on track to meet the Act 129 cost-effectiveness requirements.

4.6.1 Annual Data Request

The SWE Team will request each EDC to submit an electronic version of the model(s) used to calculate the TRC ratios in the EDC's final annual report. The TRC model(s) should contain all relevant general modeling and program-specific inputs to the B/C ratio, calculation formulas, and TRC outputs.

4.6.2 Inputs and Assumptions

Key inputs and assumptions the SWE will examine include:

- Discount rate
- Line loss factors
- Avoided costs of energy and capacity
- Incremental measure costs
- Program administration costs
- Verified savings figures
- Effective useful life of measures or measure groups
- End-use load shapes or on-peak/off-peak ratios used in benefit calculations

4.6.3 Calculations

Possible audit activities pertaining to the cost-effectiveness protocols, calculations, and evaluations may include, but are not limited to:

- A review for TRC Order compliance regarding:
    - Formulas
    - Benefits
    - Costs
    - Utility avoided costs assumptions

- A review of EDC accounting practices, including:
    - Division of costs and benefits between programs
    - Appreciation/depreciation rates

For Phase II, several EDCs are adopting the use of Cadmus' DSM Portfolio Pro for their TRC analysis. DSM Portfolio Pro is a *SQL-Server*®-based model accessed via the Web. The model uses a transparent

calculation methodology that is consistent with the *California Standard Practice Manual*. For accurate benefit estimates, DSM Portfolio Pro uses hourly load shapes and avoided costs, a built-in, yet customizable measures database, and flexible rate definitions (flat, seasonal, time-of-use, and hourly). For EDCs using DSM Portfolio Pro, the SWE Team would perform, at a minimum, a thorough one-time benchmarking of DSM Portfolio Pro's TRC calculations to verify that results are reasonable and accurate. EDCs would continue to be required to provide inputs and outputs to the SWE Team in *Excel* for annual reporting purposes.

5 Resources and Meetings

This Evaluation Framework is intended to serve as a resource for EDC program administrators and evaluation contractors. The Framework is a living document and will be updated annually in Phase II, however we suggest that stakeholders familiarize themselves with several additional resources to stay informed of the latest developments related to the evaluation of Act 129 EE&C plans.

5.1 Pennsylvania Act 129 Public Utility Commission Website

The SWE will provide documents for sharing on the PUC's public website,[105] which provides information to interested stakeholders on the actual kWh/yr and kW savings from the Act 129 programs, as well as the EDCs' expenditures on such programs. During Phase I of Act 129, the site presented information primarily in report format to interested parties. The SWE Team plans to work with the BTUS staff to identify additional content, such as SWE audit reports and statewide savings summaries, to present on the public website.

5.2 Pennsylvania Act 129 SharePoint Site

The SWE team created a PA Act 129 SharePoint site[106] to improve communication and coordination of activities among the SWE Team, the BTUS, the EDCs and their evaluator contractors, and the Energy Association. This SharePoint site serves as a repository of documents and data associated with the statewide evaluation of the EE&C Program Portfolios implemented by the seven EDCs. The structure and operation of this SharePoint site comply with the confidentiality provisions in the SWE Team contract with the PUC and the Energy Association.

The PA Act 129 SharePoint site contains several pages. Individual access to each page is based upon assigned administrator privileges and confidentiality of content and the Nondisclosure Agreement signed by all parties and referenced in the document "Contract Act 129 Statewide Evaluator" (Issuing Office: Pennsylvania Public Utility Commission Bureau of Conservation, Economics, and Energy Planning;[107] RFP-2009-1).

[105] The URL for the Act 129 directory of the PUC's website is: http://www.puc.state.pa.us/filing_resources/issues_laws_regulations/act_129_information.aspx
[106] The URL for the SWE SharePoint site is: https://sp.gdsassociates.com/sites/39701/SitePages/Home.aspx
[107] The PA PUC Bureau of Conservation, Economics, and Energy Planning (CEEP) is now known as the Bureau of Technical Utility Services (BTUS).

The PA Act 129 SharePoint site pages include:

- **SWE site (Home),** which provides a common interface for all parties directly involved in the statewide evaluation efforts and that have been granted access to the Act 129 SharePoint Site. This home page includes the following features: calendar, task lists, technical libraries, report libraries, submission logs, and discussion boards.
- **SWE Team site,** whose access is restricted to members of the SWE Team and the BTUS staff. The purposes of the SWE Team directory are to facilitate coordination of SWE Team activities, track progress, and store lists of unresolved issues.
- **Individual EDC password-protected sites,** which are tailored to each EDC's needs and include features such as submissions library, task lists, and memo libraries.

Currently, the overall SharePoint site is accessible by more than 80 unique users who have been approved by the SWE Team and the EDCs for which the users work. For Phase II of Act 129, the SharePoint site has been restructured and reorganized to facilitate navigation and access to documents and data. An archive folder has been created within each of the individual EDC directories to house all files pertaining to Phase I. The Phase I archive sections are consistent among each of the individual EDC directories and contain Level 1 folders by program year and Level 2 folders for documents such as reports, tracking data, and data requests/responses.

For the Phase II folder, the SWE will create Level 1 folders for each program year, and Level 2 folders to house documents such as reports, tracking data, and data requests/responses. The Level 1 and 2 folder structure will be consistent across the individual EDC directories. In addition, the SWE will organize the new homepage Level 1 folders by program year. Each folder will house PEG meeting minutes and agendas, as well as the final versions of the SWE semi-annual and annual reports. Additionally, the homepage will maintain all of the SWE guidance memos, the master contact list, a tracking spreadsheet for all data requests and responses, and a calendar with important dates.

5.3 Program Evaluation Group Meetings

The SWE will chair and set the agenda for quarterly meetings of the PEG and will prepare minutes of these meetings. These meetings will be conducted per the same format used during Phase I of Act 129.

5.4 Stakeholder Meetings

Key members of the SWE Team will attend stakeholder meetings and deliver presentations on the results of baseline studies, market potential studies, and recommendations for program modifications and targets for Phase III of Act 129.

6 Final Remarks

The primary objective of the EDC EE&C programs is to reach the level of savings specified in Act 129 in a meaningful, efficient, and cost-effective manner. It is the desire of the SWE to work closely and collaboratively with the PUC and EDCs in order to develop and implement an evaluation and audit process that will produce significant and standardized impact results, at the lowest cost, so that more

funds may be allocated to customer-centric savings activities. The SWE must ensure that the evaluations are accurate and represent the actual impacts of the EE&C program with a targeted level of precision and confidence.

This Evaluation Framework outlines the expected metrics, methodologies, and guidelines for measuring program performance, and details the processes that should be used to evaluate the programs sponsored by the EDCs throughout the state. It also sets the stage for discussions among a Performance Evaluation Group of the EDCs, their evaluation contractors, the SWE Team and the PUC. These discussions will help clarify the TRM, add new prescriptive measures to the TRM, and define acceptable measurement protocols for implementing custom measures in order to mitigate risks to the EDCs. The common goal requires that kWh/yr and kW savings be clearly defined, auditable, and provide a sound engineering basis for estimating energy savings.

Appendix A. Glossary of Terms

ACCURACY: An indication of how close a value is to the true value of the quantity in question. The term also could be used in reference to a model or a set of measured data, or to describe a measuring instrument's capability.

BASELINE DATA: The measurements and facts describing equipment, facility operations, and/or conditions during the baseline period. This will include energy use or demand and parameters of facility operation that govern energy use or demand.

BENEFIT/COST RATIO (B/C RATIO): The mathematical relationship between the benefits and costs associated with the implementation of energy efficiency measures, programs, practices, or emission reductions. The benefits and costs are typically expressed in dollars.

BIAS: The extent to which a measurement or a sampling or analytic method systematically underestimates or overestimates a value.

BILLING DATA: The term billing data has multiple meanings: (1) Metered data obtained from the electric or gas meter used to bill the customer for energy used in a particular billing period. Meters used for this purpose typically conform to regulatory standards established for each customer class. (2) Data representing the bills customers receive from the energy provider and also used to describe the customer billing and payment streams associated with customer accounts. This term is used to describe both consumption and demand, and account billing and payment information.

BUILDING ENERGY SIMULATION MODEL: A building energy simulation model combines building characteristic data and weather data to calculate energy flows. While hourly models calculate energy consumption at a high frequency, non-hourly models may use simplified monthly or annual degree-day or degree-hour methods.

BUREAU OF TECHNICAL UTILITY SERVICES (BTUS): Serves as the principal technical advisory bureau to the PUC regarding fixed and transportation utility regulatory matters, as well as an adviser to the PUC on technical issues for electric, natural gas, water, wastewater, and telecommunications utilities. BTUS is formerly called the Bureau of Conservation, Economics, and Energy Planning (CEEP).

CAPACITY: The amount of electric power for which a generating unit, generating station, or other electrical apparatus is rated either by the user or manufacturer. The term also refers to the total volume of natural gas that can flow through a pipeline over a given amount of time, considering such factors as compression and pipeline size.

COEFFICIENT OF VARIATION: The sample standard deviation divided by the sample mean ($Cv = \sigma/\mu$).

CONFIDENCE: An indication of how close a value is to the true value of the quantity in question. A confidence interval (CI) is a range of values that is believed—with some stated level of confidence—to contain the true population quantity. The confidence level is the probability that the interval actually

contains the target quantity. The confidence level is fixed for a given study (typically at 90% for energy efficiency evaluations).

CONSERVATION: Steps taken to cause less energy to be used than would otherwise be the case. These steps may involve improved efficiency, avoidance of waste, and reduced consumption. Related activities include installing equipment (such as a computer to ensure efficient energy use), modifying equipment (such as making a boiler more efficient), adding insulation, and changing behavior patterns.

CONSERVATION, ECONOMICS, AND ENERGY PLANNING (CEEP): Now called the Bureau of Technical Utility Services.

CONSERVATION SERVICE PROVIDER (CSP): A person, company, partnership, corporation, association, or other entity selected by the Electric Distribution Company (EDC) and any subcontractor that is retained by an aforesaid entity to contract for and administer energy efficiency programs under Act 129.

COST-EFFECTIVENESS: An indicator of the relative performance or economic attractiveness of any energy efficiency investment or practice when compared to the costs of energy produced and delivered in the absence of such an investment. In the energy efficiency field, the term refers to the present value of the estimated benefits produced by an energy efficiency program as compared to the estimated total program costs, from the perspective of either society as a whole or of individual customers, to determine if the proposed investment or measure is desirable from a variety of perspectives, such as whether the estimated benefits exceed the estimated costs.

CUSTOMER: Any person or entity responsible for payment of an electric and/or gas bill and with an active meter serviced by a utility company.

CUSTOMER INFORMATION: Non-public information and data specific to a utility customer that the utility acquired or developed in the course of its provision of utility services.

Cv: See Coefficient of Variation.

DEEMED SAVINGS: Technical Reference Manuals (TRM) provide deemed savings values that represent approved estimates of energy and demand savings. These savings are based on a regional average for the population of participants; however, they are not savings for a particular installation.

DEMAND: The time rate of energy flow. Demand usually refers to electric power and is measured in kW (equals kWh/h) but can also refer to natural gas, usually as Btu/hr, kBtu/hr, therms/day, or ccf/day.

DEMAND RESPONSE (DR): The reduction of consumer energy use at times of peak use in order to help system reliability, reflect market conditions and pricing, or support infrastructure optimization or deferral of additional infrastructure. Demand response programs may include contractually obligated or voluntary curtailment, direct load control, and pricing strategies.

DEMAND SAVINGS: The reduction in the demand from the pre-retrofit baseline to the post-retrofit demand, once independent variables (such as weather or occupancy) have been adjusted for. This term

usually is applied to billing demand to calculate cost savings, or to peak demand for equipment sizing purposes.

DEMAND SIDE MANAGEMENT (DSM): The methods used to manage energy demand, including energy efficiency, load management, fuel substitution, and load building.

EFFICIENCY: The ratio of the useful energy delivered by a dynamic system (such as a machine, engine, or motor) to the energy supplied to it over the same period or cycle of operation. The ratio is usually determined under specific test conditions.

END-USE CATEGORY (GROUPS): Refers to a broad category of related measures. Examples of end-use categories include refrigeration, food service, HVAC, appliances, building envelope, and lighting.

END-USE SUBCATEGORY: This is a narrower grouping of measure types within an end-use category. Examples of end-use subcategories include lighting controls, CFLs, LEDs, linear fluorescents, air-source heat pump (ASHP), refrigerators/freezers, central air conditioning, and room air conditioning.

ENERGY CONSUMPTION: The amount of energy consumed in the form in which it is acquired by the user. The term excludes electrical generation and distribution losses.

ENERGY COST: The total cost of energy, including base charges, demand charges, customer charges, power factor charges, and miscellaneous charges.

ENERGY EFFICIENCY: Applied to the use of less energy to perform the same function, and programs designed to use energy more efficiently. For the purpose of this Evaluation Framework, energy efficiency programs are distinguished from DSM programs in that the latter are utility-sponsored and -financed, while the former is a broader term not limited to any particular sponsor or funding source. “Energy conservation” is a related term, but it has the connotation of “doing without in order to save energy” rather than “using less energy to perform the same function”; it is used less frequently today. Many people use these terms interchangeably.

ENERGY EFFICIENCY AND CONSERVATION PLAN AND PROGRAM (EE&C): Energy efficiency and conservation plan and program for each EDC in Pennsylvania.

ENERGY EFFICIENCY MEASURE: A set of actions and/or equipment changes that result in reduced energy use—compared to standard or existing practices—while maintaining the same or improved service levels.

ENERGY MANAGEMENT SYSTEM (EMS): A control system (often computerized) designed to regulate the energy consumption of a building by controlling the operation of energy-consuming systems, such as those for space heating, ventilation, and air conditioning (HVAC); lighting; and water heating.

ENERGY SAVINGS: The reduction in use of energy from the pre-retrofit baseline to the post-retrofit energy use, once independent variables (such as weather or occupancy) have been adjusted for.

ENGINEERING APPROACHES: Methods using engineering algorithms or models to estimate energy and/or demand use.

ENGINEERING MODEL: Engineering equations used to calculate energy usage and savings. These models usually are based on a quantitative description of physical processes that transform delivered energy into useful work, such as heating, lighting, or driving motors. In practice, these models may be reduced to simple equations in spreadsheets that calculate energy usage or savings as a function of measurable attributes of customers, facilities, or equipment (e.g., lighting use = watts × hours of use).

EVALUATION: The performance of studies and activities aimed at determining the effects of a program; any of a wide range of assessment activities associated with understanding or documenting program performance or potential performance, assessing program or program-related markets and market operations; any of a wide range of evaluative efforts including assessing program-induced changes in energy efficiency markets, levels of demand or energy savings, and program cost-effectiveness.

EVALUATION CONTRACTOR (EC): Contractor retained by an EDC to evaluate a specific EE&C program and generate ex post savings values for efficiency measures.

EX ANTE SAVINGS ESTIMATE: The savings values calculated by program Implementation Conservation Service Providers (ICSP), stored in the program tracking system and summed to estimate the gross reported impact of a program. Ex ante is taken from the Latin for "beforehand."

EX POST SAVINGS ESTIMATE: Savings estimates reported by the independent evaluator after the energy impact evaluation and the associated M&V efforts have been completed. Ex post is taken from the Latin for "from something done afterward."

FREE-DRIVER: A nonparticipant who adopted a particular efficiency measure or practice as a result of a utility program but who did not receive a financial incentive from a Pennsylvania utility.

FREE-RIDER: A program participant who would have implemented the program measure or practice in the absence of the program.

GROSS SAVINGS: The change in energy consumption and/or demand that results directly from program-related actions taken by participants in an efficiency program, regardless of why they participated.

IMPACT EVALUATION: Used to measure the program-specific induced changes in energy and/or demand usage (such kWh/yr, kW, and therms) and/or behavior attributed to energy efficiency and demand response programs.

IMPLEMENTATION CONSERVATION SERVICE PROVIDERS (ICSP): Contractor retained by an EDC to administer a specific EE&C program and generate ex ante savings values for efficiency measures.

INCENTIVES: Financial support (e.g., rebates, low-interest loans) to install energy efficiency measures. The incentives are solicited by the customer and based on the customer's billing history and/or customer-specific information.

INDEPENDENT VARIABLES: The factors that affect the energy and demand used in a building but cannot be controlled (e.g., weather, occupancy).

INTERNATIONAL PERFORMANCE MEASUREMENT AND VERIFICATION PROTOCOL (IPMVP): Defines standard terms and suggests best practice for quantifying the results of energy efficiency investments and increasing investment in energy and water efficiency, demand management, and renewable energy projects.

LOAD MANAGEMENT: Steps taken to reduce power demand at peak load times or to shift some of it to off-peak times. Load management may coincide with peak hours, peak days, or peak seasons. Load management may be pursued by persuading consumers to modify behavior or by using equipment that regulates some electric consumption. This may lead to complete elimination of electric use during the period of interest (*load shedding*) and/or to an increase in electric demand in the off-peak hours as a result of shifting electric use to that period (*load shifting*).

LOAD SHAPES: Representations such as graphs, tables, and databases that describe energy consumption rates as a function of another variable, such as time or outdoor air temperature.

MARKET EFFECT EVALUATION: The evaluation of the change in the structure/functioning of a market or the behavior of participants in a market that results from one or more program efforts. Typically, the resultant market or behavior change leads to an increase in the adoption of energy-efficient products, services, or practices.

MARKET TRANSFORMATION: A reduction in market barriers resulting from a market intervention, as evidenced by a set of market effects, that lasts after the intervention has been withdrawn, reduced, or changed.

MEASURE: An installed piece of equipment or system, or modification of equipment, systems, or operations on end-use customer facilities that reduces the total amount of electrical or gas energy and capacity that would otherwise have been needed to deliver an equivalent or improved level of end-use service.

MEASUREMENT: A procedure for assigning a number to an observed object or event.

MEASUREMENT AND VERIFICATION (M&V): Activities to determine savings for individual measures and projects. This differs from evaluation, which is intended to quantify program impacts.

METERING: The use of instrumentation to measure and record physical parameters for an energy-use equipment. In the context of energy efficiency evaluations, the purpose of metering is to accurately collect the data required to estimate the savings attributable to the implementation of energy efficiency measures.

MONITORING: Recording of parameters—such as hours of operation, flows, and temperatures—used in the calculation of the estimated energy savings for specific end uses through metering.

NET PRESENT VALUE (NPV): The value of a stream of cash flows converted to a single sum in a specific year, usually the first year of the analysis. It can also be thought of as the equivalent worth of all cash flows relative to a base point called the present.

NET SAVINGS: The total change in load that is attributable to an energy efficiency program. This change in load may include, implicitly or explicitly, the effects of free-drivers, free-riders, energy efficiency standards, changes in the level of energy service, participant and nonparticipant spillover, and other causes of changes in energy consumption or demand.

NET-TO-GROSS RATIO (NTGR): A factor representing net program savings divided by gross program savings that is applied to gross program impacts to convert them into net program load impacts.

NONPARTICIPANT: Any consumer who was eligible, but did not participate in an efficiency program in a given program year. Each evaluation plan should provide a definition of a "nonparticipant" as it applies to a specific evaluation.

NON-RESPONSE BIAS: The effect of a set of respondents refusing or choosing not to participate in research; typically larger for self-administered or mailed surveys.

PARTIAL FREE-RIDER: A program participant who would have implemented, to some degree, the program measure or practice in the absence of the program (For example: a participant who may have purchased an ENERGY STAR® appliance in the absence of the program, but because of the program bought an appliance that was more efficient).

PARTICIPANT: A consumer who received a service offered through an efficiency program, in a given program year. The term "service" is used in this definition to suggest that the service can be a wide variety of services, including financial rebates, technical assistance, product installations, training, energy efficiency information, or other services, items, or conditions. Each evaluation plan should define "participant" as it applies to the specific evaluation.

PEAK DEMAND: The maximum level of metered demand during a specified period, such as a billing month or a peak demand period.

PHASE II: EE&C programs implemented by the seven EDCs in Pennsylvania subject to the requirements of Act 129 during the program years ending on May 31 in 2014, 2015, and 2016.

PORTFOLIO: Either (a) a collection of similar programs addressing the same market (e.g., a portfolio of residential programs), technology (e.g., motor efficiency programs), or mechanisms (e.g., loan programs), or (b) the set of all programs conducted by one organization, such as a utility (and which could include programs that cover multiple markets, technologies, etc.).

PRECISION: The indication of the closeness of agreement among repeated measurements of the same physical quantity.

PROCESS EVALUATION: A systematic assessment of an energy efficiency program for the purposes of documenting program operations at the time of the examination, and identifying and recommending improvements to increase the program's efficiency or effectiveness for acquiring energy resources while maintaining high levels of participant satisfaction.

PROGRAM: A group of projects, with similar characteristics and installed in similar applications. Examples could include a utility program to install energy-efficient lighting in commercial buildings, a developer's program to build a subdivision of homes that have photovoltaic systems, or a state residential energy efficiency code program.

PROGRAM EVALUATION GROUP (PEG): Created by the PUC to, among other things, provide guidance to the SWE in clarifying energy savings measurement protocols and plans by recommending improvements to the existing TRM and other aspects of the EE&C program.

PROGRAM YEAR: For Act 129, begins on June 1 and ends on May 31 of the following calendar year; impacts are reported annually, but program years are not mapped to a calendar year.

PROJECT: An activity or course of action involving one or multiple energy efficiency measures, at a single facility or site.

REGRESSION ANALYSIS: Analysis of the relationship between a dependent variable (response variable) to specified independent variables (explanatory variables). The mathematical model of their relationship is the "regression equation."

RELIABILITY: Refers to the likelihood that the observations can be replicated.

REPORTING PERIOD: The time following implementation of an energy efficiency activity during which savings are to be determined.

RETROFIT ISOLATION: The savings measurement approach defined in IPMVP Options A and B, and ASHRAE Guideline 14, that determines energy or demand savings through the use of meters to isolate the energy flows for the system(s) under consideration.

RIGOR: The level of expected confidence and precision. Greater levels of rigor increase confidence that the results of the evaluation are both accurate and precise.

SIMPLE ENGINEERING MODEL (SEM): A category of statistical analysis models that incorporate the engineering estimate of savings as a dependent variable.

SPILLOVER: Reductions in energy consumption and/or demand caused by the presence of the energy efficiency program, beyond the program-related gross savings of the participants. There can be participant and/or nonparticipant spillover.

STIPULATED VALUES: An energy savings estimate per unit, or a parameter within the algorithm designed to estimate energy impacts that is meant to characterize the average or expected value within the population.

STATEWIDE EVALUATOR (SWE): The independent consultant under contract to the PUC to complete a comprehensive evaluation of the Phase II (program years ending in 2014, 2015, and 2016) EE&C programs implemented by the seven EDCs in Pennsylvania subject to the requirements of Act 129. GDS Associates was the SWE for Phase I and is the SWE for Phase II.

STATEWIDE EVALUATION TEAM (SWE TEAM): The team, led by the Statewide Evaluator, GDS Associates, that is conducting the evaluations of the Phase II Act 129 programs. Team members are GDS Associates, Inc., Nexant, Inc., Research Into Action, Inc., and Apex Analytics, LLC.

TECHNICAL REFERENCE MANUAL (TRM): A resource document that includes information used in program planning and reporting of energy efficiency programs. It can include savings values for measures, engineering algorithms to calculate savings, impact factors to be applied to calculated savings (e.g., net-to-gross ratio values), source documentation, specified assumptions, and other relevant material to support the calculation of measure and program savings—and the application of such values and algorithms in appropriate applications.

TECHNICAL WORKING GROUP (TWG): Chaired by PUC staff and comprised of representatives from the EDCs, the SWE, and other interested parties to encourage discussions of the technical issues related to the EM&V of savings programs to be implemented pursuant to Act 129.

TIME-OF-USE (TOU): Electricity prices that vary depending on the time periods in which the energy is consumed. In a time-of- use rate structure, higher prices are charged during utility peak-load times. Such rates can provide an incentive for consumers to curb power use during peak times.

UNCERTAINTY: The range or interval of doubt surrounding a measured or calculated value within which the true value is expected to fall within some degree of confidence.

UNIFORM METHODS PROJECT SAMPLING PROTOCOLS (UMPSP): Project of the US Department of Energy to develop methods for determining energy efficiency for specific measures through collaboration with energy efficiency program administrators, stakeholders, and EM&V consultants—including the firms that perform up to 70% of the energy efficiency evaluations in the United States. The goal is to strengthen the credibility of energy efficiency programs by improving EM&V, increasing the consistency and transparency of how energy savings are determined.

VALUE OF INFORMATION (VOI): A balance between the level of detail (rigor) and the level of effort required (cost) in an impact evaluation.

Appendix B. Commercial and Industrial Measures EM&V Data Requirements

The following sections outline the EM&V data requirements by C&I program type.

B.1. Existing Facility Lighting Efficiency Programs

Example Measures: Prescriptive lighting, fluorescent lighting (Super T-8), and custom lighting

The recommendations for the EM&V plan described here have three purposes:

- To create an audit trail for the SWE, thereby enabling inspection of facilities to accurately account for hours of use of the installed fixtures,
- To determine and have a record of the pre-existing condition as may be required, and
- To enable the EDCs and the SWE to determine the actual savings.

Lighting contractors and ICSPs who are responsible for developing and implementing lighting projects must provide lighting audit information, including pre- and post-fixture types and quantities by area and an estimation of hours of use for each fixture or fixture group. This information should be included in the application process and confirmed by the program implementer. The EDC's evaluator, audited by the SWE, subsequently will sample these project files and conduct post-installation site inspections. A site-specific verification of the lighting hours of use (logging, facility interviews) for individual projects sampled by the EDC evaluator is required for projects that exceed size thresholds established in the TRM or come from facility types with high variability. As discussed in Section 3.3, the EDC evaluator may use metering studies performed by the ICSP to reduce redundancy in data collection.

ICSPs should use the TRM Appendix C calculator or create a standardized lighting audit and application template based on the TRM and TRM appendices to document the required information. This will help maintain the uniformity and organization of the documentation throughout each project. It should be the responsibility of the ICSPs to record the necessary information, and to ensure that it is complete, even when that information is provided by contractors (trade allies) or the customer. Below is a summary of data that should be collected by the ICSP and verified by the EDC evaluator (for projects in the evaluation sample).

Table B-1: C&I Lighting Program Information to Be Collected and Verified

Site Information	Required Documentation
Name of customer Address Phone number Standard building type Utility and account number (electric)	Statement on application form
Existing Equipment Information	
Lighting audit pre-condition Lamp ballast wattage from approved table When existing equipment was installed (year) or if this is a new installation	Lighting audit form (TRM Appendix C), application stage
Line item quantity	
Line item hours of use	
New Equipment Information	
Lighting audit post condition Lamp ballast wattage from approved table	Lighting audit (TRM Appendix C), as built
Line item quantity	ICSP inspection (TRM Appendix C), acceptance of as-built
Line item hours of use	ICSP inspection (TRM Appendix C), acceptance of as-built
Equipment and installation costs	

The source for the hours of use is dependent upon the level of rigor associated with the project. For example, a low-impact project would use a TRM deemed value based on the building type while a medium- or large-impact project would necessitate site-specific measuring. To avoid redundancy in data collection, EDC evaluators may use the ICSP's metering results in lieu of conducting additional, independent metering studies.

In addition to the documentation listed in Table B-1, ICSPs and evaluators are encouraged to obtain pre-installation and post-installation walk-through inspection notes from the applicant. Often, the pre-installation inspection is the basis for the application. For some of the larger lighting projects, the customer and the lighting contractor or the ICSPs may conduct a post-installation inspection and take detailed notes. Access to these documents will increase the accuracy of ex post savings estimates.

B.2. High-Efficiency Electric Motors

Example Measures: High-efficiency electric motors purchased through participating suppliers that replace standard-efficiency electric motors.

This measure covers new, general-purpose, three-phase induction motors that meet the following criteria:

- NEMA Design A or B
- From 1 to 200 rated horsepower

- Totally enclosed fan-cooled (TEFC) or open drip proof (ODP)
- Perform better than the NEMA Premium efficiency standard

Program managers should ensure that the motor end user or motor installation contractor records all information on the new motor nameplate, and the existing motor if applicable, and collect a copy of the purchase order or invoice for the new motor. The nameplate(s) information will be required to accurately assess the gross energy and demand savings. It may be necessary to measure load versus nameplate to determine load factor as described in the TRM. This measurement should be conducted by the EDC evaluation contractor if ICSP did not perform previous logging measurements to determine load factor.

ICSPs should create a standard format to document the required information. This will help maintain the uniformity and organization of the documentation. It should be the responsibility of the ICSPs to ensure that contractors record the necessary information and that it is complete. Below is a summary of data that should be collected by the ICSP and verified by the EDC evaluation contractor.

Table B-2: C&I Motor Program Information to Be Collected and Verified

Site Information	Required Documentation
Name of customer Address Phone number Standard building type Electric utility and account number	Statements on application form
Existing Motor Information	
Motor nameplate information, including manufacturer, model number, rated horsepower, enclosure type, speed, NEMA nominal efficiency Description of application Estimated annual operating hours Constant or variable speed When existing equipment was installed (year) if this is a new installation	Statements on application form (Appendix D) Photograph "Run Load Factor"
New Motor Information	
Motor nameplate information, including manufacturer, model number, rated horsepower, enclosure type, speed, and NEMA nominal efficiency Constant or variable speed	Statements on application form (Appendix D) Vendor Specification Sheet "Run Load Factor"
Equipment and Installation Costs	

Inspections should be performed on a sample of installations.

B.3. HVAC Systems

Example Measures: Electric chillers, unitary HVAC/split systems, air-to-air heat pump systems, water-source heat pumps, ground-source heat pumps, and packaged terminal systems

In certain commercial HVAC projects, the verification of savings using Equivalent Full Load Hours (EFLH) may not appropriately characterize the operation and savings in practice. For example, chillers used in process cooling applications such as ice cream manufacturing do not share the same EFLH as chillers used in more typical comfort-cooling scenarios. Therefore, enhanced rigor must be used to verify the hours of use in cases where the equipment deviates from the applications defined in the TRM. Per the same logic, calculating demand savings for measures that deviate from the TRM's defined applications also requires enhanced rigor to determine a load shape or coincidence factor. For example, a chiller that is used only at night because it makes ice for the HVAC's cooling system to melt during the day would deviate from the TRM's intended applications.

ICSPs should create a standard format to document the required information. This information is necessary to provide accountability and more accurately assess the program's gross energy and demand savings. This will help maintain the uniformity and organization of the documentation throughout each project. It should be the responsibility of the ICSPs to ensure that contractors record the necessary information and that it is complete. Below is a summary of data that should be collected by the program ICSP and verified by the EDC evaluation contractor (for projects in the evaluation sample).

Table B-3: C&I HVAC System Program Information to Be Collected and Verified

Site Information	Required Documentation
Name of customer Address Phone number Standard building type Electric utility and account number	Statements on application form
Existing Equipment Information	
Equipment number Type Manufacturer and model number Efficiency: EER, SEER, COP, kW/ton Cooling capacity Heating capacity (heat pumps) Quantity When existing equipment was installed (year) or if this is a new installation Estimated annual operating hours Year of manufacture Fuel switching?	Statements on application form Photograph

Site Information	Required Documentation
New Equipment Information	
Type Manufacturer and model number Efficiency: EER, SEER, COP, kW/ton Cooling capacity Heating capacity (heat pumps) Quantity Estimated annual operating hours Year of manufacture	Statements on application form Vendor Specification Sheet
Equipment and installation costs	

As mentioned in previous sections of this appendix and the gross impact evaluation section of this Evaluation Framework (Section 3.3), the EDC evaluation contractor may use logging or metering hours of use (for projects such as fans or pumps in the case of HVAC systems) to determine energy savings, but the ICSP must perform these calculations.

B.4. Variable Frequency Drive Programs

Measures: Variable frequency drive programs

The current TRM provides a measurement protocol for variable frequency drives (VFDs). Central to calculating savings using the fan laws is the load shape of the flow rate of the fluid before and after the installation of the VFD. This defines the difference in work required and therefore is the basis of the energy savings.

The data required to verify savings and the steps required to implement the EDC evaluation contractor's EM&V plan will, to some degree, depend on the PEG's work and any changes to the TRM the PUC approves. However the recommendations for the EM&V plan described here serve the additional purpose of creating an audit trail for the SWE, enabling inspection of facilities to accurately account for load on the VFD as installed and to determine and have a record of the pre-existing condition, if required.

Mechanical contractors and program ICSPs who are responsible for developing and implementing VFD projects are responsible for providing enough information to estimate the pre- and post-condition energy use and fluid flow rates over time. For example, this usually is done by either the ICSP or EDC, depending on the program logic Data collection occurs before installation by spot measurement of kW on a constant-speed motor to be replaced by the VFD, and subsequent modeling of the expected flow rates, taking into account the proposed control configuration and weather bin data. This information should be part of the application process and be conducted and/or confirmed by the ICSP as appropriate.

In certain custom measure applications, it may be necessary to conduct pre- or post-installation metering of the load over a representative time period to verify savings when loading is variable. The EDC evaluation contractor, audited by the SWE, subsequently will sample these project files and conduct

post-installation site inspections. The information is necessary to provide accountability and to more accurately assess the programs' gross energy and demand savings.

ICSPs should create a standardized file and application template to document the required information. This will help maintain the uniformity and organization of the documentation throughout each project. ICSPs should ensure that contractors record the necessary information and that it is complete. Below is a summary of data that the ICSP should collect, and the EDC evaluation contractor should verify (for projects in the evaluation sample).

Table B-4: C&I Variable Frequency Drive Program Information to Be Collected and Verified

Site Information	Required Documentation
Name of customer Address Phone number Standard building type Utility and account number (electric)	Statement on application form
Existing Equipment Information	
Existing motor nameplate data Pre kW, instantaneous or Pre kW over representative loading period When existing equipment was installed (year) or if this is a new installation	Motor inventory form (Appendix D), application stage Logged kW or proxy data Modeled savings
Line item quantity	Motor inventory form (Appendix D), application stage
Line item hours of use (not FLH or EFLH)	Motor inventory form (Appendix D), application stage
New Equipment Information	
Proposed motor nameplate data Drive nameplate data and efficiency Post kW over representative loading period	Motor inventory form (Appendix D), application stage Logged kW or proxy data
Line item quantity	Motor inventory form (Appendix D), application stage
Line item hours of use (not FLH or EFLH)	Motor inventory form (Appendix D), application stage
Equipment and installation costs	

Appendix C. Residential Measure EM&V Data Requirements

The following sections outline the EM&V data requirements by residential program type.

C.1. HVAC Efficiency Programs

Example Measures: Central air conditioner (A/C) and air-source heat pump (ASHP), ground-source heat pump (GSHP), GSHP de-superheater, and furnace high-efficiency fan

HVAC equipment will be provided and installed by participating installation contractors and dealers. The utilities should require the installation contractors to collect and record information about each project that receives an incentive through the program. The information is necessary to provide accountability and to support accurate assessments of the programs' gross energy and demand savings.

ICSPs should create a standardized format to document the required information. This will help maintain the uniformity and organization of the documentation throughout each project. ICSPs should ensure that contractors record the necessary information and that it is complete. Below is a summary of data that should be collected by the program ICSP and verified by the EDC evaluator (for projects in the evaluation sample).

Table C-1: Residential HVAC Efficiency Program Information to Be Collected and Verified

Site Information	Required Documentation
Name of customer Address Phone number Utility and account number (electric)	Statement on application form
Existing Equipment Information	
Manufacturer and model Efficiency: EER, SEER, COP, HSPF Capacity Original fuel type	Nameplate information on application form (applicable to early retirement scenario only). Photograph of nameplate is optional and acceptable.
Quantity	Statement on application form
Location/EFLH	Address
Type, quantity, nominal efficiency, and set points of other heating and cooling systems When existing equipment was installed (year) or if this is a new installation	Statement on application form

Site Information	Required Documentation
New Equipment Information	
Manufacturer and model Efficiency: SEER and EER (A/C), HSPF (ASHP), or EER and COP (GSHP) Capacity	1. Invoice, and 2. Application form or cut sheets
Quantity	Invoice
Location/EFLH	Address
Type, quantity, nominal efficiency, and set points of other heating and cooling systems	Statement on application form
For central air conditioners and air-source heat pumps, specify whether the measure is high-efficiency equipment, proper sizing, QIV, maintenance, or duct sealing	Statement on application form
Equipment and installation costs	

Inspections should be performed on a random sample of participating customers.

C.2. New Construction Programs

Example Measures: Insulation upgrades, efficient windows, air sealing, efficient HVAC equipment, duct sealing, efficient lighting, ENERGY STAR appliances

New homes will be constructed by qualifying builders and contractors. The utilities should require the contractors to collect and record information about the new homes that receive incentives. The information is necessary to provide accountability and to assess the programs' gross energy and demand savings more accurately.

ICSPs should create a standard format to document the required information. This will help maintain the uniformity and organization of the documentation throughout each project. ICSPs should ensure that contractors record the necessary information and that it is complete. Below is a summary of data that should be collected by the program ICSP and verified by the EDC evaluator (for projects in the evaluation sample).

Table C-2: Residential New Construction Program Information to Be Collected and Verified

Site Information	Required Documentation[108]
Name of customer Address Phone number Utility and account number (electric)	Statements on application form
Area (square feet)	Statement on application form
Dimensions and construction details	Construction drawings
Insulation thickness and type	Statements on application form
HVAC	Inventory indicating quantities, manufacturers, model numbers, and efficiency ratings for the following equipment, if applicable: furnace, boiler, combination water heater, air-source heat pump, geothermal heat pump, packaged terminal air conditioning (PTAC), packaged terminal heat pump (PTHP), central air conditioning, window air conditioner, thermostat, active solar system
Domestic hot water	Statements on application form indicating heater type (electric or gas), size (gallons), and tank insulation type and thickness, if applicable
Lighting	Inventory indicating quantities, manufacturers, model numbers, and wattages
Appliances	Inventory indicating quantities, manufacturers, model numbers, and efficiency ratings
Air infiltration	Blower door test results
Equipment and installation costs	

Any information that is available on home energy ratings or other performance ratings conducted through the program also should be collected and stored in the EDC's data tracking and reporting system.

Inspections should be performed on a random sample of new homes.

[108] Lighting and appliance data can be challenging to collect for every home. An acceptable alternative is to gather this information for the data sample and extrapolate the findings to the population.

C.3. ENERGY STAR Appliance Programs

Example Measures: Refrigerator, freezer, dehumidifier, room air conditioner, dishwasher, and clothes washer

ENERGY STAR appliances will be provided through participating retailers. The utilities should maintain records with information about each appliance that receives an incentive through the program. The information is necessary to provide accountability and to assess the programs' gross energy and demand savings more accurately.

Retailers' invoices should include the information listed below. ICSPs should ensure that retailers record the necessary information and that it is complete. Below is a summary of data that should be collected by the program ICSP and verified by the EDC evaluator (for projects in the evaluation sample).

Table C-3: Residential ENERGY STAR Appliance Program Information to Be Collected and Verified

Retailer and Equipment Information	Required Documentation
Name Address Phone number Utility providing incentive Manufacturer and model Is the model ENERGY STAR-rated? Type (see below) Quantity When existing equipment was manufactured (year) or if this is a new installation Existing equipment fuel type (if applicable) Equipment costs and installation costs (if known)	Invoices/application

Refrigerator types: Manual defrost, partial automatic defrost, top-mount freezer without door ice, side-mount freezer without door ice, bottom-mount freezer without door ice, top-mount freezer with door ice, and side-mount freezer with door ice

Freezer types: Upright with manual defrosts, upright with automatic defrosts, chest freezer, compact upright with manual defrosts, compact upright with automatic defrosts, and compact chest freezer

Dehumidifier types: 1-25 pints/day, 25-35 pints/day, 35-45 pints/day, 45-54 pints/day, 54-75 pints/day, and 75-185 pints/day

Room air conditioner – nearest location: Allentown, Erie, Harrisburg, Philadelphia, Pittsburgh, Scranton, and Williamsport

Dishwasher: Gas or electric water heater

Clothes washer: Gas or electric water heater

C.4. Appliance Retirement Programs

Example Measures: Refrigerator, freezer, and room air conditioner

The program ICSP will provide appliance turn-in incentives. The utilities should require the ICSP to maintain invoices and record information about each appliance that receives an incentive through the program. The information is necessary to provide accountability and to assess the programs' gross energy and demand savings more accurately.

ICSPs should collect the information listed below. ICSPs should record the necessary information and ensure that it is complete. Below is a summary of data the ICSPs should collect, and the EDC evaluation contractor should verify (for projects in the evaluation sample).

Table C-4: Appliance Retirement Program Information to Be Collected and Verified

Customer and Equipment Information	Required Documentation
Name Address Phone number Utility and account number (electric) Is it an ENERGY STAR-rated appliance? Quantity Vendor information Disposal site	Statement on application form
Make and model Capacity Age of the retired unit Was the unit replaced?	Nameplate information on application form

C.5. ENERGY STAR Lighting Programs

Example Measures: High-efficiency bulbs and fixtures

High-efficiency lighting measures can be distributed in a number of ways. The ENERGY STAR lighting algorithm in the 2014 Pennsylvania TRM assumes that 97% of bulbs ultimately are installed in residential sockets.[109] The TRM algorithm also does not assume any leakage in or leakage out of an EDC service territory. At the basic level of rigor, EDC evaluation contractors will apply these TRM assumptions when developing ex post savings estimates for ENERGY STAR lighting programs.

For an enhanced level of rigor, the SWE encourages EDC evaluation contractors to conduct interviews or store exit surveys to measure cross-sector sales.[110] When respondents indicate that the rebated bulbs will be installed in a nonresidential application, survey batteries should query respondents about the type of facility in which the bulbs will be installed. For the proportion of bulbs determined to be installed

[109] This estimate was updated from the 2013 TRM value of 84%.
[110] Cross-sector sales are defined as incented bulbs that are installed in non-residential applications.

in nonresidential applications, EDC evaluation contractors should assume the TRM hours of use and coincidence factor for the stated facility type.

Leakage of rebated bulbs out of an EDC service territory can be captured by exit surveys, but it is far more challenging to measure leakage into an EDC service territory effectively. Because of the data collection challenges and the likelihood that most leakage in Pennsylvania would occur back and forth between EDC service territories subject to Act 129, the SWE recommends that EDC evaluation contractors assume that leakage in and leakage out offset at both the enhanced and basic levels of rigor. If the SWE and BTUS decide that it is important to assess the leakage into and out of EDC service territories, a statewide study may be proposed and vetted with the EDCs and their evaluation contractors. Similarly, since Pennsylvania and its neighboring states have utility incentives for energy efficiency, EDCs should assume leakage between states offset. If the SWE and BTUS decide that it is important to assess the leakage into and out of Pennsylvania, a statewide study may be proposed and vetted with the EDCs and their evaluation contractors.

ICSPs should collect the information listed in Table C-5 at least every quarter. ICSPs should record the necessary information and ensure that it is complete. Below is a summary of data that the program ICSP should collect, and that the EDC evaluation contractor should verify (for projects in the evaluation sample).

Table C-5: Residential ENERGY STAR Lighting Program Information to Be Collected and Verified

Retailer and Equipment Information	Required Documentation
Name Address Phone number Utility providing incentive Manufacturer and model Type (see categories below) Wattage Quantity sold/distributed Equipment costs and installation costs (if known)	Invoices

Categories: CFL bulbs, torchières, indoor fixtures, outdoor fixtures, LED bulbs, or ceiling fans with ENERGY STAR light fixture

C.6. ENERGY STAR Windows Programs

Measure: ENERGY STAR-qualified windows

ENERGY STAR-qualified window incentives will be provided to vendors or installers. The EDCs should require the vendors or installers to collect and record information about the windows that receive incentives. The information is necessary to provide accountability and to assess the programs' gross energy and demand savings more accurately.

Program ICSPs should create a standard format to document the required information. This will help maintain the uniformity and organization of the documentation. ICSPs should record the necessary

information and ensure that it is complete. Table C-6 and Table C-7 summarize the data that the program ICSP should collect, and that the EDC evaluation contractor should verify (for projects in the evaluation sample).

Table C-6: Residential Window Program Information to Be Collected – if Vendor Receives Incentive

Vendor Information	Required Documentation
Vendor name Address Phone number	Statement on application form
Window Information	
Manufacturer, model, and type U-Factor Solar Heat Gain Coefficient (SHGC) Are the windows ENERGY STAR-rated? Number of panes in the new windows	Statement on application form
Area (square feet)	Statement on application form
Date sold	Statement on application form
Equipment and installation costs	

Table C-7: Residential Window Program Information to Be Collected – if Installer Receives Incentive

Site Information	Required Documentation
Name of customer Address Phone number Utility and account number (electric)	Statement on application form
Heating system type (heat pump, electric resistance, other)	Statement on application form
Cooling system type	Statement on application form
Existing Window Information	
Manufacturer, model, and type	Statement on application form
Area (square feet)	Statement on application form
Retrofit Window Information	
Manufacturer, model, and type U-Factor Solar Heat Gain Coefficient (SHGC) Are the new windows ENERGY STAR-rated? Number of panes in the new windows	Statement on application form
Area (square feet)	Statement on application form
Purchase and installation dates	Statement on application form
Equipment and installation costs	

Inspections should be performed on a sample of installations.

Appendix D. Mass Market Protocols

The following memos are provided in the following pages:

- CMP_Behavioral Based_OPower_Revised 11-08-2010 Approved
- CMP_Low Income WRAP_Revised_24Nov2010 Approved
- PA LEEP Component 1 Custom Protocol 2010 0930
- Residential LED Lighting Custom Protocol 2010

Custom Measure M&V Protocol
PPL Electric's OPOWER Energy Education Program

1. Measure Description

PPL Electric's OPOWER program aims to generate greater awareness about energy use and how to manage it better through energy efficiency education. The program provides information and education to utility customers about their energy use and demand and low-cost measures, practices or behaviors to reduce them. This program falls under the category of "behavior-based" energy efficiency measures and may include one or more of the following:

- Tracking and analysis of customers' energy use and demand over time.
- Benchmarking a customer's usage against comparable customers in the same geographical area and reporting the results.
- Tips, emphasizing the importance of peak load reduction during the peak load season and ways to shift energy use to off-peak periods.
- General conservation tips such as turning down the thermostat, turning off lights, shortening showers, etc.
- Encouraging the use of low-cost energy-efficiency measures such as replacing incandescent lights with CFLs, installing weather stripping and hot water saving measures, and using smart power strips.[1]
- Delivering information about tools and resources available through the utility's web site.
- Providing information about promotions of other utility residential programs.

Information about energy consumption and demand and savings tips is intended to increase awareness and understanding of energy use and savings opportunities. It is hypothesized that after customers become educated, they will be more likely to adopt practices and measures leading to more efficient energy use in their homes.

Information about energy use and savings can be delivered to customers through several different media and at different frequencies.

- Customer may obtain real time energy demand information on a continuous basis with a display placed in the customer's home.
- Customers may obtain information on demand by logging into a utility or program vendor website.

[1] Tips about energy savings measures may be tailored to the customer and generated automatically with an analytics engine that analyzes consumption history, housing characteristics, and customer demographics.

- Customers may receive reports about energy use by U.S. mail at monthly, bi-monthly, quarterly, or semi-annual frequencies.

2. M & V Methodology

This M & V protocol is based on Option C-Whole Facility of the IPMVP[2] (Billing Regression Analysis as per Section 3.3.3.3.6.2.3 of the SWE Audit Plan) for annual energy savings. Billing analysis—using data on energy use in participating and non-participating homes before and after the treatment—is the most appropriate method to use to account for savings attributable to this program. This protocol focuses on: (1) the impacts to estimate; (2) the issues that should be considered; and (3) the statistics to report.

A regression analysis of billing data will result in an estimate of the energy savings impact of education and behavioral programs in the population of customers eligible to receive the information. The estimate of the program impact will be unbiased if the model is properly specified and the error term of the model has an expected value of zero conditional on the observed covariates. Necessary conditions for obtaining an unbiased estimate of the program impact are:

- The evaluation is set up as a randomized control trial (RCT) with treatment and control groups;
- The regression analysis controls for the effects of weather on consumption; and
- The regression analysis uses consumption data from before and after the treatment.

Should the EM&V CSP determine that one or more of these conditions is not applicable to the measurement of program impacts, the EM&V CSP will include a detailed description of why the condition or conditions were not met during the calculation of verified savings.

A RCT is an experimental program design in which subjects (i.e., utility customers eligible for the program) are randomly assigned to treatment and control groups.[3] Random assignment of a treatment to customers is a key element in establishing a causal connection between the treatment and measured outcomes.[4] It facilitates identification and estimation of program impacts because it eliminates non-random, self-selection into the program. In a RCT, the treatment is independent of other observable and unobservable characteristics affecting energy use and demand.

Weather is a significant driver of energy consumption, and seasonal or annual fluctuations in temperature could mask or be mistaken for program impacts in the regression analysis. The billing regression analysis should therefore control for weather.

---

[2] *International Performance Measurement & Verification Protocol (IPMVP); Concepts and Options for Determining Energy and Water Savings: Volume 1*. Prepared by Efficiency Valuation Organization, www.evo-world.org. September 2009. EVO 10000 – 1:2009..

[3] In a RCT, the treatment group would be customers that receive information about energy usage. The control group would be customers that do not receive information (the treatment).

[4] See Levitt and List (2008).

Finally, factors affecting energy consumption unrelated to the program such as macroeconomic-driven changes in income and employment could bias estimates of program impacts. With the use of consumption data for program participants and non-participants before and after the beginning of the program, it will be possible to implement a difference-in-differences regression model to control for such factors.

3. Ex Ante Calculations

Ex ante savings are typically calculated by the program's third party implementer. These savings will be calculated based on data from OPOWER programs with verified estimates of program impacts. The program's third party implementer will report ex ante savings, along with any references and assumptions used in their calculation to the SWE prior to program implementation.

4. M & V Activities

The M&V activities will verify the reported energy savings impact of OPOWER's energy education and behavioral program. The main evaluation activities of this M & V protocol are:

- Sample selection
- Estimation of program impacts and analysis
- Reporting

Sample Selection

Sampling will follow the Sampling and Uncertainty Protocol of Section 3.3.4 of the Audit Plan[5] to achieve a 90/10 level of confidence and relative precision. Sample selection involves drawing a representative and sufficiently large sample of participants and non-participants for estimation of the program impacts. The EM&V CSP will select a random sample of participants from the treatment population and a random sample of non-participants for analysis. In an RCT, the EM&V CSP should select all customers in the program's treatment and control groups, if practicable. In selecting a non-participant control group, the EM&V CSP must choose carefully, as non-participant outcomes should represent accurately the counterfactual outcomes of participants without the treatment. Another way of saying this is that the covariate distributions for the treatment and control groups should be similar.

The selection of the estimation sample will be closely integrated with the design of the program. A randomized control trial is the preferred approach, as it facilitates estimation and ensures a causal interpretation of the estimated impacts. Implementing a RCT first requires carefully defining and selecting a population of customers eligible to participate in the program based on program objectives, utility program delivery capabilities, and program budget; selecting a random sample of customers from the population of eligible customers; and

[5] *Audit Plan and Evaluation Framework for Pennsylvania Act 129 Energy Efficiency and Conservation Programs*. 2009. Prepared by the Statewide Evaluation Team: GDS Associates, Inc., Nexant and Mondre Energy. Contracted under the Pennsylvania Public Utility Commission's RFP 2009-1 for the Statewide Evaluator. December 1, 2009.

randomly assigning eligible customers in the random sample to treatment and control groups. The estimated impact of the program is the effect of the program in the population of customers eligible for the program. The EM&V CSP should be cautious in attempting to extrapolate the results to other customer populations. In addition, the EM&V CSP should be aware of potential exposure of customers in the control group to the treatment that would undermine the validity of the trial.

The estimation sample must be sufficiently large to estimate program impacts precisely. Previous evaluations of education and behavioral utility programs have typically found savings that are between one and three percent of consumption.[6] The exact size of the estimation sample will depend on the expected program impact, the desired levels of statistical confidence and power, relative precision, the coefficient of variation of energy consumption or demand, and the population size, and will be calculated in accordance with Section 3.3.4.1 of the SWE Audit Plan.[7]

Estimation of program impacts and analysis

In this analysis, the baseline period is defined as the last 12 months of usage prior to treatment, i.e., prior to receipt of program materials. The post period is defined as the first 12 months of usage in the treatment period, i.e., the 12 months during which the customer receives program materials.

A difference-in-differences regression model will be utilized to estimate the program's impact. As the name implies, this model estimates the difference between treatment and control group customers in their differences in consumption between before and after the program. This model is valid for experimental designs where customers in the estimation sample were randomly assigned to the treatment and control groups.

A parsimonious, difference-in-differences specification for estimating the impact of the program on energy consumption is:

$$ADC_{it} = \alpha_i + \beta_1 POST_{it} + \beta_2 PARTPOST_{it} + \mu_{my} + \varepsilon_{it}$$

where, for each customer i and period t (corresponding to a month and year) :

- ADC_{it} is the average daily kWh consumption.
- $POST_{it}$ is a dummy variable that is 1 in the treatment period and 0 in the pre-treatment period.
- $PARTPOST_{it}$ is a dummy variable that is 1 in the treatment period for participants and 0 otherwise.
- ε_{it} is the random error term indicating the impact of unobservable factors on consumption.

[6] See Alcott (2009), Ayres, Raseman, and Shih (2009), Connexus Energy (2010), Costa and Kahn (2010), Kihm, Koski, and Mendyk (2010), Sipe and Castor (2009).

[7] See List et al (2010) for guidance about choosing appropriately sized samples in experimental research designs.

- μ_{my} is a month-by-year fixed effect, which represents the average impact on consumption in month m in year y. It is intended to capture weather and other monthly effects on consumption.[8]
- α_i is the intercept for home i. This corresponds to the non-weather sensitive average daily base energy use in the home.
- β_1 is the average daily change in consumption between the pre-treatment and treatment periods that is unrelated to program impacts.
- β_2 is the average daily participant savings in the population of eligible customers.

Eligible customers are randomly assigned to the treatment and control groups; thus, the error term ε_{it} will be uncorrelated with PARTPOST and other observable variables and ordinary least squares (OLS) will result in an unbiased estimate of the average daily program impact.[9] The estimated coefficient on the PARTPOST variable represents the average treatment effect (the kWh savings impact) of the program in the population of eligible customers *controlling for changes in participant usage unrelated to the program*.[10] The change in consumption unrelated to the program is given by the coefficient of the POST variable.

---

[8] If there is significant variation in weather between customers in the estimation sample, it may be desirable to drop the month-by-year fixed effects and include polynomials in heating degrees and cooling degrees as explanatory variables.

[9] A test of the random assignment of treatments would be to compare the means of observable characteristics of treatment and control groups or to regress a dummy variable for program participant on a observable variables. If assignment of treatments is random, the means should be statistically indistinguishable and the coefficients on the observable variables in the regression should be statistically insignificant. There will be correlation in a customer's consumption over time. The estimated standard errors should be corrected for this correlation. This type of statistical test will be conducted to confirm the coefficients on observable variables between treatment and control populations are indistinguishable. This type of test will also be conducted to confirm the coefficients on observable variables between the treatment population and treatment sample, and control population and control sample, are indistinguishable. The EDC will provide the SWE team with results of these tests.

[10] To see that β_2 represents average daily savings after controlling for changes in participant usage unrelated to the program, observe

$E[ADC_{it} \mid Participant_i=1, Post_i=1] = \alpha_i + \beta_1 + \beta_2 + \mu_{my}$

$E[ADC_{it} \mid Participant_i=1, Post_i=0] = \alpha_i + \mu_{my}$

where E is the expectation operator and Participant is a variable for whether the customer is a participant. Now for non-participants,

$E[ADC_{it} \mid Non\text{-}participant_i=0, Post_i=1] = \alpha_i + \beta_1 + \mu_{my}$

$E[ADC_{it} \mid Non\text{-}participant_i=0, Post_i=0] = \alpha_i + \mu_{my}$

Taking differences between pre-treatment and treatment periods for participants and non-participants yields,

$E[ADC_{it} \mid Participant_i=1, Post_i=1] - E[ADC_{it} \mid Participant_i=1, Post_i=0] = \beta_1 + \beta_2$

The difference-in-differences regression methodology requires monthly billing and program participation data to implement. The billing data should include at least the last 12 months of usage in the pre-treatment period (the baseline period) and the first 12 months of usage in the treatment period. The model can be estimated by OLS. The standard errors should be adjusted for correlation over time in a customer's consumption using Huber-White robust standard errors (Bertrand, Duflo, and Mullainathan, 2004).

The regression model estimates the average treatment effect of the program in the population of eligible customers but the model may be modified to test alternative hypotheses. For example, a common hypothesis about education and behavioral programs is that the impact of information may depend upon the frequency with which it is received. This hypothesis may be tested by interacting $PARTPOST_{it}$ with a variable for the frequency with which a customer receives reports.

Another hypothesis concerns the persistence of savings impacts of information-based programs. If participants install CFLs, high efficiency furnaces, and other measures, utility planners can be confident that any savings from the program will have typical persistence over the entire lifetime of the measure. However, if the program primarily changes behaviors, such changes may not last over their proposed lifetime unless they are reinforced. A 2009 report prepared for the California Institute for Energy and Environment noted the lack of any robust procedure for determining persistence of energy savings associated with behavioral programs.[11] This program's measured and verified savings will be applicable for the treatment period. Annual re-verification is required to confirm persistence of savings beyond the initial 12 month treatment period. In order to test persistence of savings during the treatment period, $PARTPOST_{it}$ will be interacted with indicator variables for number of months from the beginning of the treatment. The coefficients on the variables will represent the savings from the program at month 1, month 2, month 3, etc.

Reporting

The EM&V CSP will describe completely its approach and report a full complement of statistics that will allow readers to evaluate the reliability of the savings estimates:

1. Total observations, number of cross-sectional units (homes), and number of time periods;
2. Estimated coefficients and standard errors for the parameters of models;

The expected change in consumption between the treatment and pre-treatment period for participants is the sum of the program impact and any changes in consumption unrelated to the program.

Similarly, for non-participants:

$E[ADC_{it} \mid \text{Non-participant}_i=0, \text{Post}_i=1] - E[ADC_{it} \mid \text{Non-participant}_i=0, \text{Post}_i=0] = \beta_1$

Finally, the difference of the participant and non-participant differences is β_2, which is the difference-in-differences impact of the program on consumption *controlling for changes in consumption unrelated to the program*.

[11] See Skumatz (2009) pg. 11

3. Adjusted R^2;
4. Tests of the joint significance of the coefficients of the model (e.g., F test statistics).
5. Results of any tests of violations of assumptions of classical regression model. The EM&V CSP should identify and address any violations of the assumptions of the classical regression model, including serial correlation, heteroskedasticity, collinearity, and omitted variable bias.[12]

In addition, based on the regression results, the following program impact estimates should be reported:

1. Point estimate(s) of the average monthly or annual program customer impact and estimated standard errors;
2. A 95% confidence interval(s) for the average monthly or annual program customer impact;
3. Point estimate(s) of monthly or annual program savings and estimated standard errors;[13]
4. A 95% confidence interval(s) for the monthly or annual program savings.

Finally, it should be noted that one of the main challenges of evaluating information-based programs is establishing a causal link between any estimated reductions in energy use or demand and specific customer actions that led to those reductions. Information about participant behavior that could be used to substantiate regression-based savings claims may be gathered in two ways.

First, the EM&V CSP may conduct a survey or focus group of participating customers. Participants would be asked about actions taken in response to program information. A parallel survey of non-participants would provide a baseline for comparing participant and non-participant behaviors and for qualitatively establishing program impacts.

Second, the EM&V CSP may match customers in education and behavioral programs to rebate or participation lists for other utility programs. Matching will indicate rebated measures that education and behavioral program participants installed; unlike a survey, it will not, however, provide the EM&V CSP with information about behavior changes. If the program evaluation is set up as a RCT, any difference between program participants and non-participants in the installation of measures can logically be attributed to the program. The EM&V CSP can then connect differences in energy savings between customers to specific behaviors.

[12] See Greene (2007), pp. 213-222.

[13] Suppose that the estimated impact of a program on average daily electricity consumption is given by $b_2<0$. Then an estimate of average annual participant savings are equal to $-365*b_2$. Estimated annual program savings are then equal to N*estimated average annual participant savings, where N is the number of participants in the program.

5. Metering Plan

Metering is not recommended or needed to confirm impacts on energy consumption; however, because information programs are behavior based and behavior changes may not be permanent, it is recommended that the billing analysis be repeated periodically to assess the persistence of savings.

Interval metered data of a statistically valid sample is required to measure and verify program impacts on peak load reduction.

6. Calculations and Adjustments

A proper sample design and properly-specified regression model will yield an unbiased estimate of the program savings impact. However, the utility may not be able to take 100 percent of the credit for the savings because some savings may be counted and credited to other utility programs. When a customer installs a measure in response to program information and receives a rebate from the utility, the savings appear not only in the regression-based estimate of the program impact but also in the savings of the utility rebate program. In order to avoid the double counting of savings, the EM&V CSP will have to estimate how much of the education and behavioral program savings have been counted by and credited to other utility programs.

The amount of savings overlap is relatively straightforward to calculate in a RCT. For simplicity, suppose that there are an equal number of customers in the treatment and control groups and that the EM&V CSP has information about the installation of Measure A, which is promoted by the utility, for both groups. Customers in the behavioral and education program treatment and control group are assumed to receive the same treatment from the utility for the program promoting Measure A (i.e., they face the same marketing and incentives). Because customers were randomly assigned to the treatment and control groups, any difference between the groups in the installation of Measure A can be attributed to the behavioral program. Suppose the difference is Δn_A and the per unit savings are s_A. Then the amount of savings counted by the behavioral and education program and the other utility program would be $\Delta n_A * s_A$. To avoid double counting, these savings must be subtracted from either the education and behavioral program savings or the other utility program savings.

For measures promoted by utility programs *and tracked at the customer level*, accounting for overlapping savings claims is fairly straightforward. The EM&V CSP would match behavioral and education program participants and non-participants to participation data for other utility programs and calculate the savings attributable to education and behavioral program. For measures not tracked at the customer level such as upstream CFL programs, accounting for savings overlap is more complicated. The EM&V CSP will have to survey participants and non-participants about their participation in upstream programs and other utility programs not tracking participation at the customer level to estimate the extent of savings overlap. These surveys will be conducted in accordance with Section 3.3.4 of the SWE Audit Plan in order to obtain a statistically significant estimate of treatment and non-treatment group participation in upstream programs.

References

Allcott, Hunt, 2009. Social Norms and Energy Conservation. MIT Center for Energy and Environmental Policy Research working paper.

Ayres, Ian, Sophie Raseman, and Alice Shih, 2010. Evidence from Two Large Field Experiments that Peer Comparison Feedback Can Reduce Residential Energy Usage. National Bureau of Economic Research Working Paper No. 15386. Available at http://www.nber.org/papers/w15386.pdf

Bertrand, Marianne, Ester Duflo, and Sendhil Mullainathan, 2004. How Much Should We Trust Difference-in-Differences Estimates. Quarterly Journal of Economics, 119 (1), pp. 249-275.

Connexus Energy, 2010. Measurement and Verification Report of OPower Energy Efficiency Pilot Program. Report prepared by Power System Engineering.

Costa, Dora and Matthew E. Kahn (2010). Energy Conservation "Nudges" and Enviromentalist Ideology: Evidence from a Randomized Residential Electricity Field Experiment. National Bureau of Economic Research working paper 15939. Available at http://www.nber.org/papers/w15939.pdf

Greene, William, 2007. Econometric Analysis (6th edition). New York: Prentice-Hall.

Kihm, Steven, Karen Koski, and Andy Mendyk, 2010. Focus on Energy – Powercost Monitor Study: Final Report, April 16, 2010. Report prepared by Energy Center of Wisconsin on behalf of Focus on Energy. *www.focusonenergy.com/files/.../powercostmonitorstudy_finalreport.pdf*

Levitt, Steven D. and John List, 2008. Field Experiments in Economics: The Past, The Present, and the Future. National Bureau of Economic Research working paper 14356. Available at http://www.nber.org/papers/w14356.

List, John A., Sally Sadoff, and Mathis Wagner, 2010. So You Want to Run an Experiment, Now What? Some Simple Rules of Thumb for Optimal Experimental Design. National Bureau of Economic Research working paper 15701. http://www.nber.org/papers/w15701Sipe, Brien and Sarah Castor, 2009. The Net Impact of Home Energy Feedback Devices. Prepared for Energy Program Evaluation Conference, Portland, Oregon.

Skumatz, Lisa A. Lessons Learned and Next Steps in Energy Efficiency Measurement and Attribution: Energy Savings, Net to Gross, Non-Energy Benefits, and Persistence of Energy Efficiency Behavior. Prepared for the CIEE Behavior and Energy Program; November 2009.

Custom Measure Protocol
Low Income Usage Reduction Program
Winter Relief Assistance Program (WRAP)
PPL Electric

1. Measure Description

PPL Electric Utilities provides measures and services to reduce the electric bills and improve the comfort of low-income customers in conjunction with its Winter Relief Assistance Program ("WRAP")[1].

The Low-Income WRAP Program offers:

- Energy audits
- Installation of energy saving measures
- Energy education
- Minor health and safety repairs

WRAP (Winter Relief Assistance Program) is an existing PPL Electric program designed to reduce electric consumption and improve living comfort for low-income customers. Eligible customers receive a free energy audit and their home is evaluated for eligible energy saving measures. A pre-approved list of cost-effective measures is used along with other criteria to determine if appliances and other larger equipment can be cost-effectively replaced. Implementer agencies have either in-house contractors and/or they contract-out installation of the energy saving measures. Outdated and inefficient equipment in customer homes is replaced with energy-efficient equipment. Energy education is also offered through the Low-Income WRAP Program to encourage customers to conserve energy. Minor health and safety repairs are made so that agencies implementing the program do not have to deny services altogether if a structure can be treated with minor repair. This minimizes "walk-away" situations.

The Act 129 WRAP program targets low-income residential customers who are at or below 150% of the federal poverty level; low income customers 200% of federal poverty level are eligible for the existing WRAP program. The Act 129 WRAP program is available to customers in existing single-family housing and in existing multi-family housing, where 50% or more of the tenants are low-income qualified. The Act 129 WRAP program seeks to reach new participants, PPL Electric customers who received WRAP assistance in the past and may be in need of further WRAP services, as well as customers that may not have been eligible for low-income assistance due to eligibility rules such as requiring more than nine months residence in a dwelling and minimum levels of kWh usage. Customer eligibility parameters are below.

The WRAP program activities are divided into three different measure groups, to appropriately differentiate estimated energy savings. The measure groups are defined as follows:[2]

[1] "WRAP" is PPL's Low Income Usage Reduction Program (LIURP).
[2] APPRISE, PPL Electric Utilities Universal Service Programs Final Evaluation Report, Oct. 2008. Pg 50.

1. Base Load: Customers with no electric heat receive this type of service. Measures include CFLs, refrigerator replacement, air conditioner replacement, dryer venting, waterbed replacement, heating filter changing or cleaning, water heater set-back, and other measures that meet the PUC payback criteria. Effective 2008, base load recipients may receive up to $200 in comfort measures such as weather stripping and door sweeps.

2. Low Cost: In addition to the base load measures, customers with electric hot water are eligible for water heater replacement, Gravity Film Exchange (GFX), repairs of plumbing leaks, water pipe insulation, showerheads/aerators, and solar water heating.[3] Contractors can replace a washing machine with PPL approval.

3. Full Cost: Customers are eligible for full cost WRAP if the home has installed electric heat and the customer uses installed electric heat as the main heating source. The auditor can upgrade a base load job to full cost when full cost measures will reduce electric energy usage. This may include homes with defacto electric heat and high cooling usage. (The WRAP Standards and Field Guide defines defacto heating as when a household uses electric space heaters as the primary heating source, either in the entire home or a portion of the home.) In addition to the base load and water heating measures, they may receive heating and/or cooling measures, as well as additional follow-up energy education (site or phone). The additional measures for full cost customers include blower-door guided air sealing, insulation, heating repair/retrofit/replacement, cooling system repair and replacement, duct insulation and repair, caulking and weather stripping, and thermostat replacement.

2. EM&V Methodology

These EM&V guidelines are based on Option C – Whole Facility (Billing Regression Analysis as per Section 3.3.3.3.6.2.3 of the Audit Plan4), of the IPMVP5 for annual energy savings and coincident peak demand savings respectively. The results from the billing regression analysis will provide statistically adjusted engineering estimates (SAE) of ex-ante savings from this protocol and inform the inputs to the calibrated simulation model.

Ongoing evaluation has been built into the LIURP process since its initial implementation. Accordingly, PPL Electric collects data on each LIURP household for the thirteen month period prior to and following the installation of weatherization treatments. These data are reported to the Pennsylvania Public Utility Commission on a yearly basis.

A stipulated energy savings value will be used to impute savings until enough post-participation data are available to conduct the billing analysis, typically 12 months of post data. Savings for

[3] PPL does not require a payback for the solar water heating.

[4] *Audit Plan and Evaluation Framework for Pennsylvania Act 129 Energy Efficiency and Conservation Programs*. 2009. Prepared by the Statewide Evaluation Team: GDS Associates, Inc., Nexant and Mondre Energy. Contracted under the Pennsylvania Public Utility Commission's RFP 2009-1 for the Statewide Evaluator. December 1, 2009.

[5] *International Performance Measurement & Verification Protocol (IPMVP); Concepts and Options for Determining Energy and Water Savings: Volume 1*. Prepared by Efficiency Valuation Organization, www.evo-world.org. September 2009. EVO 10000 – 1:2009.

the regular WRAP are determined by an independent evaluator (APPRISE) and reports are submitted to and approved by the Pa PUC. These energy savings are reported in three measure groups, or job types discussed above: base load, low cost and full cost. The stipulated savings value for each job type will be based on the most recent PUC-approved average savings for each LIURP WRAP job type.

The following table shows annual results for PPL Electric's WRAP evaluations. In this table, 2006 results were based on billing analysis conducted by a third party. In 2007 and 2008 results are from annual reports based on the tracking system results. These figures are average gross savings per case. About a third of WRAP recipients gain energy during the post-period. PPL Electric enforced CAP benefit limits in 2008 which may have contributed to higher savings for non-heating customers.

Table 1. PPL Electric Savings by Job Types (2006 to 2008)

Year	Measure	Base load	Low Cost (Water Heating)	Full Cost (Heating)
2006[6]	Savings (kWh)	388	454	1,475
	% Reduced	4%	5%	7%
2007[7]	Savings (kWh)	316	693	914
	% Reduced	3%	5%	4%
2008[8]	Savings (kWh)	1,042	1,588	1,306
	% Reduced	8%	11.5%	6%

In 2008, PPL Electric provided usage reduction measures to 1,474 customers with installed electric heat and/or high cooling usage (Full Cost WRAP), 256 water-heating customers (Low Cost WRAP) and 1,351 lighting and appliance or base load customers (Base load WRAP). Customers in all three categories received base load measures and energy education.[9] The annual energy savings shown in Table 1 and Table 2 reflect the customer's actual bills based on consumption. Table 2 is based on 2008 electric distribution and generation rates.

Table 2. 2008 WRAP Average Bill Reduction

Job Type	Number of Customers	Annual Energy Savings (kWh)
Full Cost WRAP Jobs	1474	1306
Low Cost WRAP Jobs	256	1588
Base load WRAP Jobs	1351	1042

[6] Apprise, "PPL Electric Utilities Universal Services Programs Final Evaluation Report, October 2008.
[7] PPL Electric Utilities "Low Income Usage Reduction Program" 2007 Annual Report Winter Relief (WRAP) Program, April 30, 2009.
[8] PPL Electric Utilities "Low Income Usage Reduction Program" 2008 Annual Report Winter Relief (WRAP) Program, April 29, 2010.
[9] PPL, PPL Electric Utilities, Low Income Usage Reduction Program, 2008 Annual Report, April 29, 2010.

3. M&V Activities

The LIURP WRAP database (WRAP V) records three job types. The three Act 129 WRAP job types will each be considered a unique measure group with a stipulated savings value. Data for Act 129 participants will be uploaded to the PPL Act 129 data tracking system (EEMIS) and included in the M&V analyses.

Consumption histories will be used to determine Act 129-specific stipulated savings job type. The census of program participants will be included in the billing analysis.

4. Sampling and Metering Plan

Sampling of Installed Measures

Sampling of installed measures for verification will follow the Sampling and Uncertainty Protocol of Section 3.3.4 of the Audit Plan[10] to achieve a 90/10 level of confidence and relative precision. In the existing WRAP program, PPL Electric conducts QA/QC activities as one aspect of their M&V activities. An independent contractor conducts a site inspection for 60% of all full-cost jobs (expenditures over $750) to verify the installations. PPL Electric will continue this process, which includes Act 129 WRAP jobs. Sampling 60% of all full cost jobs far exceeds sampling that meets 90/10 levels of confidence and precision. For example, 1,474 full cost jobs were completed in 2008. A 90/10 sample requires about 70 verification site visits; 60% of the 1474 full cost jobs results in over 800 on-site verifications. No additional sampling of installed measures is planned beyond that already conducted through the existing WRAP.

Metering Plan

Metering is not recommended or needed; however, it is recommended that the billing analysis be repeated periodically to assess the persistence of savings.

5. Calculations and Adjustments

Energy Savings

For PPL Electric's PY 1 and PY2 reporting (Program Years 2009 ending 5/31/2010 and 2010 ending 5/31/2011), the *ex ante* and *ex post* savings for each Act 129 WRAP project will equal the evaluated gross energy savings for the applicable job type as reported in the 2008 LIURP Evaluation Final Report to the Pa PUC. The stipulated energy savings values for PY 1 and PY 2 will be based on the 2008 LIURP, differentiated by measure group and type as shown in Table 2. The savings values in the reports were determined by APPRISE's billing regression analysis as discussed above.

In PY 2 (6/1/10 - 5/31/11) the WRAP V database download into EEMIS will include the LIURP job type for each Act 129 WRAP transaction. The *ex ante* and *ex post* savings for each Act 129

[10] *Audit Plan and Evaluation Framework for Pennsylvania Act 129 Energy Efficiency and Conservation Programs*. 2009. Prepared by the Statewide Evaluation Team: GDS Associates, Inc., Nexant and Mondre Energy. Contracted under the Pennsylvania Public Utility Commission's RFP 2009-1 for the Statewide Evaluator. December 1, 2009.

WRAP project will equal the 2008 LIURP savings for the applicable job type. The job types and their stipulated savings are provided in Table 1.

In PY 3 (6/1/11 - 5/31/12) the gross impact evaluated energy savings will be determined using a billing regression analysis of the then available billing data of previous program years. New stipulated savings will be determined for the three job types and they will be used for the *ex ante* and *ex post* savings. The a billing regression analysis will be conducted if there is a statistically valid number of PY 1 participants (this is a short year because the program launched 11/09) and if there is enough time to complete the billing analysis before PY 3 starts. If there are not enough cases or other issues arise, the PY 3 savings will be based on the 2009 LIURP job types stipulated values (if those values are approved by the PUC in time); or some type of blended results based on 2008/2009 LIURP job types and/or a billing analysis for Act 129 WRAP transactions (PY 1 or PY 1 plus partial PY 2).

In PY 4 (6/1/12 - 5/31/13) new stipulated savings will be determined for the three job types and they will used for the ex-ante and ex-post savings. The new stipulated savings values for each job type will be determined from a billing regression analysis of Act 129 WRAP PY 2 participants. PY 2 4th quarter participants may need to be excluded to allow enough time to complete the 12 month pre- and post-participation billing analysis in time to apply the results commencing 6/1/12.

Coincident Peak Demand Savings

ACT 129 has specific requirements to compute and report demand savings. An effort must be made to quantify these savings for inclusion towards the ACT 129 goals. However, the Pa LIURP does not require EDCs to determine peak load reductions. Therefore, for Act 129 WRAP, peak load impacts will be determined using end-use load shapes in combination with PPL's top 100 hours to arrive at an estimated kW impact. The EM&V CSP will model the impacts specific to the WRAP participants given data are available to update load shapes for this population.

Custom Measure Protocol: Low-Income Energy Efficiency Program (LEEP) – Component 1

SWE Approved September 30, 2010

1. Measure Description

Part of the PECO portfolio of new energy efficiency (EE) and demand response (DR) programs is the Low-Income Energy Efficiency Program (LEEP), which educates and assists eligible residential customers to make their homes more energy efficient. The program consists of three components. This protocol will address the energy savings for LEEP Component 1 audits only.

Component 1 audits done as part of LEEP will be similar to the Low Income Usage Reduction Program (LIURP) which has provided energy efficiency services and energy education to PECO's low-income customers since 1988[1]. One difference is that LIURP addresses both electric and gas energy usage; whereas LEEP addresses only electricity demand and energy savings. Another is that the LIURP program income eligibility is 200% of the Federal Poverty Level (FPL) whereas LEEP is focusing on households below 150% of the FPL. However, customers with income up to 200% FPL may be enrolled in LEEP. These differences should not affect the potential electricity savings versus those of LIURP.

LEEP will expand the LIURP program and emphasize the following:

- Target the highest usage eligible customers, focusing on customers at or below 150% of the FPL, for program services.
- Expand installation of CFLs to include all lighting used for more than three hours a day.
- Address supplemental electric heating that may be due to many factors, including inoperable central heating systems.
- Continue customer energy education during the audits and in follow-up contacts and provide customers additional information about other energy efficiency programs.
- Conduct impact evaluations using techniques that isolate and estimate the energy bill effects of the program services provided to participants.

For LEEP Component 1, the expansion of the LIURP program to accommodate more participants is expected to double the number of participants over the 2008 LIURP level of 9,000 households by May 2013. This will result in up to an additional 20,000 participants receiving installations by May 2013.

Participating households will receive two types of assistance:

[1] PECO Energy has implemented a set of Universal Services Programs to meet requirements set by Pennsylvania's electric and gas restructuring legislation and various Public Utility Commission orders and agreements. The Universal Services Programs include: 1) A CAP payment assistance Program that is designed to make energy bills more affordable by furnishing payment subsidies; 2) A LIURP Program that is designed to make energy bills more affordable by helping to reduce usage; and 3) A CARES Program that is designed to assist households in developing appropriate strategies for maintaining energy service.

- **In-home Audits and Education**—These are on-site audits of appliances and energy use and tests used to identify the applicability of energy-savings measures the program offers and to educate residents about ways to reduce their energy usage.
    - Trained auditors perform on-site audits (air leak testing and home inspection) and assess the energy performance of the house; i.e., identify where energy is used and where there are inefficiencies and determine which measures are appropriate to install.
    - The auditors discuss the opportunities to reduce energy use and bills with residents.
    - Follow-up contacts with the participants reinforce the message of the benefits of energy-saving behaviors (e.g., turning off lights in unoccupied rooms) and adoption of energy-savings measures offered by the auditors.
- **Direct Installation of Measures**—Install measures to reduce energy use in the home at no charge to residents.
    - This aspect of the LIURP-type program will be expanded to bring services to more households. Applicable measures will continue to be installed, at no cost to residents, in the same way as they have been in past LIURP programs.
    - Focus mainly on measures consistent with PECO's current LIURP program. CFLs will be expanded to include all lights used more than three hours daily, with an expected increase of CFL installations from an average of 4 bulbs up to 10 bulbs per household.
    - Evaluate repair or replacement of non-working gas heating units to remove electric space heaters from use.
    - Install ENERGY STAR appliances as applicable.

LEEP Component 1 (audit) is subdivided into 3 different measure groups (electric baseload, electric heat, gas heat) and 2 measure types (basic and major), plus additional CFLs, to appropriately differentiate estimated energy savings. The measure groups are defined as follows:

- *Electric Baseload – Basic Measure*: includes measures such as CFLs (4), refrigerator removal, AC maintenance, faucet aerator, showerhead, water heater pipe insulation, water heater tank insulation, etc.
- *Electric Baseload – Major Measure*: includes same measures as the Electric Baseload – Basic Measure plus room/wall AC replacement, refrigerator replacement, electric water heater replacement[2], and water heater timers (electric water heaters only).
- *Electric Heat – Basic Measure*: includes same measures as the Electric Baseload – Basic Measure plus duct and pipe insulation, programmable thermostats, etc.
- *Electric Heat – Major Measure*: includes same measures as the Electric Heat – Basic Measure plus blower door guided air sealing, heat pump installation/replacement, and insulation installation
- *Gas Heat – Basic Electric Measure*: includes measures such as CFLs (4), refrigerator removal, and AC maintenance
- *Gas Heat – Major Electric Measure*: includes same measures as the Gas Heat – Basic Measure plus refrigerator replacement and room/wall AC replacement

[2] Note that most existing water heaters are gas, therefore water heater replacement will be a small overall component of the program.

- *CFLs*: standard "screw-in" compact fluorescent bulbs beyond the initial 4 included as part of the primary measure groups

2. EM&V Methodology

These guidelines are based on Option C – Whole Facility (Billing Regression Analysis as per Section 3.3.3.3.6.2.3 of the Audit Plan[3]), and Option D – Calibrated Simulation (Whole Building Simulation as per Sections 3.3.3.3.6.2.2 and 3.3.3.3.7.2 of the Audit Plan[4]) of the IPMVP[5] for annual energy savings and coincident peak demand savings respectively. The results from the billing regression analysis will provide statistically adjusted engineering estimates (SAE) of ex-ante savings from this protocol and inform the inputs to the calibrated simulation model. Additionally the calibrated simulation model will serve as a verification of the billing regression analysis results.

The use of both Option C and Option D will classify the gross energy and demand impact evaluations of LEEP within the IPMVP Enhanced Level of Rigor.

3. M&V Activities

An audit report shall be completed for each participant/unit on a standard form. At a minimum, the following information should be included for each participant/unit:

- Participant characteristics (name, address, phone, etc.)
- Existing home characteristics
- Measure group implemented: Electric Baseload – Basic, Electric Baseload – Major, Electric Heat – Basic, Electric Heat – Major, Gas Heat – Basic Electric, Gas Heat – Major Electric, CFLs[6]
- List of individual measures implemented within the measure group such as: AC replacement, AC maintenance, number of CFLs, refrigerator removal, refrigerator replacement, faucet aerator, showerhead, water heater pipe insulation, water heater tank insulation, water heater replacement, etc.
- Details on individual measures. For example:
    - Existing lamp and replacement CFL wattage, estimated hours of use
    - Existing and replacement air conditioner capacity, model number, efficiencies, etc.

---

[3] *Audit Plan and Evaluation Framework for Pennsylvania Act 129 Energy Efficiency and Conservation Programs*. 2009. Prepared by the Statewide Evaluation Team: GDS Associates, Inc., Nexant and Mondre Energy. Contracted under the Pennsylvania Public Utility Commission's RFP 2009-1 for the Statewide Evaluator. December 1, 2009.

[4] *Ibid.*

[5] *International Performance Measurement & Verification Protocol (IPMVP); Concepts and Options for Determining Energy and Water Savings: Volume 1*. Prepared by Efficiency Valuation Organization, www.evo-world.org. September 2009. EVO 10000 – 1:2009.

[6] The intention is to target replacement of existing incandescent lamps that are on 3 or more hours per day. This inherently requires estimation of hours of use by determining if a lamp uses more than 3 hours, but it does not include actual metering using logging equipment which would not be cost effective, nor does it include an estimation of the actual usage beyond 3 hours. For example if a lamp is identified as on for more than 3 hours it will be replaced and recorded, but the implementer will not make an estimate that it is on for 3 or 4 or 5 hours, etc. If it turns out it was on for less than 3 hours, that will be picked up in the billing analysis. As will savings for more than 3 hours. Thus in the long run the actual savings will be picked up rather than an estimate.

- Existing and replacement refrigerator type, model number, wattage, etc.
- Number of faucet aerators and showerheads
- Replacement insulation R-values

4. Sampling and Metering Plan

Sampling

Sampling of installed measures for verification will follow the Sampling and Uncertainty Protocol of Section 3.3.4 of the Audit Plan[7] to achieve a 90/10 level of confidence and relative precision. Sampling population size should follow the Simple Random Sampling method as described in Section 3.3.4.1 of the Audit Plan. Sampling will be conducted on a quarterly basis for both phone surveys and on-site verification.

Spot Metering (SP)

When appropriate (for example blower door test, refrigerator power, HVAC power, etc), spot metering of measures will be used to inform calibrated simulation model inputs for verification of energy savings and determination of demand savings. Sampling for spot metering will follow the Sampling and Uncertainty Protocol of Section 3.3.4 of the Audit Plan.

Short Term Metering (ST)

Similar to spot metering, short term metering of measures will be used as necessary to inform calibrated simulation model inputs for verification of energy savings and determination of coincident demand savings. Sampling for short term metering will follow the Sampling and Uncertainty Protocol of Section 3.3.4 of the Audit Plan.

5. Calculations and Adjustments

Energy Savings

Component 1 (audits) of LEEP is delivered using a customized approach for each participant, including behavioral measures and technical performance measures that are specific to a participant. Billing analysis—comparing the home's energy use pre- and then post-treatment—is the most appropriate method to use to account for this customized approach and has been applied in several jurisdictions[8,9,10].

---

[7] *Audit Plan and Evaluation Framework for Pennsylvania Act 129 Energy Efficiency and Conservation Programs*. 2009. Prepared by the Statewide Evaluation Team: GDS Associates, Inc., Nexant and Mondre Energy. Contracted under the Pennsylvania Public Utility Commission's RFP 2009-1 for the Statewide Evaluator. December 1, 2009.

[8] Evaluation of the Washington State Weatherization Assistance Program. 2001. Prepared by Oak Ridge National Laboratory, Energy Division, ORNL/CON-478. March 2001.

[9] *Ohio Electric Partnership Program Impact Evaluation, Final Report*. 2006. Prepared by M. Blasnik & Associates. Prepared for the Ohio Office of Energy Efficiency. June 30, 2006.

[10] *New Hampshire Weatherization Program Impact Evaluation Report*. 2007. Prepared by M. Blasnik & Associates. Prepared for New Hampshire Office of Energy and Planning. April 9, 2007.

APPRISE has been using billing analysis to do the impact evaluation for LIURP for many years[11]. The APPRISE approach has been reviewed and found to be robust and consistent with the International Performance Measurement and Verification Protocols.

A similar billing analysis approach is recommended for the LEEP gross energy impact evaluation. This approach, however, requires one year of post-participation data. This would result in a 2011 fourth quarter completion of the billing analysis of Program Year 2009 (PY2009 and ending 5/31/2010) participants. Due to the ACT 129 annual reporting requirements, PY2009 evaluation results being finalized at the end of 2011 will not work. Therefore, the following approach is recommended.

For Program Year 2009 (PY2009 ending 5/31/2010), evaluated gross energy savings for LEEP will be based on a stipulated energy savings value. The stipulated energy savings values for PY2009 will be based on the 4-year average of the 2005-2008 PECO Energy LIURP Evaluation Final Reports, differentiated by measure group and type as shown in Table 1 below. The savings values in the reports were determined by APPRISE's billing regression analysis as mentioned above.

For subsequent program years (PY2010 and thereafter) the gross impact evaluated energy savings will be determined using a 4-year rolling average using the latest 4-years of available annual billing regression analysis of previous program years. For PY2010, this 4-year rolling average will use LIURP data from 2006-2009. PY2011 savings will be determined using a 4-year average of billing analysis results from the 2007-2009 LIURP programs and PY2009 LEEP data (Note that LIURP program years are with the calendar year whereas LEEP program years are from June 1 to May 31. This affects when the billing analysis results are available for inclusion in an average). PY2012 savings will be determined using a 4-year average of billing analysis results from the 2008-2009 LIURP programs and PY2009 and PY2010 LEEP data. The rolling 4-year average stipulated savings will be adjusted annually thereafter using the additional yearly program data. Eventually the stipulated value will be based entirely on LEEP program data.

The proposed LEEP Component 1 (audits), PY2009 stipulated energy savings values are based on the results from the 2005, 2006, 2007, and 2008 APPRISE billing analyses of LIURP and shown in Table 1 below. The stipulated energy savings are determined using the average annual savings by measure group and type (Table 2 below) for the three measure groups of LIURP jobs—Electric Baseload, Electric Heat, and Gas Heat, using both Basic and Major measure types. The analysis includes consideration of household and home characteristics, and adjusts for the impact of weather. Note that Electric Baseload jobs included all types of heating; Electric Heat jobs were those involving weatherization and/or electric heating equipment measures.

The SWE Team has determined that it will be necessary for PECO to report information on the number and types of measures installed each year. This report is necessary to verify that savings analysis is consistent between years and that the 4-year average measure distribution is representative of the current program year.

[11] 2005-2008 PECO Energy LIURP Evaluation Final Reports. Prepared by APPRISE (Applied Public Policy Research Institute for Study and Evaluation).

Table 1 Stipulated Audit Savings by Measure Group

Table 1a: Stipulated Energy Savings for Electric Baseload Jobs

Electric Baseload (Appliances & Lighting)	
Basic: -(1-4) CFL bulbs -Pipe & water heater wrap -Efficient showerheads -Efficient faucet aerators -Smoke detectors -Refrigerator Removal -AC Maintenance	**Major:** -Includes Basic measures and -Refrigerator replacement -Air conditioner replacement -Water heater replacement -Electric water heater timer
738 kWh	**1,644 kWh**

Table 1b: Stipulated Energy Savings for Electric Heat Jobs

Electric Heat (Weatherization, Insulation, Heating system repair/replacement)	
Basic: -(1-4) CFL bulbs -Pipe & water heater wrap -Efficient showerheads -Efficient faucet aerators -Smoke detectors -Refrigerator Removal -AC Maintenance -Duct insulation -Programmable thermostats	**Major:** -Includes Basic measures and -Refrigerator replacement -Air conditioner replacement -Water heater replacement -Electric water heater timer -Heat pump installation/replacement -Insulation -Blower door guided air sealing
623 kWh	**1,840 kWh**

Table 1c: Stipulated Energy Savings for Gas Heat Jobs

Gas Heat (Weatherization, Insulation, Cooling system repair/replacement)	
Basic: -(4) CFL bulbs -Pipe & water heater wrap -Efficient showerheads -Efficient faucet aerators -Smoke detectors -Refrigerator replacement -AC Maintenance	**Major:** -Includes Basic measures and -Refrigerator replacement -Air conditioner replacement
338 kWh	**1,640 kWh**

Table 2 below shows the derivation of the stipulated audit electric energy savings (kWh) for each of the measure groups and measure types (basic, major). The primary data is available (sample sizes, annual weather data, home and customer characteristics, etc.) from the 2005-2008 LIURP evaluation reports and the evaluator for these savings values.

Table 2: LEEP Component 1 (kWh) Stipulated Savings Calculations

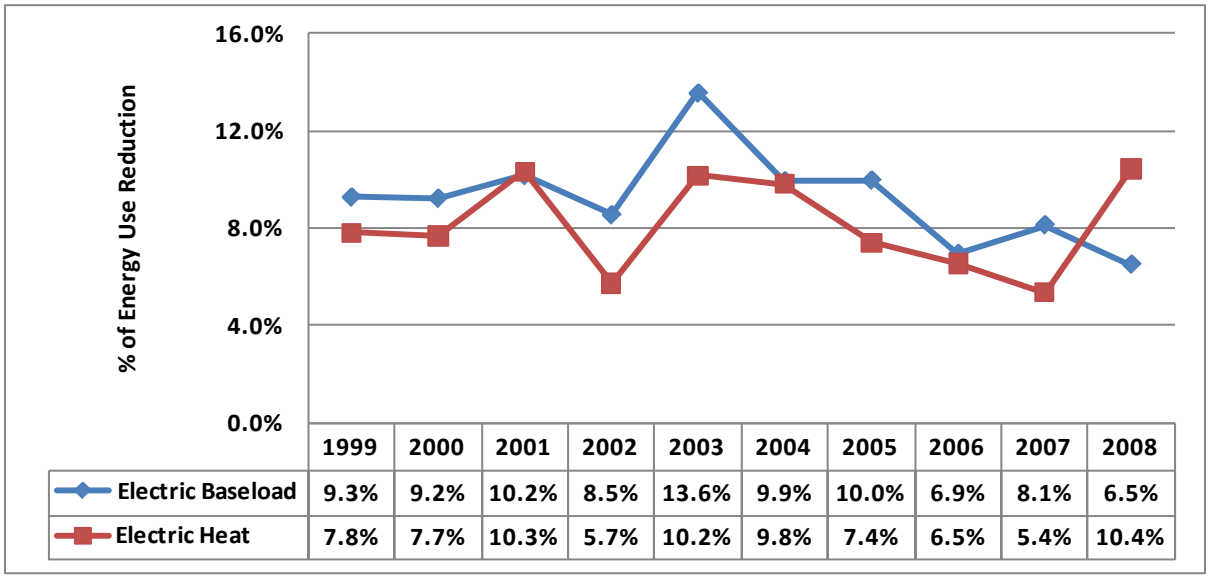
Program Year	Electric Baseload		Electric Heat		Gas Heat	
	Basic	Major	Basic	Major	Basic	Major
2005	955	2,049	532	1,789	478	2,834
2006	614	1,442	560	1,817	104	691
2007	775	1,504	382	1,374	339	1,705
2008	606	1,581	1018	2,381	429	1,328
Average	**738**	**1644**	**623**	**1840**	**338**	**1640**

Sources:

1. 2005-2008 PECO Energy LIURP Evaluation Final Reports. Prepared by APPRISE (Applied Public Policy Research Institute for Study and Evaluation).

Results from the long history of billing analyses done for the LIURP program indicate basic stability in the estimate of savings by customer. While billing analysis results may show some fluctuation in savings estimates from year to year, this is normal variability and is to be expected. The stability of the savings averages over the past nine years is a strong endorsement of their use for stipulated savings by measure group and type. Figure 1 shows that while there is fluctuation around the average from year to year, there is no strong indication of a changing level of percent savings for any of the measure groups. Use of the long-term average energy use reduction (kWh saved) is a reasonable extension of customer-level savings from LIURP to stipulated energy savings for LEEP Component 1.

Figure 1 Percent of Electricity Savings from 1999 to 2008 by Measure Group



Source: PECO Energy 2008 LIURP Evaluation Report.

Coincident Peak Demand Savings

With all of its strengths, billing analysis does not provide enough information to develop coincident peak demand savings (herein referred to as "demand savings"). However, ACT 129 has specific demand savings requirements. Knowing there will be demand savings associated with the LEEP Component 1 measures, an effort must be made to quantify these savings for inclusion towards the ACT 129 goals. Therefore, to determine demand savings the following approach is recommended.

For Program Years 2009 (PY2009 ending 5/31/2010) and 2010 (PY2010 ending 5/31/2011), ex ante gross demand savings for LEEP Component 1 audits will be based on a stipulated demand savings value. Ex ante demand savings are differentiated by measure group and type as shown in Table 3 below. The ex ante demand savings are calculated by multiplying the stipulated energy savings from Table 1 above by a coincident peak demand savings conversion factor of 0.000122 kW/kWh. This factor was derived from Global Energy Partner's BEST model results for PECO's approved Energy Efficiency and Conservation Plan[12]. BEST is an internally developed user interface for the DOE2.2 simulation engine. The model was developed using a prototypical North-East residential building and Philadelphia weather data. A representative sample of measures within the LEEP Component 1 program were applied to the baseline model and the simulation results were used to derive the LEEP Component 1 kW to kWh conversion factor. The model results provide a reasonably accurate, best available, demand savings for the various LEEP Component 1 measure groups and types until a more robust model can be developed using LEEP program specific data.

The ex post gross impact evaluated demand savings will be determined for PY2011 and beyond using Option D – Calibrated Simulation (Whole Building Simulation as per Sections 3.3.3.3.6.2.2 and 3.3.3.3.7.2 of the Audit Plan[13]) of the IPMVP[14]. Data from the LIURP program evaluations will be used to inform the inputs to the calibrated simulation model, as will LEEP program tracking information on measures installed, and spot metering and short term metering sampling results. A load shape will then be developed from the calibrated model and applied to the annual energy savings to determine the ex post demand savings.

Additionally the calibrated simulation model will serve as a verification of the billing regression analysis results. Ex ante demand savings can be adjusted annually thereafter using the additional yearly program data.

Savings Summary

Table 3 shows a summary of the measure characteristics, including savings, life, incremental cost and unit definitions, for the various Component 1 audit measure groups and types. Because the LIURP program consistently installed a maximum of 4 CFL bulbs in each participant's home, the LIURP billing analysis data can only be extended to the LEEP program energy savings from the first 4 CFL's installed. For participants that are given more than 4 CFLs as part of Component 1 or Component 2, or given any CFLs

---

[12] PECO Energy Efficiency and Conservation Plan (Program Years 2009-2012). Prepared July 1, 2009.
[13] *Audit Plan and Evaluation Framework for Pennsylvania Act 129 Energy Efficiency and Conservation Programs*. 2009. Prepared by the Statewide Evaluation Team: GDS Associates, Inc., Nexant and Mondre Energy. Contracted under the Pennsylvania Public Utility Commission's RFP 2009-1 for the Statewide Evaluator. December 1, 2009.
[14] *International Performance Measurement & Verification Protocol (IPMVP); Concepts and Options for Determining Energy and Water Savings: Volume 1*. Prepared by Efficiency Valuation Organization, www.evo-world.org. September 2009. EVO 10000 – 1:2009..

as part of Component 3, the deemed savings for these CFLs will be based on the PA TRM. For additional CFL's that are part of Component 1, 2, or 3, the energy and demand savings must be recorded separately following the savings methodology defined in the 2010 PA TRM with a realization rate adjustment to the In-Service Rate (ISR) default based on the evaluation results.

Table 3: LEEP Component 1 Measure Characteristics

Measures	Annual kWh Savings per Unit	Coincident Peak kW Savings per Unit	Measure Life (years)	LIURP 2008 Measure Cost	Unit Definition	Sources
Electric baseload - basic measures (C1)	738	0.090[6]	6.4	$ 109	per installation	2,3,4,5
Electric baseload - major measures (C1)	1644	0.201[6]	13	$ 855	per installation	2,3,4,5
Electric heat - basic measures (C1)	623	0.076[6]	6.4	$ 756	per installation	2,3,4,5
Electric heat - major measures (C1)	1840	0.225[6]	13	$ 2,617	per installation	2,3,4,5
Gas heat - basic electric measures (C1)	338	0.041[6]	6.4	$ 1,128	per installation	2,3,4,5
Gas heat - major electric measures (C1)	1640	0.200[6]	13	$ 1,955	per installation	2,3,4,5
Extra CFLs (C1, C2, C3):						
60W to 13W CFL	Per TRM	Per TRM	6.4	$ 4	per lamp	
75W to 19W/20W CFL	Per TRM	Per TRM	6.4	$ 4	per lamp	1

Sources:

1. *Technical Reference Manual (TRM) for Pennsylvania Act 129 Energy Efficiency and Conservation Program and Act 213 Alternative Energy Portfolio Standards*, Pennsylvania Public Utility Commission, June 2010

2. *PECO Energy 2005 LIURP Evaluation Final Report*. 2007. Prepared by APPRISE (Applied Public Policy Research Institute for Study and Evaluation). April 2007

3. *PECO Energy 2006 LIURP Evaluation Final Report*. 2008. Prepared by APPRISE (Applied Public Policy Research Institute for Study and Evaluation). April 2008

4. *PECO Energy 2007 LIURP Evaluation Final Report*. 2009. Prepared by APPRISE (Applied Public Policy Research Institute for Study and Evaluation). April 2009

5. *PECO Energy 2008 LIURP Evaluation Final Report*. 2010. Prepared by APPRISE (Applied Public Policy Research Institute for Study and Evaluation). April 2010

6. Multiply annual kWh Savings per unit by a coincident peak demand savings conversion factor of 0.000122 kW/kWh, derived from Global Energy Partner's BEST model results for PECO's approved *Energy Efficiency and Conservation Plan (Program Years 2009-2012)*. July 1, 2009.

The CFL measure savings, cost, and useful life values are the same as in the CFL Initiative program—TRM savings and life values and current prices in the PECO service territory. Basic measures are assigned the same EUL as CFLs and major measures are assigned the EUL for refrigerator replacement of 13 years as per the TRM. CFLs are the most common baseload measure installed and refrigerators the most common major measures installed.. . All the savings are gross values and do not include any free-rider or free-driver effects.

Custom Measure Protocol: Residential LED Lighting

1. Measure Description

This measure is the replacement of inefficient residential incandescent lamps with more efficient LED lamps.

This new protocol is being proposed because the June 2010 TRM[1] only covers Residential ENERGY STAR Lighting (Section 4.2), but does not address these products which provide an efficient alternative to inefficient incandescent lamps, but did not have an Energy Star Rated Category.

This measure is being submitted as a temporary Custom Measure Protocol with stipulated savings similar to an Interim TRM protocol. As new products become available which meet the new Energy Star Criteria for Residential LED Lighting released in August 2010, a new Interim TRM Protocol will be developed to include those products.

2. Algorithms

The measure savings are based on the June 2010 TRM[2] Section 4.2 Residential ENERGY STAR Lighting algorithms with the following adjustments. Note that these LED lamps are intended by the manufacturers to replace specific incandescent lamps, therefore there is a specific savings value for each lamp type.

Residential LED Candelabra Lamp (2W, 2.5W, or 3W LED)

- Energy Impact (kWh) = (LC_{Watts} * (LC_{Hours} * 365) / 1000) * ISR_{LED} = 6 kWh
- Peak Demand Impact (kW) = (LC_{Watts} / 1000) * CF_{LC} = 0.0005 kW

Residential LED Globe Lamp (2W G25 or 2W G16.5 LED)

- Energy Impact (kWh) = (LG_{Watts} * (LG_{Hours} * 365) / 1000) * ISR_{LED} = 13 kWh
- Peak Demand Impact (kW) = (LG_{Watts} / 1000) * CF_{LG} = 0.0014 kW

Residential LED A15 Lamp (2.5W A15 LED)

- Energy Impact (kWh) = (LA_{Watts} * (LA_{Hours} * 365) / 1000) * ISR_{LED} = 9 kWh
- Peak Demand Impact (kW) = (LA_{Watts} / 1000) * CF_{LA} = 0.0008 kW

Residential LED R20 Lamp (7W R20 LED)

- Energy Impact (kWh) = (LR_{Watts} * (LR_{Hours} * 365) / 1000) * ISR_{LED} = 24 kWh
- Peak Demand Impact (kW) = (LR_{Watts} / 1000) * CF_{LR} = 0.0022~~1~~ kW

[1] *Technical Reference Manual (TRM) for Pennsylvania Act 129 Energy Efficiency and Conservation Program and Act 213 Alternative Energy Portfolio Standards*, Pennsylvania Public Utility Commission, June 2010

[2] Ibid

Residential LED PAR20 Lamp (7W PAR20 LED)

- Energy Impact (kWh) = ($LP20_{Watts}$ * ($LP20_{Hours}$ * 365) / 1000) * ISR_{LED} = 31 kWh
- Peak Demand Impact (kW) = ($LP20_{Watts}$ / 1000) * CF_{LP20} = 0.00297 kW

Residential LED PAR30 Lamp (11W PAR30 LED)

- Energy Impact (kWh) = ($LP30_{Watts}$ * ($LP30_{Hours}$ * 365) / 1000) * ISR_{LED} = 28 kWh
- Peak Demand Impact (kW) = ($LP30_{Watts}$ / 1000) * CF_{LP30} = 0.00264 kW

Residential LED PAR38 Lamp (16W PAR38 LED)

- Energy Impact (kWh) = ($LP38_{Watts}$ * ($LP38_{Hours}$ * 365) / 1000) * ISR_{LED} = 21 kWh
- Peak Demand Impact (kW) = ($LP38_{Watts}$ / 1000) * CF_{LP38} = 0.00198 kW

Residential LED MR16 Lamp (4W MR16 LED)

- Energy Impact (kWh) = ($LMR16_{Watts}$ * ($LMR16_{Hours}$ * 365) / 1000) * ISR_{LED} = 12 kWh
- Peak Demand Impact (kW) = ($LMR16_{Watts}$ / 1000) * CF_{LMR16} = 0.001109 kW

3. Definition of Terms

These savings numbers are derived from the following assumptions:

- LC_{Watts} = Average delta watts per purchased LED Candelabra lamp
- LC_{Hours} = Average hours of use per day per purchased LED Candelabra lamp
- CF_{LC} = Peak Coincidence Factor for LED Candelabra lamps
- LG_{Watts} = Average delta watts per purchased LED Globe lamp
- LG_{Hours} = Average hours of use per day per purchased LED Globe lamp
- CF_{LG} = Peak Coincidence Factor for LED Globe lamps
- LA_{Watts} = Average delta watts per purchased LED A15 lamp
- LA_{Hours} = Average hours of use per day per purchased LED A15 lamp
- CF_{LA} = Peak Coincidence Factor for LED A15 lamps
- LR_{Watts} = Average delta watts per purchased LED R20 lamp
- LR_{Hours} = Average hours of use per day per purchased LED R20 lamp
- CF_{LR} = Peak Coincidence Factor for LED R20 lamps
- $LP20_{Watts}$ = Average delta watts per purchased LED PAR20 lamp
- $LP20_{Hours}$ = Average hours of use per day per purchased LED PAR20 lamp
- CF_{LP20} = Peak Coincidence Factor for LED PAR20 lamps
- $LP30_{Watts}$ = Average delta watts per purchased LED PAR30 lamp
- $LP30_{Hours}$ = Average hours of use per day per purchased LED PAR30 lamp
- CF_{LP30} = Peak Coincidence Factor for LED PAR30 lamps
- $LP38_{Watts}$ = Average delta watts per purchased LED PAR38 lamp
- $LP38_{Hours}$ = Average hours of use per day per purchased LED PAR38 lamp
- CF_{LP38} = Peak Coincidence Factor for LED PAR38 lamps

- $LMR16_{Watts}$ = Average delta watts per purchased LED MR16 lamp
- $LMR16_{Hours}$ = Average hours of use per day per purchased LED MR16 lamp
- CF_{LMR16} = Peak Coincidence Factor for LED MR16 lamps
- ISR_{LED} = Residential LED In Service Rate 100%[3]

Table 0-1. Residential LED Lamp Default Savings

Variable	Type	Delta Watts	Incandescent Baseline Lamp	LED Lamp	Energy Impact (kWh)	Demand Impact (kW)
LC_{Watts}	Fixed	13.2	15W Candelabra Base	2W, 2.5W or 3W Candelabra LED (Average Input wattage = 1.8W)	6	0.00048
LG_{Watts}	Fixed	23.0	25W G25 or G16.5 Lamp	2W G25 or 2W G16.5 LED (Average Input wattage = 2.0W)	13	0.001453
LA_{Watts}	Fixed	12.5	15W A15 Lamp	2.5W A15 LED (Input wattage = 5.5W)	9	0.0008378
LR_{Watts}	Fixed	34.5	40W R20 Halogen Lamp	7W R20 LED (Input wattage = 5.5W)	24	0.002407
$LP20_{Watts}$	Fixed	44.3	50W PAR20 Halogen Lamp	7W PAR20 LED (Input wattage = 5.7W)	31	0.002866
$LP30_{Watts}$	Fixed	40.4	50W PAR30 Halogen Lamp	11W PAR30 LED (Input wattage = 9.6W)	28	0.002642
$LP38_{Watts}$	Fixed	29.6	45W PAR38 Halogen Lamp	16W PAR38 LED (Input wattage = 15.4W)	21	0.001978
$LMR16_{Watts}$	Fixed	16.7	20W MR16 Halogen Lamp	4W MR16 LED (Average Input wattage = 3.3 W)	12	0.00107094

[3] California Public Utilities Commission, Energy Division. "Final Evaluation Report: Upstream Lighting Program; Volume 1." Prepared by KEMA, Inc., The Cadmus Group, Inc., Itron, Inc., PA Consulting Group, and Jai J. Mitchell Analytics. February 8, 2010.

Table 0-2. Residential LED Variables

Variable	Type	Value	Source
LC_{Hours}	Fixed	1.2	1
CF_{LC}	Fixed	0.036	1
LG_{Hours}	Fixed	1.5	1
CF_{LG}	Fixed	0.063~~2~~	1
LA_{Hours}	Fixed	2.0	1
CF_{LA}	Fixed	0.066~~2~~	1
LR_{Hours}	Fixed	1.9	1
CF_{LR}	Fixed	0.065~~0~~	1
$LP20_{Hours}$	Fixed	1.9	1
CF_{LP20}	Fixed	0.065~~0~~	1
$LP30_{Hours}$	Fixed	1.9	1
CF_{LP30}	Fixed	0.065~~0~~	1
$LP38_{Hours}$	Fixed	1.9	1
CF_{LP38}	Fixed	0.065~~0~~	1
$LMR16_{Hours}$	Fixed	1.9	1
CF_{LMR16}	Fixed	0.0645~~6~~	1
ISR_{LED}	Fixed	100%	1

Sources:

1. California Public Utilities Commission, Energy Division. "Final Evaluation Report: Upstream Lighting Program; Volume 1." Prepared by KEMA, Inc., The Cadmus Group, Inc., Itron, Inc., PA Consulting Group, and Jai J. Mitchell Analytics. February 8, 2010.

4. Measure Life

- Residential LED Measure Life = 15 yrs[4]

[4] The measure life for each LED lamp type was determined by dividing the manufacturers' estimated lifetime in hours by the annual hours of use for each lamp type. All lifetimes were greater than 15 yrs (range of 16-65 yrs), however, the PA 2010 TRM (Sections 1.14 and 8.1) effectively caps the lifetime of a measure at 15 yrs by not allowing savings to be claimed beyond 15 yrs.

Appendix E. Common Approach for Measuring Net Savings for Appliance Retirement Programs

See memo on the following pages.

research into action inc

Research Into Action, Inc.
PO Box 12312
Portland, OR 97212

www.researchintoaction.com

503 287 9136 (main)
503 281 7375 (fax)
888 492 9100 (toll free)

Memorandum

To: PA Program Evaluation Group (PEG)

From: Scott Dimetrosky and Noah Lieb, Apex Analytics

Jane Peters and Ryan Bliss, Research Into Action

Date: March 14, 2014

Re: Common Approach for Measuring Net Savings for Appliance Retirement Programs

1. Recommended Standard Net Savings Protocol

Appliance retirement programs (ARP) typically offer some mix of incentives and free pickup for the removal of old-but-operable refrigerators, freezers, or room air-conditioners. These programs are designed to encourage the consumer to:

- Discontinue the use of secondary or inefficient appliances
- Relinquish appliances previously used as primary units when they are replaced (rather than keeping the old appliance as a secondary unit)
- Prevent the continued use of old appliances in another household through a direct transfer (giving it away or selling it) or indirect transfer (resale on the used appliance market)

Because the program theory and logic for appliance retirement differs significantly from standard "downstream" incentive programs (which typically offer rebates for the purchase of efficient products) the approach to estimating free-ridership is also significantly different. Consistent with the Pennsylvania TRM, which relies on the Department of Energy Uniform Methods project as the default inputs for estimating gross savings, the SWE Team recommends that the Pennsylvania EDCs also follow the UMP guidelines for estimating program net savings.[1]

In the following sections we present the UMP approach, adding in clarifying explanations/diagrams where applicable.

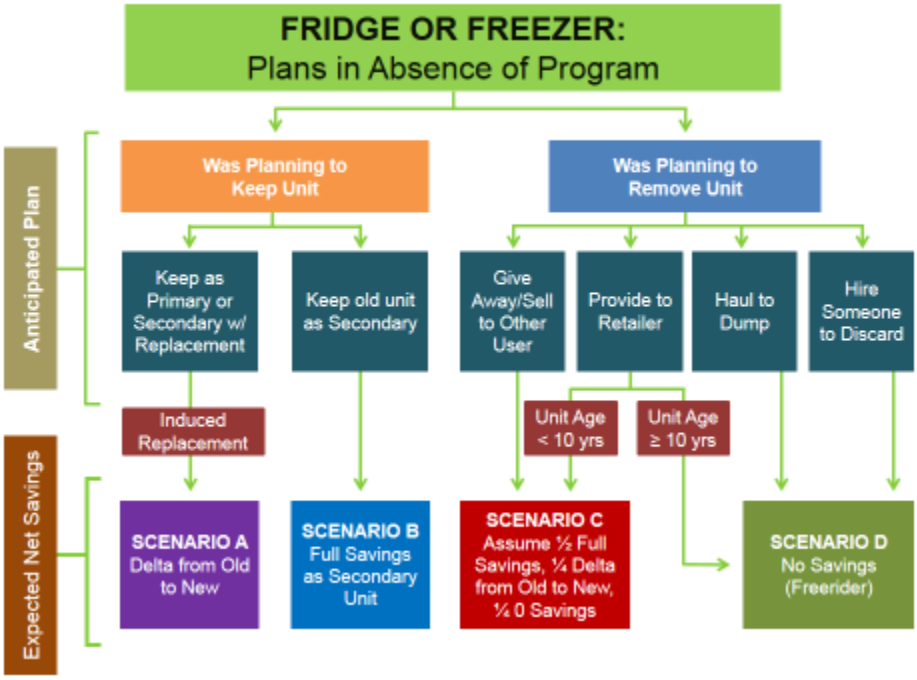
[1] See The Uniform Methods Project: Methods for Determining Energy Efficiency Savings for Specific Measures, Chapter 7: Refrigerator Recycling Evaluation Protocols, National Renewable Energy Laboratory, March 2013 (Download available at: *http://www1.eere.energy.gov/wip/pdfs/53827-7.pdf*)

1.1. General Free-ridership Approach

The nature of the appliance retirement program requires a unique approach to estimating free-ridership, and ultimately, net savings. Free-ridership is based on the participants anticipated plans had the program not been available – a free-rider is classified as one who would have removed the unit from service irrespective of the program. Net savings for the appliance retirement program is therefore based on the participants' anticipated continued operation of the appliance– either as primary or secondary unit, within their home or transferred to another home (either directly or indirectly).

The general approach to estimating net savings for the appliance retirement program is a several-step process to segment the participants into different groups, each with unique savings attributable to them. Participants should first be classified as either "keepers" or "removers." The "keepers" segment, defined as those who were planning on keeping the unit, should be further segmented into groups based on whether they replaced the unit. The "removers" segment, defined as those who were planning on removing the unit, should be further segmented into groups based on whether the unit would have continued operating or been permanently removed from service. Each respondent is then assigned a net savings value, and overall program net savings is calculated in aggregate across the sample population. A simple flow chart, included below in Figure 1, shows how the net savings are derived for the appliance retirement program. A more detailed discussion follows below.

Figure 1. Diagram to Determine Appliance Retirement Net Savings[2]



1.1.1. Classifying the Participant as "Keeper or Remover"

The first step is to classify each participant as a keeper or remover. This first classification is assessed through a series of questions used to determine what the participant likely would have done if the appliance had not been removed by the program. The following example shows the basic approach:[3]

1. Were you using the recycled unit as your primary [**appliance**], or had it been a secondary or spare?
   a. Primary
   b. Secondary

2. If the appliance pickup program was not available, would you have still removed the [**appliance**], or would you have kept it?
   a. Removed it
   b. Kept it

[2] The savings in this figure are the deemed value from the 2013 PA TRM, and are presented as an example only as updates to these values have been proposed for the 2014 PA TRM.

[3] Note these questions are provided as examples of questions to derive the information needed to classify participants into the various scenarios. EDCs can adapt these questions as long as they can provide the same information.

1.1.2. Classifying the status of the "Keeper"

The "keepers" segment, as discussed previously, would be assigned to those who answered question 2 above as "b. kept it". The keepers segment are not qualified as free-riders, assuming in absence of the program their [**appliances**] would have continued operating normally. These respondents should be further segmented into groups based on whether they replaced their unit – this helps define the deemed savings values to be assigned to them. The following question is an example of what to ask:

3. Did you replace the removed [**appliance**] with a different unit?
   a. Yes, replaced the unit (Scenario A, below)
   b. No, did not replace the unit (Scenario B)

The "keeper" respondents that indicate the [**appliance**] was not replaced with a different unit (are assigned the full savings (scenario B below). The "keeper" respondents that indicate the [**appliance**] was replaced with a different unit and the replacement is determined to have been induced by the program are assigned the replacement TRM-based deemed savings values. This is typically a small percentage of participants, however, as the incentive usually covers only a very small percentage of the incremental cost of purchasing a new unit. The following set of questions helps to further classify the net replacement savings based on the replacement type of unit.

4. Did you replace the removed [**appliance**] with another one?
   c. Yes
   d. No

5. Is the replacement [**appliance**] an ENERGY STAR or high efficiency model?
   a. Yes
   b. No

6. Was this replacement [**appliance**] brand new or used?
   a. Brand New
   b. Used

As previously mentioned, the proportion we expect to have been induced is very small, so it is critical that the respondents answer is repeated to them and asked again, clarifying questions include:[4]

7. Would you have purchased your replacement [**appliance**] if the recycling program had not been available?

8. I would like to confirm your answer, are you saying that you chose to purchase a new appliance because of the appliance recycling program, or are you saying that you would have purchased the new [**appliance**] regardless of the program?

If the respondent confirms that they would have purchased the new [appliance] regardless of the program, then by definition they cannot have been induced to purchase by the program and are not classified as scenario A (induced replacement). At this point we still need to determine what they would have done with the old unit. You can then ask if they would have kept it (classified as scenario B) or removed it (continue on to next section regarding "remover"). If the respondent confirms that they would not have purchased the replacement unit without the program then they are considered an induced replacement and get the appropriate replacement savings value from the TRM.

1.1.3. Classifying the status of the "Remover"

The "remover" segment, as discussed previously, would be assigned to those who answered question 2 above as "a. remove it". The remover segment is potentially qualified as free-riders, assuming in absence of the program their appliance would have been removed from service. These respondents should be further segmented into groups based on whether the unit would have continued operating or been permanently removed from service – this helps define the deemed savings values to be assigned to them. The following questions are an example of what to ask:

9. If the appliance pickup program was not available, which one of the following alternatives would you have most likely done with your [**appliance**] when you were ready to dispose of it? Would you have:
   a. Sold it
   b. Given it away for free
   c. Had it removed by the dealer you got your replacement [**appliance**] from
   d. Took it to a dump or recycling center
   e. Hired someone else to haul it away

Ask the following question if the answer to question 9 above is "a – Sold it":

10. You said you would have most likely sold your [**appliance**]. Would you have sold the [**appliance**] to an appliance dealer, or to a private party (like a friend, relative or by running an ad)?
    a. Dealer
    b. Private party (friend, relative, or by running ad)

If the anticipated plan was to sell the unit to a dealer, then the age of the unit needs to be discovered. Ask the following question if the answer to question 9 above is "a. Dealer":

11. You said you would have most likely sold your [**appliance**] to a Dealer. Was your [**appliance**] less than 10 years old?
    a. Yes, less than 10 years old
    b. No, at least 10 years old

We can assume that operable units less than 10 years old (answer "a. Yes, less than 10 years old" to question 11) are likely to be resold on the open market (qualified as scenario C), whereas units at least 10 years old (answer "b. No, at least 10 years old" to question 11) are likely to be removed from service (scenario D) since there is little probability of them having any retail value greater than the cost of attempting to sell the unit.[5]

Ask the following question if the answer to question 9 above is "b – Given it away for free":

12. You said you would have most likely given away your [**appliance**]. Would you have given it to a private party (like a friend, relative or by running an ad), or to a charitable organization?
    a. Private party (friend, relative or by running an ad)
    b. Charitable organization

Ask the following question if the answer to question 9 above is "d – Took it to a dump or recycling center":

13. You said you would have most likely taken away the [**appliance**] yourself. Would you have taken it to a dump, or to a recycling center?
    a. Dump
    b. Recycling Center

If the respondent was planning on transferring the unit by selling (if unit is less than 10 years old) or giving it away (answers a or b for question 9), then we can assume the unit would likely continue operating and therefore the respondents are not classified as free-riders. The savings attributable to these participants are the most difficult to estimate, because this scenario (Scenario C) is attempting to estimate what the prospective buyer of the used appliance did in absence of finding the program recycled unit in the market place (i.e., the program took the unit off the grid, so the prospective purchaser faced, in theory, a smaller market of used refrigerators). The UMP uses, and this guideline recommends in absence of primary data collection, a composite value for this scenario, assuming half of the respondents would receive full savings (assuming unit would have served as secondary unit for a different household), a quarter of the respondents receive the delta between a new and old unit (non-ENERGY STAR), and the remaining quarter of the respondents receive zero savings (assuming different household was able to find alternative similar old unit).

If the respondent was planning on removing the unit from service, either through recycling it, hauling it to the dump, or hiring someone to remove it (answers c, d, e to question 9), or if they

[5] Note that the 10 year age cutoff for resale value was derived from the following study: Navigant Consulting, January 22, 2013: *Energy Efficiency/Demand Response Plan: Plan Year 4 Evaluation Report: Residential Fridge and Freezer Recycle Rewards Program*; Prepared for Commonwealth Edison Company

planned on giving it to a retailer but the unit was at least 10 years old, then they are classified as full free-riders and not allocated any savings (scenario D, net savings = 0).[6] One final consideration with respect to the free rider scenario is the availability of disposal options in the service area in question. Evaluators may want to include viability/logistics of alternative options (whether there is even possibility of this service in participant's area) in advance of fielding the survey. If it is discovered that no such option exists, then additional options need to account for alternative possibilities in the survey.

1.1.4. Estimating Net Savings

Net savings should be assigned individually to each respondent based on the responses provided to the questions outlined above. The net savings should be averaged across all respondents to calculate program level net savings. Table 1 demonstrates the proportion of a sample population that are classified into each of the potential seven categories and the resulting weighted net savings.

Table 1: Net Savings Example for a Sample Population[7]

Primary Classification	Secondary Classification	Replacement TRM value	Population (%)	UEC (kWh) w/out Program	UEC (kWh) w/ Program	kWh Savings
Would have kept unit	Scenario A: Induced Replacement	Non-ES unit	3%	1,026	520	506
	Scenario A: Induced Replacement	ES unit	2%	1,026	404	622
	Scenario B: Secondary unit w/out replacement	No replacement	25%	1,026	0	1,026
Would have removed unit	Scenario D: Removed from service	No replacement	20%	0	0	0
	Scenario C: Transferred	No replacement or unit age >= 10 years	12.5%	0	0	0
		Non-ES unit, unit age < 10 years	12.5%	1,026	520	506
		No replacement	25. 0%	1,026	0	1,026
Net Savings (kWh)						**604**

6 Note Scenario D assumes that the retailers that picked up the unit would have discarded the unit, rather than selling it on the secondary market.

7 The percent values presented in this table are just examples; actual research should be conducted to determine the percentage of units that fall into each of these categories. Note that the UEC values presented in the table represent the default values from the 2014 TRM, factoring in part-use.

1.1.5. Data Sources

A Participant Survey, comprised of a random sample of participants, should be the primary source of data collected for estimating net-to-gross for the appliance recycling program. Per the UMP, a secondary source of supporting data may come from a non-participant sample. Non-participants do not have the same perceived response bias as participants, and can help offset some of this potential bias in estimating the true proportion of the population that would have recycled their unit in absence of the program. To maintain consistency with the UMP, we recommend averaging the results of the non-participant survey with that of the participant survey. The use of a non-participant sample is recommended but not required given budget and time considerations.

Appendix F. Common Approach for Measuring Free-Riders for Downstream Programs

See memo on the following pages.

research into action inc

Research Into Action, Inc.
PO Box 12312
Portland, OR 97212

www.researchintoaction.com

503 287 9136 (main)
503 281 7375 (fax)
888 492 9100 (toll free)

Memorandum

To: The PA EDCs, TUS, and SWE Team

From: Jane Peters and Ryan Bliss, Research Into Action

Date: October 4, 2013 (revised November 12, December 4, and December 23, 2013)

Re: Common Approach for Measuring Free-riders for Downstream Programs

The PA PUC Implementation Order specifies that the net-to-gross ratio (NTG) for Phase II of Act 129 is to be treated in the same way as for Phase I. Specifically, for compliance purposes the NTG ratios for Phase II programs continues to be set a 1.0 – basing compliance with energy and demand reduction targets on gross verified savings. However, the PUC order also states that the EDCs should continue to use net verified savings to inform program design and implementation.

There are two reasons to consider having a uniform NTG approach for the EDCs. One is that if NTG measurement for a program is consistent across time, comparisons of the NTG metric across time will be reliable and a comparisons therefore are valid. If the NTG metric is measured the same way every year or every quarter, program staff can use the NTG metric to inform their thinking because it provides a consistent metric over time. Of course, programs often change across years: measures may be added or taken away, and rebate amount or technical services may vary, consistent measurement of NTG is even more valuable in these situations because it permits better understanding of how the changes affect NTG.

The second reason to consider having a uniform NTG approach for the EDCs is the value that can be obtained from comparisons across utilities. Just as programs change year to year, it is clear that the programs offered by the EDCs vary from each other. When there are different metrics, no one can discern whether different NTG values are due to program differences, external differences, or differences in the metric. By using a consistent metric, we can at least rule out the latter.

The variability in the types of services/measures offered by the programs, the different delivery strategies, and the variability of the customer projects themselves makes it necessary to tailor the attribution assessment appropriately. The need for comparability of results between years and between EDCs, however, requires a consistent overall approach to assess attribution. The challenge is in allowing flexibility/customization in application yet still maintaining a consistent approach.

1. Sources for Free-ridership and Spillover Protocols

Under the Uniform Methods Project (UMP) funded by DOE, The Cadmus Group and its subcontractors are in the process of developing a framework and a set of protocols for determining the energy savings from specific energy efficiency measures and programs. The

Phase I report, published in April 2013, outlines methods for evaluating gross energy savings for common residential and commercial measures offered in ratepayer-funded initiatives in the United Sates.[1] In Phase II, The UMP is planning to develop a crosscutting protocol for determining NTG. However, according to the Phase I report, variation in definitions of net savings (for example, whether it includes participant and/or nonparticipant spillover) and policies regarding NTG vary across jurisdictions. This means that the UMP protocol, due to be published in 2014, will not offer specific recommendations on how NTG is applied nor will it offer specific survey questions and analysis techniques.

While the UMP may be a future resource for guidance on assessing NTG, the *Framework* provides the following general guidance as a good starting place for assessing free-ridership and spillover. Overall, the "SWE recommends that EDCs use standard sampling techniques, data collection approaches, survey questions, survey instruments, and analysis methodology for NTG studies."[2]

Furthermore, the SWE recommends standardization – at a minimum within the EDCs' measurement activities and ideally across all EDCs – for provision of consistency in explaining program effects. Among several free-ridership methods mentioned, the SWE recommends an approach similar to that chosen by the Energy Trust, which uses a concise battery of questions to assess *intention* and *program influence*, which is the focus of the rest of this memo.

The *Framework* also defines participant and nonparticipant spillover and recommends the consideration of trade ally surveys and reports for assessing the nonparticipant portion of a programs spillover impact. A future memo will discuss measuring participant and nonparticipant spillover.

2. Sampling

The sampling approach for estimating free-riders should use confidence and precision levels at least equivalent to the approach for gross savings being estimated for a specific program. The SWE further encourages sampling and reporting free ridership and spillover at the end-use level or stratifying for high-impact end-uses in much the same way as for gross savings estimates whenever possible. EDCs are encouraged to use higher confidence and precision levels, and to conduct the sampling at the measure level when more detailed information is needed for program assessment.

---

1 The Uniform Methods Project: Methods for Determining Energy Efficiency Savings for Specific Measures, January 2012 – March 2013, NREL/SR-7A30-53827 published April 2013 by The Cadmus Group, contract no. DE-AC-08GO28308 and found at http://www.nrel.gov/docs/fy13osti/53827.pdf

2 SWE Team. *Evaluation Framework for Pennsylvania Act 129 Phase II Energy Efficiency and Conservation Programs*. June 30, 2013. Page 61

3. Recommended Standard Free-Ridership Protocol

The following discussion presents a standard, yet flexible, approach to assessing free-ridership for the EDCs to use during Phase II. This method applies to downstream programs, typically using some incentive or direct installation.[3] Research Into Action and Energy Trust of Oregon developed this approach for telephone and on-site assessment of NTG (by project and by measure) across residential, commercial, industrial, and government sectors including:

- Rebates and grants for energy efficiency improvements
- Rebates and grants for renewable energy sources
- Technical assistance
- Education and outreach

The assessment battery is brief to avoid survey burden yet seeks to reduce self-report biases by including two components of free-ridership: 1) *intention* to carry out the energy efficient project without program funds; and 2) *influence* of the program in the decision to carry out the energy efficient project. When scored, each component has a value ranging from zero to 50, and a combined total FR score that ranges from zero to 100. These components are potentially subject to different and opposing biases: as a result, the intention component typically indicates higher free-ridership than the influence component. Therefore, combining them decreases the biases.[4]

In the following subsections, we describe a Common Method for a standard retrofit incentive program, including both the question battery and scoring. We describe how the Common Method can be adapted for different types or variations of program or measure types (e.g., EDC direct install and custom programs). We finally address several questions and concerns that EDCs and their evaluation contractors raised in response to earlier versions of this memo.

3.1. Intention

Intention is assessed through a few brief questions used to determine how the upgrade or equipment replacement likely would have differed if the respondent had not received the program assistance. The initial question asks the respondent to identify of a limited set of options that best describe what most likely would have occurred without the program assistance. Note that "program assistance" often includes more than just the incentive or rebate – it may also include audits, technical assistance, and the like.

---

[3] At the November 2013 PEG meeting, the SWE offered a memo on NTG approach for Appliance Recycling Programs based on the Uniform Method Project. Finally, when self-report questions are used for upstream and mid-stream programs those questions should use the same structure described herein. However, self-report methods are typically insufficient and additional data sources should be used but are not prescribed at this time.

[4] See Section 2.4.4 for detailed discussion.

The offered response options (typically four or five, and preferably no more than six) capture the following general outcomes:

- Would have canceled or postponed the project, upgrade, purchase, etc., beyond the current program cycle (typically at least one year).
- Would have done something that would have produced savings, but not as much as those achieved through the upgrade or equipment replacement as implemented.
- Would have done the upgrade or equipment replacement as implemented.
- Don't know

The first outcome (canceled or postponed beyond the program cycle) indicates zero free-ridership and thus results in a score of 0. The second option indicates some free-ridership, but not total free-ridership (a score ranging from 12.5 to 37.5 for the *intention* component). The level of free-ridership depends on two factors: a) the level of savings that the respondent would have achieved without the program's assistance; and b) in the case of nonresidential programs, whether the respondent's business or organization would have paid the entire cost of the equipment replacement or upgrade without the program assistance. The third outcome (done project as implemented) indicates total free-ridership (a score of 50 for the *intention* component).

In previous implementations of this approach, "don't know" responses to this question were assigned the midpoint score of 25 for the *intention* component. Alternative treatments that have been proposed for "don't know" responses are to assign the mean of non-missing responses or to exclude the case and replace it with another. Both those treatments may be problematic, as they assume that "don't know" responders are otherwise similar to the rest of the sample, when there may be reasons for the "don't know" response that make them dissimilar.[5] Generally, imputing the mean for missing responses is not considered best practice.[6]

We recognize that imputing the midpoint may be considered arbitrary (but see *Section 5.4, Treatment of "Don't Know" Responses*). Moreover, our experience is that "don't know" responses are infrequent, and so the way in which they are handled likely will not have a great impact on the resulting free-ridership estimates. Evaluators may implement alternative approaches to handling "don't know" responses in addition to assigning the midpoint and report both results. As an alternative approach, we recommend using linear regression to predict the *intention* score from each respondent's *influence* score.

As discussed below, the assessment of the above factors will depend somewhat on the nature of the program, but the overall approach is guided by several considerations:

- The instrument should be as brief as possible to avoid survey burden.

---

5 *Section 5.4, Treatment of "Don't Know" Responses*, discusses evidence for this.

6 Enders, C.K. *Applied Missing Data Analysis*, New York: The Guilford Press, 2010.

- Challenging a respondent's consistency can make the respondent feel defensive and may not produce more accurate data[7] – therefore, the instrument should avoid overt "consistency checks."
- The instrument should recognize the limits of reporting a counterfactual, particularly in assessing cases in which respondents that report they would have saved some, but less, energy without the program.

Any tailoring of the approach should take the above considerations into account.

The following subsections describe, in turn, how *intention* typically has been assessed with the Common Method in nonresidential and residential programs and how it can be further tailored if needed.

3.1.1. Assessment of Intention in Nonresidential Programs

In this section, we describe how the Common Method typically is applied and scored in standard, nonresidential incentive programs. We also discuss tailoring or modification of the Common Method.

3.1.1.1. General Application of Intention Assessment in Nonresidential Programs

Typically, the nonresidential battery begins with the following question:

- *Which of the following is most likely what would have happened if you had not received [the program assistance]?*

The battery has included the following options in multiple evaluations of a wide range of nonresidential programs:

- Canceled or postponed the project at least one year
- Reduced the size, scope, or efficiency of the project
- Done the exact same project
- Don't know

Respondents that select the second option are asked:

- *By how much would you have reduced the size, scope, or efficiency? Would you say…*
  - *a. a small amount,*
  - *b. a moderate amount, or*

[7] See *Section 5.5* for a more detailed discussion.

c. a large amount

Note that the intent is *not* to separately assess reduction in size, scope, *and* efficiency – it is simply to assess whether, in the respondent's opinion, in absence of the program the project would have been reduced in size, scope, or efficiency by a small, moderate, or large amount. Under the above assumption that that a precise estimate of counterfactual savings is not likely to be achievable, this approach makes no effort to establish such an estimate. Instead, the approach simply attempts to obtain the respondent's best general estimate of the counterfactual.

In response to the initial draft of this memo, some evaluators have noted that a small, moderate, or large reduction in a given project's size would not necessarily have the same energy impact as a small, moderate, or large reduction in the project's scope or the efficiency level of the equipment used. This is understood, but the purpose is to balance the desire to obtain some estimate of savings reduction with the desire to avoid response burden and reduce the risk of false precision.

Nevertheless, evaluators may propose alternative response options. The SWE requests that those evaluators provide their rationale for such alternatives.

Respondents that report they would have done exactly the same project without the program's assistance are asked:

› *Would your business have paid the entire cost of the upgrade?*

This question is used to help mitigate a bias to overstate the likelihood that the respondent would have done the same project without program assistance.[8] Respondents get the highest free-rider score *only* if they report that they would have done the same project without program assistance and that their business would have paid the entire cost. Otherwise, a lower free-rider score is assigned, as shown below.

It is important to note that the above question is not a consistency check. That is, respondents who report they would have done the same project without program assistance but do not confirm that their business would have paid the entire cost are *not* confronted with the apparent inconsistency and asked to resolve it. Nor does the method assume that the second response is the correct one. Instead, the method assumes that neither response provides the full picture and that further questioning could not *reliably* provide the complete picture. The method thus assigns a free-rider value that is intermediate to both: that is, it assumes that the best estimate is that the project would have produced some savings but not as much as were actually produced through the program.

3.1.1.2. Scoring of Intention Assessment in Nonresidential Programs

An *intention* free-ridership score of 0 to 50 is assigned as follows:

[8] See *Section 5.2*, *Controlling for "Socially Acceptable" Response Bias*, for a more complete discussion of this potential bias.

- A project that would have been canceled or postponed beyond the program cycle is assigned an intention score of 0.
- A project that would have been done exactly as it actually was done, with the cost born entirely by the respondent's business or organization, is assigned an intention score of 50.
- A project that would have resulted in fewer savings than the project actually done is assigned an intermediate score based on the responses to the applicable follow-up question(s).

Interviewers (or web surveys) should make reasonable attempts to get a response to the questions. If respondents cannot select an option, "don't know" responses are assigned a score that represents the midpoint of the range of possible values for that question (as illustrated below).[9]

Table 1 summarizes the possible response combinations to the questions described above and the *intention* score assigned to each unique combination.

Table 1. General Free-Ridership Intention Component Scoring

QUESTION	RESPONSE	INTENTION SCORE
1. Which of the following is most likely what would have happened if you had not received [the program assistance]?	Postponed / cancelled	0
	Reduced size, scope, efficiency	Based on response to Q2
	No change	Based on response to Q3
	Don't know	25*,**
2. By how much would you have reduced the size, scope, or efficiency?	Small amount	37.5
	Moderate amount	25
	Large amount	12.5
	Don't know	25*
3. Would your business have paid the entire cost of the upgrade?	Yes	50
	Don't know	37.5*
	No	25**

* Represents the midpoint of possible values for this question.

** Infrequent response.

3.1.1.3. Tailoring of Intention Assessment in Nonresidential Programs

The above approach has been used to assess *intention* with a range of retrofit incentive programs. Evaluators may propose other modifications as needed, but such modifications should be

[9] *Section 5.4, Treatment of "Don't Know" Responses*, discusses the rationale for this treatment of "don't know" responses rather than alternatives, such as assigning a mean value. In fact, "don't know" responses are infrequent.

informed by the general principles described above, of keeping the instrument brief, recognizing the limits of counterfactual questioning, and avoiding consistency checks.

Tailoring of Question Wording

The specific wording of the questions and the response options provided should be tailored to the specific program, measure type, or sample group. As indicated above, the general form of the initial *intention* question is: "Which of the following is most likely what would have happened if you had not received [the program assistance]?" Therefore, it is important to identify the primary type or types of program assistance that are considered important in reducing the key barriers to carrying out the targeted behavior (e.g., an upgrade to more energy efficient equipment). In other words, it is important to clearly indicate what participating in the program meant and what program they were participating in.

Example: A program operated through a State agency helped businesses obtain contracts with an ESCO to finance efficiency upgrades. In this case, the "intention" question was:

> "What do you think your organization most likely would have done if the [Name of Office] had not helped you obtain the contract with an ESCO like ...?"

As noted above, the "influence" question should include the range of program elements or services. Evaluators should be careful not to ask about services that a particular program does not provide. For example, it would be confusing to ask how influential the rebate was if there was no rebate attributable to the program/measure. Logic models, program theory, and staff interviews typically inform the list of program elements to ask about.

Tailoring of Response Options

As noted above, one area in particular where modification may be proposed is in the specification of equipment replacement or upgrade alternatives to identify differing levels of counterfactual energy savings (i.e., in place of asking whether the respondent would have done something that reduced energy a small, moderate, or large amount). In such cases, the counterfactuals options should reflect the likely range of activities that likely would have occurred absent program assistance, with points assigned to reflect the amount of energy savings each would provide.

For example, the following alternatives could be specified for a lighting program that incents LEDs:

1. Put off replacing the [X type of] lights with LEDs for at least one year or cancelled it altogether.
2. Kept some of the existing lights and replaced some lights with LEDs.
3. Installed different lights. If so, what kind? _____
4. Installed the same number and type of LED lights anyway.

5. Done something else. If so, what? _____
6. Don' Know or no answer

Follow-up questions are needed for some responses. In this case, for respondents who report they would have installed fewer lights, a follow-up question is needed to assess the savings reduction – specifically, what percentage of lights would they have replaced with LEDs? For respondents who said they would install the same number, a follow-up question should be used to verify that the respondent would have paid the entire cost without program support.

Other Tailoring or Modifications

In response to the initial draft of this memo, some additional types of modifications have been suggested:

- Preceding the initial counterfactual question with one asking whether the respondent had already carried out the equipment replacement or upgrade before applying for the incentive. Evaluators may include such a question but should still ask the counterfactual question as described above.
- Specifying the value of each respondent's incentive in the initial counterfactual question. This is acceptable, but evaluators should keep in mind that the incentive often is not the only program assistance received and other program assistance may also have had a role in driving the project. So, for example, the question may refer to "the incentive of $X and other assistance, such as identification of savings opportunities."

We provide further discussion of tailoring the general free-ridership approach for programs other than standard retrofit type programs in *Section 4, Tailoring the Free-ridership Approach*, below.

3.1.2. Assessment of Intention in Residential Programs

The assessment of *intention* for residential programs is similar to that for nonresidential programs. However, the response option "reduced the size, scope, or efficiency of the project" is not likely to be as meaningful to a residential respondent as to a nonresidential one, nor is a residential respondent expected to be able to estimate whether the reduction would be small, moderate, or large. Evaluators, rather, should attempt to provide a list of meaningful counterfactual options.

Table 5 shows examples of counterfactual response options used with three types of residential measures: appliances, air or duct sealing or insulation, and windows. As this shows, the goal is to cover the range of likely alternatives to carrying out the incented upgrade, with intention scores that reflect the degree of free-ridership. Reporting an alternative that likely would have produced no energy savings results in a score of 0; reporting something that likely would have produced some energy savings, but lower savings than the incented upgrade or purchase results in an intermediate score of .25; and reporting the same outcome as the incented upgrade or purchase results in a score of .5.

Table 2. Example Counterfactual Response Options for Various Residential Measure Types

PROGRAM	COUNTERFACTUAL RESPONSES	INTENTION SCORE
Appliance	Cancel/postpone purchase	0
	Repair old appliance	0
	Buy used appliance	0
	Purchase less expensive appliance	.25
	Purchase less energy efficient appliance	.25
	Purchase same appliance without the rebate	.5
	Don't know	.25
Air/Duct Sealing, Insulation	Cancel/postpone	0
	Do by self (if program incents only contractor-installation)	.25
	Reduce amount of sealing/insulation	.25
	Have the same level of sealing/insulation done without the rebate	.5
	Don't know	.25
Windows	Cancel/postpone purchase	0
	Replace fewer windows	.25
	Purchase less expensive windows	.25
	Purchase less energy efficient windows	.25
	Do same window replacement without the rebate	.5
	Don't know	.25

A difference from the nonresidential instrument is that, respondents that report they would have done the same thing without the incentive are not then asked whether they would have paid the cost of the upgrade. A question that may seem perfectly reasonable in the context of a decision about allocating a business's resources may not seem reasonable in the context of personal decisions. Instead, the "would have done the same thing" response may include the words "without the rebate [or incentive]."

Issues relating to tailoring the *intention* component are the same as for nonresidential assessments (see *Section 3.1.1.3*, above).

3.2. Influence (Nonresidential and Residential)

Assessing program influence is the same for nonresidential and residential programs.

Program influence may be assessed by asking the respondent how much influence – from 1 (no influence) to 5 (great influence) – various program elements had on the decision to do the project the way it was done.

The number of elements included will vary depending on program design. Logic models, program theory, and staff interviews typically inform the list. Among the more typical elements programs use to influence customer decision making include: information; incentives or rebates; interaction with program staff (technical assistance); interaction with program proxies, such as members of a trade ally network; building audits or assessments; and financing.

The program's influence score is equal to the maximum influence rating for any program element rather than, say, the mean influence rating. The rationale is that if any given program element had a great influence on the respondent's decision, then the program itself had a great influence, even if other elements had less influence.

Table 3. General Free-Ridership Influence Component

Calculation of the Influence Score is demonstrated in the following example:

- Rate influence of program elements.

	Not at all influential				*Extremely influential*		
– Incentive	1	2	**3**	4	5	DK	NA
– Program staff	1	2	3	**4**	5	DK	NA
– Audit/study	1	2	3	4	**5**	DK	NA
– Marketing	1	**2**	3	4	5	DK	NA
– Etc.	1	2	3	4	5	DK	NA

In this example the highest score (a '5' for the influence of the audit/study) is used to assign the influence component of the FR score. High program influence and FR have an inverse relationship – the greater the program influence, the lower the free-ridership, as see in Table 3.

Table 4. General Free-Ridership Influence Component Scoring

PROGRAM INFLUENCE RATING	INFLUENCE SCORE
1 – not at all influential	50
2	37.5
3	25
4	12.5
5 – extremely influential	0
DK	25

3.3. Total Free-ridership Score

Total free-ridership is the sum of the *intention* and *influence* components, resulting in a score ranging from 0 to 100. This score is multiplied by .01 to convert it into a proportion for application to gross savings values.

4. Applying the Common Method to Other Program Types

Evaluators should be able to use the Common Method, described above, with most retrofit incentive programs. Evaluators may tailor the approach for use with programs that do not fit the general retrofit incentive mold.

In programs where the primary program approach is to provide assistance (e.g., rebate/incentive, technical assistance, direct install) to the program participant to reduce barriers to undertaking energy efficient upgrades or improvements, it typically should be sufficient to tailor question wording and response options while maintaining the overall approach. In such cases, the *intention* component may require more tailoring than the *influence* component.

In programs that must influence multiple actors to achieve the desired outcomes or carry out their influence through more complex forms of assistance, it may be necessary to tailor the method more extensively or to propose an alternative approach. Section 5.1 discusses the process for proposing methods in the above cases.

The following examples show how the method has been applied for some programs that do not fit the standard retrofit incentive model. The purpose of these examples is not to show the only possible ways in which the Common Method may be modified to use with different program types, but are here for illustrative purposes. EDCs and their evaluators should propose an approach that is consistent with the considerations outlined in *Section 3.1*, above.

The first example illustrates a case for which the modification is relatively simple; the second example illustrates a more complex case requiring more extensive modification.

4.1. Direct Install (DI) Program

Direct install (DI) programs are different from most programs in that the program is offered directly to potential participants via program representatives. In applying the Common Method to a DI program, the battery sought to verify whether the respondent was even considering the directly installed measure(s) prior to program contact. Where the respondent was not even considering the measures before being contacted by the program, the total free-ridership score was set to 0 (i.e, both the intention and influence scores were 0). For respondents who were planning an upgrade, the method mirrors the general approach described above.

Assessment of program influence was as described above, but included potential program influences reflecting the unique elements of the DI program. For example, in a case where the program included a building assessment along with DI measures, the influence question included

"assessment results," along with "interactions with the assessor or contractor," and "the fact that the measure was free."

4.2. Financing an Energy Performance Contract (EPC)

Some programs will require more extensive and *ad hoc* tailoring of the Common Method, such as when a program works with third-party entities to assists with project financing. In one example, a program helped building owners establish and implement energy performance contracts (EPCs) with program-administrator-approved energy service companies (ESCOs). Since the program administrator worked with both the building owner and the ESCO, neither alone could accurately describe what would have happened without the assistance. Therefore, for each sampled project, the evaluators surveyed both the building owner and the ESCO.

The building owner instrument included the standard *intention* question of what would have happened (postpone/cancel, smaller project, same upgrade) without program support and the standard "influence" question.[10] The evaluators calculated building owner *intention* and *influence* following the standard approach, described above.

The instrument for ESCOs asked:

- How likely they would have known about the client without the program's assistance.
- What likely would have happened without the program's assistance (same EPC, lower-savings EPC, no EPC).

The evaluators calculated only ESCO intention, using the algorithm shown in Table 6.

Table 5. Algorithm for ESCO Intention Score

WOULD LIKELY HAVE KNOWN ABOUT CLIENT	COUNTERFACTUAL	INTENTION SCORE
Yes, likely would have known about client's needs without program assistance	Same EPC	50
	Lower-savings EPC	25
	No EPC	0
No, likely would not have known about client's needs without program assistance	N/A	0

To aid in determining how to combine the building owner and ESCO scores, the building owner instrument also asked:

- Whether they had ever worked with an ESCO before.
- Whether they would have used an ESCO without program assistance.

[10] Influencers were program information, interaction with program staff, the list of prequalified ESCOs, and program assistance in selecting an ESCO.

The evaluators used the algorithm shown in Table 7 to calculate the intention component score based on responses by both the building owner and the ESCO. The algorithm assumed that the ESCO responses were not relevant if: 1) the building owner was experienced with ESCOs and so could accurate predict what would have happened without the program assistance; 2) the owner indicated that without program assistance they would have cancelled or postponed the project or would not have used an ESCO.

Table 6. Algorithm for Combining Building Owner and ESCO Intention Score

Would Have Used ESCO?	Bldg. Owner experienced with ESCO	ESCO responses considered?	Bldg. Owner Response to Intention Questions	ESCO Response to Intention Questions	Final intention score
No/DK	N/A	No[1]	Free-rider, Partial or Not Free-rider	N/A	Client score
Yes	Yes	No[2]			
Yes	No	Yes	Free-rider (would have done same project)	Free-rider	50
				Partial free-rider	37.5
				Not free-rider	25
			Partial Free-rider (would have done less efficient project)	Free-rider	25
				Partial free-rider	25
				Not free-rider	12.5
		No[3]	Not Free-rider (would have cancelled or postponed)	N/A	0

1 Since the building owner would not have used an ESCO without program assistance, ESCO responses are not relevant.

2 Since the building owner was experienced with ESCOs, it was assumed that they could accurately predict what would have happened without program assistance.

3 Since the building owner indicated they would have cancelled or postponed the project without program assistance, the ESCO responses are not relevant.

In other cases, where there may be reason to question the building owner's ability to provide an accurate intention response, then the ESCO's response was also considered and could be used to adjust the building owner's score.

5. Response to Questions and Concerns Raised About the Common Method

In response to the initial and revised drafts of this document, some evaluators raised questions or concerns concerning the Common Method described above. We have revised the above sections to address those concerns. We also provide additional information and clarification here in reference to specific questions or concerns raised.

5.1. Inapplicability to Some Program Types

By January 2, 2014, EDCs should submit a list of all programs to the SWE, identifying which programs:

- Can use the Common Method or the Appliance Recycling Program (ARP) Method
- May be able to use the Common Method or the ARP Method
- Cannot use the Common Method or the ARP Method

The SWE will establish a working group to review and provide feedback on the proposed approaches. EDCs will use the working group's feedback to finalize their proposed approaches.

5.2. Controlling for "Socially Acceptable" Response Bias

One concern is that respondents' self-reports are likely to be tainted by a bias toward reporting that they would have done the energy-saving project even without the program. This assumption has variously been ascribed to a "social desirability" bias (where energy conservation is the "socially desirable" response) or to an attribution bias (in which we tend to make internal attributions for "good" decisions or outcomes and external attributions for poor ones).

Above, we argued that the two components of free-ridership that the battery assesses – *intention* to carry out the energy efficient project and *influence* of the program – are likely subject to different and opposing biases, which are at least partly canceled out by combining the components. While the *intention* component is subject to biases that would increase the estimate of free-ridership, the *influence* component may be subject to biases that would decrease the estimate of free-ridership. Specifically, rated influence may reflect satisfaction with the program such that participants who are satisfied with the program may report greater program influence. If so, a program with high participant satisfaction may appear to have lower free-ridership on that basis.

Analysis of responses to the battery tend to support the above suppositions. We analyzed responses to the battery from 158 participants in nonresidential retrofit and new construction programs and 1,252 participants in a range of residential programs (appliances, shell measures, home performance, and refrigerator recycling).[11] First, the two components positively correlated in both the nonresidential and residential samples (.40 and .37, respectively), indicating shared measurement variance. However, the *intention* component yielded higher mean scores than did the *influence* component for both the nonresidential (95% CI: 16.8 ± 3.4 vs. 5.3 ± 1.5) and residential (95% CI: 26.4 ± 1.3 vs. 10.5 ± 0.8) samples. If the shared variance between the two components indicates they are both measuring free-ridership, these findings are consistent with

[11] The responses were collected in May through July of 2010, as part of the evaluation of roll-out of the Energy Trust Fast Method for collecting participant feedback. *Fast Feedback Program Rollout: Nonresidential & Residential Program Portfolio*. Submitted to Energy Trust of Oregon by Research Into Action, Inc., December 31, 2010.

the idea that *intention* may over-estimate free-ridership and *influence* may under-estimate it. Absent any compelling evidence that one of these components by itself yields a truer estimate of free-ridership, it is safest to conclude that combining them provides the best assessment.

5.3. *Intention* Counterfactual Indicates Reduced Energy Savings

As described above in Section 2.2.1, the Common Method provides three counterfactual options: 1) the upgrade would have been canceled or postponed at least one year; 2) the upgrade's size, scope, or efficiency would have been reduced; and 3) the same upgrade would have been done. Respondents who report a reduction in size, scope, or efficiency are then asked whether the reduction would be small, moderate, or large.

Three questions have been raised about the treatment of a reported reduction in size, scope, or efficiency:

- *Does the method ask separately about the reduction in size, in scope, and in efficiency and, if so, how does it combine or weight the responses?*
- *Does the Common Method allow for asking about specific changes in size, scope, or efficiency? For example, in the case of a lighting project, could the instrument ask if the respondent would have installed different kinds of lights and, if so, what kind?*
- *If the Common Method allows for asking about specific changes in size, scope, or efficiency, how should the response be scored if the respondent does not provide enough information to determine a counterfactual difference in energy savings?*

The underlying concern is whether the approach is capable of accurately capturing the difference in energy savings between the project-as-implemented and the counterfactual case where some energy savings would have been achieved.

As noted in *Section 3.1.1.1*, above, the intent is *not* to separately assess reduction in size, scope, *and* efficiency – it is simply to assess whether, in the respondent's opinion, in absence of the program the project would have been reduced in size, scope, or efficiency by a small, moderate, or large amount. Under the assumption that that a precise estimate of counterfactual savings is not likely to be achievable, this approach makes no effort to establish such an estimate. Instead, the approach simply attempts to obtain the respondent's best general estimate of the counterfactual.

It is understood that a small, moderate, or large reduction in a given project's size would not necessarily have the same energy impact as a small, moderate, or large reduction in the project's scope or the efficiency level of the equipment used. The purpose is to balance the desire to obtain some estimate of savings reduction with the desire to avoid response burden and reduce the risk of false precision.

Nevertheless, evaluators may propose alternative response options. In the event that the respondent does not provide enough information to determine a counterfactual difference in

energy savings, the recommended approach is to assign the midpoint value of 25. However, evaluators may also propose an alternative approach. The SWE requests that those evaluators provide their rationale for such alternatives.

5.4. Treatment of "Don't Know" Responses

As described above, in the case of "don't know" responses to one of the free-ridership questions, the Common Method assigns the appropriate midpoint score. For example, if a respondent cannot provide any response to the main counterfactual question for the *intention* component, the method assigns the midpoint value of 25 for that component.

One objection raised was that assigning a midpoint value will inflate the free-ridership estimate in cases where mean free-ridership is less than 50%. For example, Section 2.4.2 (*Controlling for "Socially Acceptable" Response Bias*), showed a mean *intention* value of 16.8 for nonresidential programs. If the midpoint value of 25, rather than the mean of 16.8, is substituted for a "don't know" response to the *intention* component, the resulting total free-ridership value will be inflated.

A proposed alternative to imputing the mean of non-missing responses is to exclude cases with "don't know" responses and replace them with another. Both those treatments may be problematic, as they assume that "don't know" responders are otherwise similar to the rest of the sample. However, the mere fact that they could not answer the *intention* counterfactual suggests they may differ from other respondents in some important respects that might affect their overall free-ridership level. Generally, imputing the mean for missing responses is not considered best practice.[12]

We could not use the nonresidential data described above to reliably investigate the question of whether "don't know" responders differ from others, as only three nonresidential respondents (2% of the sample of 158) gave a "don't know" response to the *intention* question. However, in the residential dataset, 70 respondents (6% of the sample of 1,252) gave "don't know" responses.[13]

We therefore investigated whether respondents that had *intention* "don't know" responses differed from other respondents on the *influence* component of the free ridership battery. On average, respondents that gave an *intention* response (n = 1,164) indicated a maximum program influence of 4.4 on a 1-to-5 scale, while those that gave an *intention* "don't know" response (n = 70) indicated a maximum program influence of 4.1. This difference was marginally significant (F = 3.2, p = .07). While this finding does not conclusively show that "don't know" respondents differ from others, it argues against assuming no difference.

---

12 Enders, C.K. *Applied Missing Data Analysis*, New York: The Guilford Press, 2010.

13 The percentage of respondents who gave "don't know" responses to the *influence* component was even lower – 1% for both residential and nonresidential samples. Similarly, in a dataset of 228 nonresidential respondents from a different evaluation conducted in Ontario, 2% of respondents gave *intention* "don't know" responses and none gave *influence* "don't know" responses.

We recognize that imputing the midpoint may be considered arbitrary. Moreover, our experience is that "don't know" responses are infrequent, and so the way in which they are handled likely will not have a great impact on the resulting free-ridership estimates. Evaluators may implement alternative approaches to handling "don't know" responses in addition to assigning the midpoint and report both results. As an alternative approach, we recommend using linear regression to predict the *intention* score from each respondent's *influence* score.

5.5. Consistency Checks and Related Issue

Consistency checks are frequently used in social and epidemiological research, but a Google search found many more references to using consistency checks to aid data cleaning after survey completion than to resolve seemingly inconsistent response while the survey is ongoing. There are reasons not to include consistency checks in a free-ridership survey.

The assumption that the inconsistency can be resolved accurately may be unfounded. That assumption is based on the belief that the questioner can accurately and reliably determine which of two inconsistent responses is the correct one. A respondent confronted with inconsistent responses may seek to resolve the consistency, but that does not mean that the final response will be accurate. Instead, the response may be influenced by "self-enhancement" motivation.[14]

Other reasons not to confront respondents with inconsistent responses are that doing so may make respondents feel uncomfortable; as a result, it could color later responses; and it lengthens the survey. Lengthening the survey, and perhaps even inducing some discomfort, may be acceptable if the result is better data. However, as argued above, there is reason not to believe that is will do so. Further, the need to assess which response is correct brings more evaluator subjectivity into the assessment. Therefore, we recommend against consistency checks.

5.6. Incorporation of Trade Ally Responses

One evaluator asked how an algorithm for a residential program might incorporate trade ally responses in a manner similar to the ESCO example given in *Section 4.2*, above.

The evaluator may propose an approach for SWE review (see *Section 5.1*).

5.7. Influence from Previous Program Years or Cycles

One evaluator asked whether influence to participate in a program that comes from participation in a previous year (or previous phase) is considered free-ridership.

[14] Swann, William B., Jr. "Self-Verification Theory." In P. Van Lange, A.W. Kruglanski, and E.T. Higgins (eds.), *Handbook of Theories of Social Psychology*. Thousand Oaks, CA: Sage Publications, 2011.

Our experience has been that most regulators limit consideration to the current year or phase. In practice, it may be difficult to determine whether program influence was from the current year or phase or from an earlier year or phase.

Appendix G. Common Approach for Measuring Spillover for Downstream Programs

See memo on the following pages.

research into action inc

Research Into Action, Inc.
PO Box 12312
Portland, OR 97212
www.researchintoaction.com

503 287 9136 (main)
503 281 7375 (fax)
888 492 9100 (toll free)

Memorandum

To: The PA EDCs, TUS, and SWE Team

From: Jane Peters and Ryan Bliss, Research Into Action

Date: January 31, 2014, Finalized February 28, 2014.

Re: Common Approach for Measuring Spillover (SO) for Downstream Programs

The PA PUC Implementation Order specifies that the net-to-gross ratio (NTG) for Phase II of Act 129 is to be treated in the same way as for Phase I. Specifically, for compliance purposes the NTG ratios for Phase II programs continues to be set a 1.0 – basing compliance with energy and demand reduction targets on gross verified savings. However, the PUC order also states that the EDCs should continue to use net verified savings to inform program design and implementation.

The *Framework* states that the "SWE recommends that EDCs use standard sampling techniques, data collection approaches, survey questions, survey instruments, and analysis methodology for NTG studies."[1] Furthermore, the SWE recommends standardization – at a minimum within the EDCs' measurement activities and ideally across all EDCs – for provision of consistency in explaining program effects. The *Framework* also defines participant and nonparticipant spillover ("spillover" or "SO") and recommends the consideration of trade ally surveys and reports for assessing the nonparticipant portion of a program's spillover impact. However, the SWE has determined that while estimation of nonparticipant spillover is desirable, it is not required. If assessed, nonparticipant spillover may be assessed through either a general population (nonparticipant) survey or through a survey of trade allies.

Research Into Action has prepared a memorandum on a common approach for measuring free-ridership for downstream programs (the "FR memo") and has submitted it to the EDCs, TUS, and SWE Team. In it, we discuss the reasons for having a uniform NTG approach for the EDCs. The FR memo has received comments and is undergoing final revision.

To inform the development of a common approach to assessing participant spillover, the SWE solicited descriptions of spillover approaches from the EDCs' evaluation contractors and drafted a memo based on the feedback from the evaluation contractors. The following sections describe the draft common approach, incorporating comments made on the initial draft.

As is the case with the common approach to free-ridership estimation, EDCs and their evaluation contractors may, if they wish, use alternative approaches in parallel with the common approach to assessing participant spillover through self-report surveys or add elements to the common approach, but they should be able to report results from the common approach as described

[1] SWE Team. *Evaluation Framework for Pennsylvania Act 129 Phase II Energy Efficiency and Conservation Programs*. June 30, 2013. Page 61

below in addition to reporting results from alternative or modified approaches to assessing participant spillover. Moreover, EDCs and their evaluation contractors may propose alternative approaches for programs for which the common method may not be applicable, such as approaches focusing on midstream or upstream influences for nonparticipant spillover.

1. Sampling

The *Framework* does not specify confidence and precision levels for estimating spillover. The SWE recommends – but does not require – that the evaluation strive to achieve confidence and precision levels sufficient to provide meaningful feedback to EDCs.

As noted above, the SWE has determined that while estimation of nonparticipant spillover is desirable, it is not required. If assessed, the sampling approach should produce a sample that is representative of the target population (nonparticipants or trade allies) or capable of producing results that can be made representative through appropriate weighting of data. In the case of trade ally surveys, the sampling plan should take trade ally size (e.g., total sales, total program savings) and type of equipment sold and installed (e.g., lighting or non-lighting) into consideration.

2. Participant Spillover

The following provides a description of the SWE's recommended approach for assessing participant spillover. It begins with an overview of the recommended approach. Following are detailed descriptions of the specific approaches for residential and nonresidential participant spillover. The latter cover the SWE's recommended questions and response options to include in participant surveys as well as recommended computational rules for converting survey responses to inputs to the formulas for calculating spillover. The residential and nonresidential participant surveys are slightly different.

2.1. Overview of Recommended Common Protocol

For both the residential and nonresidential sectors, the participant spillover approach will assess, for each participant:

- The number and description of non-incented energy efficiency measures taken since program participation.
  - This may include all energy efficiency measures, even if not eligible for program incentives. However, EDCs should distinguish between program-eligible and other types of measures (including measures that are in the TRM but not eligible for a specific program and energy efficient measures not in the TRM) in their analyses. See further discussion in Section 2.2, below.
- An estimate of energy savings associated with those energy efficiency measures. (Details in Section 2.2, below.)

- The program's influence on the participant's decision to take the identified measures, assessed with a rating scale and converted to a proportion, with possible values of 0, .5, and 1. (Details in Section 2.2, below.)

The specific methods for the residential and nonresidential sector will differ somewhat in details of program influence assessment and estimation of the measure-specific energy savings.

As detailed below, evaluators will calculate spillover savings in four categories:

- For program-eligible measures.
- For measures in the TRM but not eligible for incentives for the program in question.
- For measures not in the TRM but for which the EDC's evaluator can provide reasonable documentation of savings.
- For all measures in any of the above categories.

For each of the above categories, the evaluators will:

- Calculate total spillover savings for each participant as the sum of measure savings by number of units by influence score.
- Total the savings associated with each program participant, to give the overall participant SO savings.
- Multiply the mean participant SO savings for the participant sample by the total number of participants to yield an estimated total participant SO savings for the program.
- Divide that total savings by the total program savings to yield a participant spillover percentage.

2.2. Residential Participant Spillover: Detailed Methods

The residential participant spillover survey will include questions to assess, for each participant: the number and description of non-incented energy efficiency measures taken since program participation; and the program's influence on the participant's decision to take those measures.

2.2.1. Identification of Non-rebated Residential Measures

The survey will assess the purchase and installation of any energy efficient measures, whether eligible for program rebates, in the TRM but not eligible, or not in the TRM. The survey will ask participants a series of questions similar to the following to determine whether they installed any additional energy efficient measures without receiving a rebate:

1. You received a rebate for installing [list of rebated measures]. Since participating in the program, have you installed any additional [list of rebated measures] for which you did not receive a rebate?

   a. [IF YES:] How many/how much have you installed?[2]

2. Since participating in the program, have you installed any other energy efficient products or equipment, or made any energy efficiency improvements for which you did NOT receive a program rebate?

3. [IF YES:] What type of other energy efficient improvements, products, or equipment did you install? [Record description of each additional installed measure]

   a. [FOR EACH MEASURE:] How many/how much did you install?

2.2.2. Assessment of Program Influence on Residential Measures

The survey will ask respondents about the level of influence the prior program participation had on their decision to install the additional measures. The survey may apply a single influence assessment to all measures, under the assumption that residential respondents are not likely to report different levels of program influence for different measures. At the evaluator's discretion, the survey may assess influence for each measure identified.

The SWE recommends that the influence question identify various ways in which the program participation might have influenced the decision to install additional measures. For example, evaluators may consider a question similar to the following:

4. On a 0 to 5 scale, with 0 meaning "not at all influential" and 5 meaning "extremely influential," how influential were each of the following on your decision to [vary wording as appropriate:] install the additional equipment/product(s)/improvement(s)?[3]

   a. Information about energy savings from utility marketing, program representatives, retailers, or contractors

   b. Your satisfaction with the equipment for which you had received a rebate

   c. Your installation of [rebated measure(s)] made you want to do more to save energy

Program influence is assessed as the maximum influence rating given to the four program elements.

- **Example:** A respondent gives influence ratings of 3, 5, and 3, respectively, energy savings information, satisfaction with equipment, and desire to do more. Therefore, the

[2] Ask "how many" for unit items, such as lamps, appliances, and so forth. Ask "how much" for items installed by quantity, such as weather sealing or insulation.

[3] The survey should ask about all three of the above items, as they may have had differing levels of influence. Assessments of "overall program influence" may incorporate the lower ratings of some program elements. However, the final program influence rating will be the maximum influence of any single program element. Moreover, a single question about overall "program influence" may not incorporate influence from information that a program-influenced retailer or contractor provided and does not get at the possible cognitive processes that may have resulted from having undertaken program-induced energy savings.

program influence rating is 5 because at least one program element was "extremely influential."

The maximum influence rating is assigned a value that determines what proportion of the relevant measures' savings is attributed to the program:

- A rating of 4 or 5 = 1.0 (full savings attributed to the program).
- A rating of 2 or 3 = 0.5 (half of the savings attributed to the program).
- A rating of 0 or 1 = 0 (no savings attributed to the program).

At the evaluator's discretion, to provide additional relevant feedback to the program, the survey may ask participants whether there was a reason that they did not receive an incentive for the additional energy efficient technologies.

2.2.3. Assessment of Energy Savings for Residential Spillover

Where applicable, the savings for each additional measure installed will be calculated per the TRM for a rebated measure installed through the program. For partially-deemed measures, a working group of the PEG will develop conservative working assumptions for any required inputs (e.g., square footage of home, R-value improvement, replaced wattage). As an alternative, the PEG working group may identify average verified savings for such measures.

For measures not in the TRM, the evaluator should identify the source and methodology used to assess per-item savings.

2.2.4. Calculation of Total Residential Spillover and Savings Rate

Evaluators will calculate summed spillover savings in four categories:

- For program-eligible measures.
- For measures in the TRM but not eligible for incentives for the program in question.
- For measures not in the TRM but for which the EDC's evaluator can provide reasonable documentation of savings.
- For all measures in any of the above categories.

Evaluators will first calculate spillover savings for each spillover measure reported as the product of the measure savings, number of units, and influence score:

$$Measure\ SO = Measure\ Savings * Number\ of\ Units * Program\ Influence$$

For each of the above categories, the evaluators then will:

- Total the savings associated with each program participant, to give the overall participant SO savings.

$$Participant\ SO = \Sigma Measure\ SO$$

- Multiply the mean participant SO savings for the participant sample by the total number of participants to yield an estimated total participant SO savings for the program.

$$\Sigma Participant\ SO\ (population) = \frac{\sum Participant\ SO\ (sample)}{Sample\ n} \text{ x Population } N$$

- Divide that total savings by the total program savings to yield a participant spillover percentage:

$$\%\ Participant\ SO\ = \frac{\sum Participant\ SO\ (population)}{Program\ Savings} \text{ x } 100$$

2.3. Nonresidential Participant Spillover: Detailed Methods

The participant spillover survey includes questions to assess, for each participant: the number and description of non-incented energy efficiency measures taken since program participation; and the program's influence on the participant's decision to take those measures. The approach for nonresidential participant spillover is similar to that for residential, but differs in some details.

2.3.1. Identification of Non-rebated Nonresidential Measures

The survey will assess the purchase and installation of any energy efficient measures, using questions similar to the following:

1. Since your participation in the program, did you install any ADDITIONAL energy efficiency products or equipment, or made any energy efficiency improvements that did NOT receive incentives through any utility program?

2. [IF YES:] Please describe the energy efficiency equipment installed or energy efficiency improvement? [Probe for measure type, size, and quantity]

The questioner should attempt to document all additional, non-rebated equipment installed since program participation, whether eligible for program rebates, in the TRM but not eligible, or not in the TRM.

2.3.2. Assessment of Program Influence on Nonresidential Measures

The survey will ask respondents about the level of influence the prior program participation had on their decision to install the additional measures. For example, evaluators may consider a question similar to the following:

3. On a 0 to 5 scale, with 0 meaning "not at all influential" and 5 meaning "extremely influential," how influential was your participation in the [NAME OF PROGRAM] on

your decision to [vary wording as appropriate:] install the additional equipment/complete the energy efficiency improvement(s)?

At the evaluators' discretion, the survey may ask the above influence question only once to cover all additional energy efficient installations or improvements or separately for different energy efficient installations or improvements.[4] In the event that a respondent reports many (e.g., more than three) additional non-rebated measures, evaluators have the option of assessing influence for some of them (e.g., the three that deliver the greatest energy savings) and assigning the mean influence score from those measures to the remaining ones.

For each additional energy efficient installation or improvement, the influence rating is assigned a value that determines what proportion of the measure's savings are attributed to the program:

- A rating of 4 or 5 = 1.0 (full savings attributed to the program).
- A rating of 2 or 3 = 0.5 (half of the savings attributed to the program).
- A rating of 0 or 1 = 0 (no savings attributed to the program).

At the evaluator's discretion, to provide additional relevant feedback to the program, the survey may ask participants whether there was a reason that they did not receive an incentive for the additional energy efficient technologies.

2.3.3. Assessment of Energy Savings

Where applicable, the savings for each additional measure installed will be calculated per the TRM for a rebated measure installed through the program. For partially-deemed measures, a working group of the PEG will develop conservative working assumptions for any required inputs (e.g., square footage of home, R-value improvement, replaced wattage). As an alternative, the PEG working group may identify average verified savings for such measures.

For measures not in the TRM, the evaluator may conduct a brief engineering analysis to assess savings or, if applicable, identify an alternative source and methodology for assessing savings.

2.3.4. Calculation of Total Nonresidential Spillover and Savings Rate

The calculation of nonresidential spillover and savings rate is essentially the same as for residential.

Evaluators will calculate summed spillover savings in four categories:

- For program-eligible measures.
- For measures in the TRM but not eligible for incentives for the program in question.

[4] It may be more likely in the nonresidential sector that equipment upgrade decisions will be made on the basis of rational economic considerations, and so the influence of previous program participation may be more likely to vary from measure to measure.

- For measures not in the TRM but for which the EDC's evaluator can provide reasonable documentation of savings.
- For all measures in any of the above categories.

Evaluators will first calculate spillover savings for each spillover measure reported as the product of the measure savings, number of units, and influence score:

$$Measure\ SO = Measure\ Savings * Number\ of\ Units * Program\ Influence$$

For each of the above categories, the evaluators then will:

- Total the savings associated with each program participant, to give the overall participant SO savings.

$$Participant\ SO = \Sigma Measure\ SO$$

- Multiply the mean participant SO savings for the participant sample by the total number of participants to yield an estimated total participant SO savings for the program.

$$\Sigma Participant\ SO\ (population) = \frac{\sum Participant\ SO\ (sample)}{Sample\ n} \text{ x Population } N$$

- Divide that total savings by the total program savings to yield a participant spillover percentage:

$$\%\ Participant\ SO\ = \frac{\sum Participant\ SO\ (population)}{Program\ Savings} \text{ x } 100$$

3. Nonparticipant and Total Spillover

The SWE has determined that while estimation of nonparticipant spillover is desirable, it is not required. Nonparticipant spillover may be assessed through either a general population (nonparticipant) survey or through a survey of trade allies.

3.1. Nonparticipant Survey

If a general population survey is selected, it should assess, for each survey respondent:

- The number and description of non-incented energy efficiency measures taken in the program period.
- An estimate of energy savings associated with those energy efficiency measures.
- The program's influence on the participant's decision to take the identified measures, assessed with a rating scale and converted to a proportion, with possible values of 0, .5, and 1.

Evaluators should submit draft survey questions to the SWE.

3.2. Trade Ally Survey

The following provides an overview of the SWE's recommended approach to assessing spillover through a trade ally survey, followed by the SWE's recommended questions and response options to include in participant and trade ally surveys to assess residential and non-residential SO as well as recommended computational rules for converting survey responses to inputs to the formulas for calculating SO, described above. The residential and nonresidential participant surveys are slightly different and are described in separate subsections. The residential and nonresidential trade ally surveys are essentially identical and are described in a single subsection.

3.2.1. Overview of Recommended Trade Ally Approach

If an evaluator chooses to assess nonparticipant spillover through trade ally surveys, separate surveys should be conducted for the residential and nonresidential sectors. Each survey should assess, for each sampled respondent:

- The number of program-qualified measures sold or installed within the specified sector, in the specified utility's service territory, in the specified program year.
- The percentage of such installations that received rebates from the specified program.
- The trade ally's estimate of the proportion of their sales or installations of non-rebated measures that went to prior program participants.
- The trade ally's judgment of the specified program's influence on sales of the common program-qualified but not rebated measures, assessed with a rating scale and converted to a proportion, with a minimum value of 0 and a maximum value of 1.

The survey should estimate total sales of all program-qualified measures by asking TAs to report sales of their most commonly sold program-qualifying measures and determining what proportion of their total sales of high-efficiency products those measures made up (details in Section 3.2.6, below).

Trade ally survey questions should ask about sales within a specific sector (residential or nonresidential). If an evaluation plan calls for a single trade ally survey in a given sector to provide SO figures across multiple programs within that sector, that survey should be worded to ensure that the trade ally understands that responses should refer to the multiple programs. The SWE will provide guidelines for allocating SO savings across the multiple programs at a later time.

3.2.2. Identification of Non-rebated Measures

The trade ally surveys will ask about sales or installations of the respective programs' most common qualified measures. Theoretically, the survey should assess sales or installations of all program-qualified measures. Otherwise, it will undercount SO. However, doing so would create

unreasonable burden on the respondents and would not likely produce reliable results. Therefore, the recommended common method takes the following approach.

First, evaluators should identify each sampled *trade ally's* most commonly rebated measures as well as other commonly rebated program measures of the type pertinent to the trade ally.[5]

The survey should assess the number of non-rebated units sold of each of the respondent's most commonly rebated measures within the territory of the EDC in question. The introduction to the survey should make it clear to respondents that questions about sales of measures pertain to measures sold within that EDC's territory and that responses should refer to a given sector (residential or nonresidential) and to all of that EDC's applicable programs within that sector.

To prevent undue burden, the survey should restrict the number of measures investigated to no more than four. For each of those measures, the survey should ask respondents questions similar to the following:

1. During the program year, how many [measure] did you sell/install within the service territory of [EDC]?

2. Approximately what percentage of your [measure] installations in [EDC] service territory received rebates through the program?

By subtraction, the response to Question 2 provides the percentage of non-rebated units, of a specific type, sold/installed.

For each of the respondent's most commonly sold program-rebated measures, the number of non-rebated units will be estimated as total number of units sold/installed multiplied by the non-rebated percentage.

As indicated above, it is impractical for the survey to attempt to estimate the number of units sold of *all* program-qualified measures that a respondent sold. This means that the above procedure will underestimate spillover. As a way of providing some information on the possible degree to which spillover is underestimated, the survey should ask respondents to estimate the percentage that their most commonly-rebated products, combined, comprise of their total sales/installations of high-efficiency products, using a question like:

3. Thinking about those types of products together, what percentage do they make up of your total dollar sales of high-efficiency products?

The purpose of this question is not to inform a precise and reliable estimate of additional spillover, but rather to provide information on the possible degree to which spillover is underestimated.

[5] For example, for HVAC trade allies, identify common program-rebated HVAC measures. However, note that many trade allies deal in a range of measure types.

3.2.3. Assessment of Program Influence

For each of the identified measures, the survey will ask respondents about the level of influence the program had on their sales/installations of non-rebated program-qualified measures, using a question similar to the following:

4. Using a 0 to 5 likelihood scale, where 0 is "not all influential" and 5 is "extremely influential," how influential was the program on your sales of non-rebated high efficiency products of that type to your customers?

For each measure identified, the maximum influence rating is assigned a value that determines what proportion of the measure's savings is attributed to the program:

- A rating of 4 or 5 = 1.0 (full savings attributed to the program).
- A rating of 2 or 3 = 0.5 (half of the savings attributed to the program).
- A rating of 0 or 1 = 0 (no savings attributed to the program).

3.2.4. Assessment of Energy Savings

The savings for each additional measure installed will be calculated per the TRM for a rebated measure installed through the program. For partially-deemed measures, a working group of the PEG will develop conservative working assumptions for any required inputs (e.g., square footage of home, R-value improvement, replaced wattage). As an alternative, the PEG working group may identify average verified savings for such measures.

3.2.5. Calculation of Trade-Ally-Reported Spillover (SO)

For each surveyed trade ally, the total SO of each reported measure (i.e., the commonly rebated measures) will be calculated as:

$$Reported\ Measure\ SO = Measure\ Savings * Number\ of\ Units * Program\ Influence$$

The SO from each measure will be summed for each surveyed trade ally to calculate the total SO for that trade ally. Total trade-ally-reported SO for a program can be estimated one of two ways:

- Calculate the mean total SO per trade ally and multiply it by the total number of trade allies, if known, to estimate total SO for the program.
- Calculate the mean SO percentage for each sampled trade ally as the trade ally's total SO divided by the trade ally's total program savings; calculate the mean SO percentage across sampled trade allies (weighted by trade ally size; see below) and multiply that mean SO percentage by the total program savings (from the program database) to estimate total SO for the program.

In either case, the mean total SO or mean SO percentage for trade-ally-reported measures should be weighted by trade ally size using total program sales of non-rebated high-efficiency

equipment (if available) or by a reasonable proxy, such as total program incentives. The means also should be weighted by trade ally type (e.g., lighting or non-lighting).

Total trade-ally-reported SO can be divided by the total program savings to yield a total SO percentage, as:

$$\%\ Total\ Trade\ Ally\ (TA) Reported\ SO\ = \frac{\sum Total\ TA\ Reported\ SO\ Across\ all\ Program\ TAs}{Program\ Savings}$$

The evaluators should calculate and report the weighted mean percentage of total sales of high-efficiency equipment that the reported SO measures constitute. The percentage should be weighted by total sales of high-efficiency equipment (if available) or by a reasonable proxy, such as total program incentives. (Again, the purpose is not to yield a precise and reliable estimate of additional spillover, but to provide a "best available" indication of the degree to which spillover may be undercounted.)

3.2.6. Total and Nonparticipant Spillover

The above approach theoretically yields (but underestimates) total SO because it does not differentiate between sales of non-rebated measure to program participants and nonparticipants.

If responses to the trade ally survey indicate that the trade-ally-identified commonly sold program-rebated measures comprise a large percentage (e.g., 90% or more) of all high-efficiency equipment sold, then evaluators should attempt to determine what percentage of the total trade-ally-identified SO is from nonparticipants by subtracting the total participant SO for that sector from the total trade-ally-reported SO, as:

$$\sum Nonparticipant\ SO = \sum Total\ TA\ Reported\ SO - \sum Participant\ SO$$

That total, divided by the total program savings, yields a non-participant SO percentage, as:

$$\%\ Nonparticipant\ SO\ = \frac{\sum Nonparticipant\ SO}{Program\ Savings}$$

If the trade-ally-identified commonly sold program-rebated measures do not comprise a large percentage (e.g., 90% or more) of all high-efficiency equipment sold, then subtracting participant SO likely will not yield an accurate estimate of nonparticipant SO. In that case, evaluators should report the total trade-ally-reported SO and participant SO.

Appendix H. Quarterly Data Request

PA Act 129 Statewide Evaluation Team

Memo

To: All EDCs

From: The SWE Team

CC: BTUS Staff

CC: Donna Clark, Energy Association

Date: October 7, 2013

Re: Act 129 Quarterly Data Requests – Phase II Updates

The SWE Team is requesting information and data pertaining to the following:

- General implementation and evaluation information;
- Residential Program Data; and
- Commercial & Industrial Program Data.

The objective of this request is to obtain the necessary data and supporting documentation to spot-check the savings reported by each EDC. Please note that the requests detailed in the following sections of this memo may not apply to all EDCs based on variations in program design and implementation. We ask that each EDC supply us with the information requested, or similar information, as applicable for each program with reported savings in the most recent quarterly report. If an EDC has any questions pertaining to the SWE Team's expectations regarding data to be provided for a particular program, please contact a SWE Team member to clarify. Contact information for the SWE Team members, depending on the type of program and data in question, is provided below.

- Residential: Andrea Jester, GDS – andrea.jester@gdsassociates.com and Jeffrey Huber, GDS Jeffrey.huber@gdsassociates.com
- Low Income: Andrea Jester, GDS – andrea.jester@gdsassociates.com and Jeffrey Huber, GDS Jeffrey.huber@gdsassociates.comCommercial and Industrial: Jesse Smith, Nexant – jsmith@nexant.com

All information provided in response to this data request should correspond to the summary data reported in each quarterly report submitted to the SWE Team. Please provide this same type of information with the submission of each quarterly report. Specifically, please provide the SWE Team with the data requested in Sections (2) and (3) of this data requests – this includes all of the data necessary for the SWE Team to conduct our audit of the corresponding quarterly report.

Requests pertaining to program data should be provided to the SWE Team in conjunction with the submission of all quarterly reports (i.e., 45 days after the close of each quarter).

Please upload all information requested to the EDC specific password protected SWE SharePoint Site page under the folder "PYX Data Request Responses". For questions regarding the SharePoint Site, file transfers or the requests of this memo in general, please contact Andrea Jester [andrea.jester@gdassociates.com]

1) General Updates

Please provide the SWE Team with updates on EM&V Activity Plans and Schedules. Going forward, please provide any updates to general program implementation and scheduled EM&V activities on a quarterly basis.

2) Residential Programs

Please provide the following information for all residential programs reporting unverified and/or verified gross savings in the corresponding quarterly report.

a) Appliance Recycling Programs

i. Please upload your EDC's entire database for all program participants for the corresponding quarter.

ii. The SWE Team will choose a random sample of 10 customers from the uploaded database and will provide the EDC with the names of those customersat the end of the program year. Please respond by providing the corresponding JACO work orders for the ten customers selected.

b) Efficient Equipment Rebate Programs

i. Please upload your EDC's entire database for all program participants for the corresponding quarter.

ii. The SWE Team will choose a random sample of 10 customers from the uploaded database and will provide the EDC with the names of those customers at the end of the program year. Please respond by providing the corresponding rebate applications for the ten customers selected.

c) CFL Programs

i. Please upload your EDC's entire database for all bulbs distributed during the corresponding quarter. The database should include the following data fields as applicable:

Table 1: CFL Lighting Data to Be Collected

New Equipment Information	Required Documentation
Name (if applicable) Address (if applicable) Phone number (if applicable) Promotion (Sales, Giveaway)	
Date Type (CFL Standard, Specialty) Manufacturer and Model	Invoice: Giveaway Event

Wattage Quantity Pre-Event Quantity Post-Event Equipment Costs Staffing Costs Misc. Costs	
Date Type Manufacturer and Model Wattage Quantity – Purchase Agreement Quantity – Sales Receipts Equipment Costs (Retail) Incentives Offered	Invoice: Buy-Downs • Purchase Agreement
Date Type (CFL Standard, Specialty) Manufacturer and Model Wattage Quantity – Sales Receipts Equipment Costs (Retail) Incentives Offered	Invoice: Upstream Incentives • Purchase Agreement
Distribution Method Pertinent Documentation of Distribution Date Type (CFL Standard, Specialty) Manufacturer and Model Wattage Quantity – Sales Receipts Equipment Costs (Retail) Incentives Offered	Other Distribution Methods

ii. Please provide us with a copy of all invoices corresponding to the CFLs distributed in the corresponding quarter and represented in the database provided. This should include invoices for both (a) CFL buy-downs and (b) giveaway events. The invoices should include the necessary information to verify the data presented in the database.

iii. Please provide us with a sample of twenty rebate applications, if applicable, corresponding to customer rebates provided in the corresponding quarter for the purchase of CFL bulbs.

iv. If the EDCs' evaluation contractor has performed a quarterly audit and trueup for the same period under review, then the SWE team requests the results of the audit.

d) HVAC and Tune-Up Programs

i. Please upload your EDC's entire database for all program participants for the corresponding quarter.

ii. Please provide us with a random sample of ten rebate applications corresponding to projects claiming savings for the installation of high efficiency HVAC equipment.

iii. Please provide us with a random sample of ten work orders/invoices/rebates corresponding to projects claiming savings for tune-up work done to existing residential HVAC equipment.

e) Home Performance/Audit Programs

i. Please upload your EDC's entire database for all program participants for the corresponding quarter. The database should contain, at a minimum, the data fields specified in the first column of the following table:

Table 2: Other Residential Program Data

New Equipment Information	Required Documentation
Customer Name Address Phone number Program Type (on-line audit, in person audit, etc.)	
Date Measure Type(s) Manufacturer and Model • (identifying info. to determine savings) Measure Count by Type Customer Costs (e.g., audit, equipment installed) Utility Costs (e.g., audit rebates, measures purchased)	Purchase Receipt, Audit Invoice, Audit Report, etc.

ii. Please provide us with a random sample of ten Audit Reports for projects claiming savings based upon audit findings. Please provide us with a sample of ten invoices identifying the measures distributed or installed directly through this program.

f) Whole Building/Weatherization Programs

i. Please upload your EDC's entire database for all program participants for the corresponding quarter. The database should contain, at a minimum, the data fields specified in the first column of the following table:

Table 3: Other Residential Program Data

New Equipment Information	Required Documentation
Customer Name Address Phone number Program Type (e.g., audit, direct install, weatherization, etc.)	
Date Measure Type(s) Manufacturer and Model • Identifying info. to determine savings Measure Count by Type Customer Costs (e.g., equipment installed) Utility Costs (e.g., rebates)	Purchase Receipt, Audit Invoice, Audit Report, etc.

ii. As applicable, please provide us with a random sample of ten work orders and rebate forms corresponding to projects claiming savings based upon whole building and/or weatherization improvements. Please provide us with a sample of ten invoices identifying the measures distributed or directly installed through this program.

g) New Construction

i. Please upload your EDC's entire database for all program participants for the corresponding quarter. The database should contain, at a minimum, the data fields specified in the first column of the following table:

Table 4: Other Residential Program Data

New Equipment Information	Required Documentation
Customer Information Contractor Information	
Date Efficiency Improvements • Identifying info. to determine savings Measure Count by Type Customer Costs (e.g., equipment installed) Utility Costs (e.g., rebates)	Rebate Form, Purchase Receipt, Contractor Invoice, Audit/Modeling Report, etc.

ii. As applicable, please provide us with a random sample of ten invoices/receipts and rebate forms corresponding to projects claiming savings based upon high efficiency improvements installed in the new home. Additionally, please provide all supporting documentation for the same ten projects used to determine the savings associated with the efficient new construction project.

h) Other Residential Programs (as applicable)

Programs falling under this category could include, but are not limited to, give-away events (non-CFL) or kit distribution programs/projects.

i. Please upload your EDC's entire database for all participants in the corresponding quarter. The database should contain, at a minimum:

Table 5: Other Residential Program Data

New Equipment Information	Required Documentation
Customer Name (if applicable) Address (if applicable) Phone number (if applicable) Promotion (Sales, Giveaway, Rebate, etc.)	
Date Measure Type(s) Manufacturer and Model (identifying info.) Measure Count by Type Equipment Costs Rebate/Buy-down Costs (e.g., utility costs)	Purchase Receipt, Invoice, etc.

ii. Please provide us with a random sample of ten work orders/invoices/rebates, as applicable, which correspond to program activities occurring in this past quarter. The invoices should include the necessary information to verify the data presented in the database specifically: measure types; counts; and costs, specifically EDC costs (e.g., rebates, buy-down prices, etc.)

i) Low-Income Programs

i. Please provide all spreadsheets and supporting calculations presenting the total participants and deemed savings for each measure for the reported kWh and kW savings for each low-income program.

ii. Please provide the detailed breakdown of the number of telephone surveys or site inspections performed per quarter by the Implementation Contractor, Evaluation Contractor or other entity.

iii. Please provide a summary of the discrepancies found through all on site or telephone inspections. This summary should provide the number and types of measures that were not installed or are not operating

iv. Please provide the percent of sites that fail inspection.

v. If available please provide an Excel file containing the results of all site inspections.

vi. Please provide the data collected and examined by the EDC's Evaluation Contractor to determine the in service rate for low income measures.

3) Commercial and Industrial Program Requests

The C&I quarterly data requests are divided into five parts. The expected due dates for each part are listed in Table 6. Please note that parts 3c and 3d are dependent on evaluation activities and due dates for responses may vary. Responses to 3e are not required unless a supplemental request is made, typically for ride-along and independent site inspections.

Table 6: C&I Data Request Due Dates

Part	Name	Due Date
3a	Program Database Audit	45 days after quarter close
3b	Project Files Audit	45 days after quarter close
3c	Evaluation Sample Audit	After evaluation sample is selected
3d	Evaluation Audit	After evaluation activities are completed
3e	Ad Hoc Audit	10 business days after request

a) Program Database Audit

For each program, please provide the program database in spreadsheet form at the participant-level for all participants reporting savings in the corresponding quarter.[1] List one participant per line. Each line item must include, at minimum, key fields listed in Table 7. Notes have been provided to clarify the intent of each field. If data for a listed field were not available or not applicable to the program, leave the field blank. If needed, please provide a short write-up to explain missing data and/or program specific characteristics that affect the reliability of the program database.

To enable the SWE team to compile and check the submitted databases efficiently, follow the naming conventions for key fields, shown in Table 7. Following this naming convention ensures that the SWE team processes the data correctly and eliminates the possibility of certain fields being misinterpreted. Each of the required headers should be included in all databases, even if it is not applicable for a particular program.

If an Incremental Monthly Program Extract that contains all the pertinent information with the proper naming conventions above is produced, it can be used as a substitute for the data request response for the Program Population Audit portion (item 3.a).

This information will be cross-checked with figures stated in the EDC quarterly/annual reports.

[1] If measure-level information is readily available, please provide in a supplemental database.

Table 7: C&I Data Request Key Fields

Column	Data Point	Required Field Name	Format
1	Operating Company	EDC	Text
2	Tracking Number (Unique Identifier)	UNIQUEID	Text
3	Program Year	YEAR	Numeric
4	Program Quarter	QUARTER	Numeric
5	Program	PROGRAM	Text
6	Sector	CUSTSEGMENT	Text
7	Building Type	BLDGTYPE	Text
8	Measure	MEASURE	Text
9	Quantity	QTY	Numeric
10	Installation Date	INSTALLDATE	MM/DD/YYYY
11	Recorded Date	RECORDDATE	MM/DD/YYYY
12	Rebate Paid Date	REBATEDATE	MM/DD/YYYY
13	Reported Energy Savings	REPORTEDKWH	Numeric
14	Reported Demand Savings	REPORTEDKW	Numeric
15	Rebate Amount	REBATEAMOUNT	Numeric

Notes (numbers correspond to the column number):

1. EDC: Name of the EDC.
2. Unique tracking number: Identifier for the participant. Participants should be defined in the same way they are defined in EDC reports.
3. Program Year: Program Year during which the participant was “recorded”, as defined in GM-006 (see note #11 for “recorded date” below).
4. Quarter: Quarter during which the participant was “recorded”, as defined in GM-006 (see note #11 for “recorded date” below).
5. Program: Name of the program, as listed in EDC Reports.
6. Sector: Sector of the participant (e.g., Small C&I, Large C&I, Government/Non-Profit/Institutional).
7. Building Type: Building type per TRM categories.
8. Measure: Name of measures.
9. Quantity: Number of measures.
10. Installation Date: When the project was installed and commercially operable, as defined in GM-006.
11. Recorded Date: When the project was “recorded”, as defined in GM-006 (the date when a project is ready to be included in quarterly reports).
12. Rebate Paid Date: When the rebate check was issued to the participant.
13. Reported Energy Savings: Total reported gross energy savings for participant.
14. Reported Demand Savings: Total reported gross demand savings for participant.
15. Rebate Amount: Total amount of rebate issued for participant.

b) Project Files Audit

For each program, please provide information for projects reporting savings during the corresponding quarter. For programs with less than 50 participants during the quarter, please provide information for 5 randomly selected projects. For all other programs, please provide information for 10 randomly selected projects.

- Provide participant-submitted documentation (electronic forms or scans) for sampled projects, including:
    - Application forms, approval forms, installation confirmation
    - Equipment specifications, invoices, certifications
    - Savings calculations sheets
    - Other pertinent information

c) Evaluation Sample Audit

For each program, please provide information regarding the projects in the sample drawn by the EDC evaluator.

- Provide description of process used by the EDC evaluator to select the sample. If a batch process is used to combine quarters, please describe rationale. (This is only required once per program year).
- Provide list of samples and alternates selected by the EDC evaluator. This list must include key fields listed in Table 8:
- Provide participant-submitted documentation (electronic forms or scans) for sampled projects, including:
    - Application forms, approval forms, installation confirmation
    - Equipment specifications, invoices, certifications
    - Savings calculations sheets
    - Other pertinent information

Table 8: Sample List Key Fields

Column	Data Point	Required Field Name	Format
1	Tracking Number (Unique Identifier)	UNIQUEID	Text
2	Reported Energy Savings	REPORTEDKWH	Numeric
3	Reported Demand Savings	REPORTEDKW	Numeric

d) Evaluation Audit

For each program, please provide verification data for projects in the sample drawn by the EDC evaluator. This request may be fulfilled as documentation is made available by the EDC evaluator and is not required at the stated due date of this data request.

- Provide evaluation reports including:
    - Site-specific M&V plans
    - Site-findings
    - Savings calculations
    - Other pertinent information

e) Ad-Hoc Audit

Please provide the following information for projects as requested by the SWE Team on an as-needed basis. Situations where this information is required may include when specific issues are identified on a particular project, reviews for ride-along site inspections, or preparation for independent site inspections.

- Provide complete project documentation (electronic forms or scans) including:
    - Application forms, approval forms, installation confirmation
    - Equipment specifications, invoices, certifications
    - Savings calculations sheets
    - Other pertinent information
- Provide evaluation reports including:
    - Site-specific M&V plans
    - Site-findings
    - Savings calculations
    - Other pertinent information

Appendix I. Low-Income Measures Memo Clarification

PA Act 129 Statewide Evaluation Team

Memo

To: All EDCs
From: PUC Staff and the SWE Team
CC: TUS/BCS Staff
CC: Donna Clark, Energy Association
Date: October 10, 2012; (Revised May 15, 2014)
Re: Act 129 Low-Income Measure Reporting –Clarification

STATEMENT OF ISSUE

The reporting of low-income measures has presented challenges throughout Phase I. Currently, there is not a consistent methodology across EDCS to define a low-income measure, to group similar measures, or to differentiate a low-income measure from a non-low-income measure. The legislation defines "measures" in terms cited below, but each EDC has interpreted this definition differently. When the SWE sought to capture the number of specific low-income measures and total measures offered by the EDCs, the inconsistencies were revealed.

DISCUSSION OF ISSUE

PUC Staff, jointly with the SWE, has determined that it would benefit all Act 129 participants to reconcile this issue for Phase II of Act 129. In order to standardize and make consistent the reporting in Phase II of Act 129, it will be necessary to use an agreed upon methodology for grouping similar measures and for distinguishing low-income specific measures. For Phase II, there are two separate compliance criteria: meeting the "proportionate number of measures" and achieving the 4.5% kWh savings target.

Act 129 defines an Energy Efficiency and Conservation (EE&C) measure (in the definitions section; 66 Pa.C.S. 2806.1(m)) as follows:

> Energy efficiency and conservation measures.
> (1) Technologies, management practices or other measures employed by retail customers that reduce electricity consumption or demand if all of the following apply:
> (i) The technology, practice or other measure is installed on or after the effective date of this section at the location of a retail customer.
> (ii) The technology, practice or other measure reduces consumption of energy or peak load by the retail customer.

(iii) The cost of the acquisition or installation of the measure is directly incurred in whole or in part by the electric distribution company.
(2) Energy efficiency and conservation measures shall include solar or solar photovoltaic panels, energy efficient windows and doors, energy efficient lighting, including exit sign retrofit, high bay fluorescent retrofit and pedestrian and traffic signal conversion, geothermal heating, insulation, air sealing, reflective roof coatings, energy efficient heating and cooling equipment or systems and energy efficient appliances and other technologies, practices or measures approved by the commission.

Staff proposes that EDCs refer to the PA TRM when determining the appropriate level of granularity at which to list measures when calculating the "proportionate number of measures." Technologies which are addressed by a single algorithm section in the TRM should not be further subdivided. Measure divisions should be based on equipment types, not differences in equipment efficiency or sizing of the same type of equipment. For example, Central AC units should be considered a separate measure from High Efficiency Heat Pumps because they are two distinct equipment types addressed in Section 2.1.1 of the PA TRM. However, all residential Central AC units should be identified as a single measure regardless of SEER value or tonnage. Similarly, EDCs should not separate compact fluorescent light bulbs into multiple measures based on wattage.

An exception to this TRM based measure definition is Commercial lighting retrofits. Section 3.2.2 of the 2013 PA TRM offers a single series of algorithms for non-residential lighting improvements. The SWE and PUC Staff recognize that this section should be divided into a number of technologies including linear fluorescents, LED exit signs, LED traffic signals, street lighting, occupancy sensors and de-lamping. A grouping approach which distinguishes between equipment types, but not sizes or efficiency levels should be employed for measures which are not addressed in the PA TRM.
With regard to determining which measures can be classified as specific Low-Income measures, the legislation states the following:

(G) The plan shall include specific energy efficiency measures for households at or below 150% of the federal poverty income guidelines. The number of measures shall be proportionate to those households' share of the total energy usage in the service territory. The electric distribution company shall coordinate measures under this clause with other programs administered by the commission or another federal or state agency. The expenditures of an electric distribution company under this clause shall be in addition to expenditures made under 52 pa. Code ch. 58 (relating to residential low income usage reduction programs).

For Phase I, the Low-Income Working Group (LIWG) Report calculated the methodology for determining the total energy usage in an EDC's territory and provided those figures in the report. The usage calculation figures have been updated with 2010 Census data for Phase II of the Act and are provided below in Table 1. The LIWG Report did not provide the EDCs with a means of distinguishing the low-income measures from non-low-income measures, but did allow for design adjustments in EE&C plans beyond May 31, 2013. *See*, LIWG Report at Docket M-2009-2146801.

Table 1: Phase 2 Low-Income Target Proportions by EDC

EDC	Percent 2011 kWh Usage Low-Income Households vs. Total Consumption
Duquesne	8.402%
PECO	8.799%
PPL	9.950%
Met-Ed	8.787%
Penelec	10.231%
Penn Power	10.639%
West Penn Power	8.794%

PROPORTIONATE NUMBER OF MEASURES CRITERIA

Staff proposes that in order to provide standardization across the EDCs, that a measure be determined specific to low-income if the measure is:

a) Provided at no cost to the low-income customer and
b) Targeted to the low-income customer (defined as a customer at or below 150% FPIG).

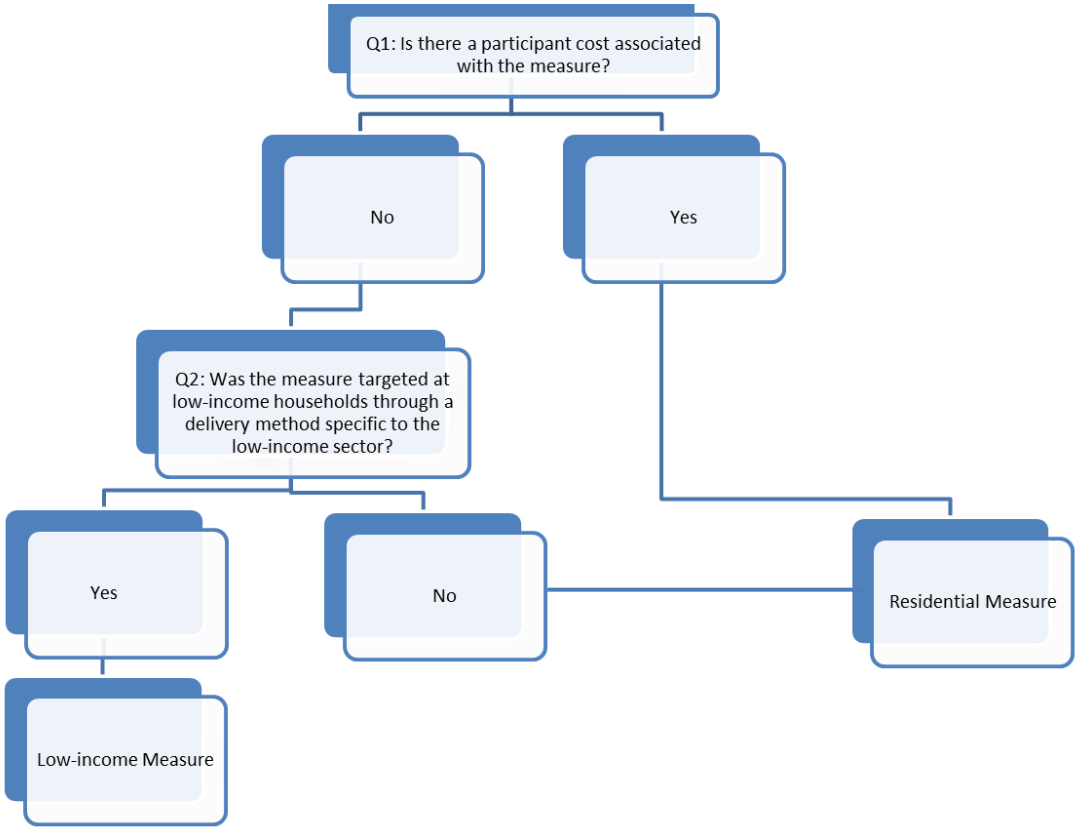
Staff is unaware of any current or past specific low-income measure that requires a participant cost to a low-income customer. For Phase II, only specific low-income measures will be counted towards satisfying the "proportionate number of measures" compliance component.

Please note that our proposed definition does *not* require that the measure/measure type be installed to be counted. Under the definition discussed above, the measure would count if it is targeted to low-income customers and is offered at no cost to low-income customers. If an EDC offers a measure under a specific Low-Income Program (example, mattress), but no customers end up having the measure (mattress) installed, it would still count towards satisfying the "proportionate measures" requirement.

Staff recognizes the possibility of a single measure being classified as both a low-income and a non-low-income measure if it is offered in two different programs with different levels of financial responsibility for the participant. For example, an EDC may offer an HVAC tune-up measure in its standard residential portfolio where it pays homeowners a $50 rebate toward the cost of the service. The balance of the cost of implementing this measure is the responsibility of the homeowner. This same EDC may offer an HVAC tune-up measure in its low-income program where 100% of the cost of the improvement is paid by the EDC. In this example, "HVAC tune-up" should be included twice in the EDC's list of measures offered, but only one occurrence is considered a specific low-income measure. Figure 1 provides a methodology EDC s can use to determine whether a given measure in its portfolio is:

1. A low-income measure (no cost to the participant and targeted to the low-income sector)
2. A general residential measure
3. Offered via two different delivery mechanisms or two different levels of participant cost (free/not free). Therefore the measure counts once in the numerator of the 'proportionate number of measures' ratio and twice in the denominator.

Figure 1: Process Map for Determining Low-Income Measures



During Phase I of Act 129, several EDCs provided 'kits' to customers in their low-income programs. Staff believes that each distinct equipment type within these kits should be counted as a separate measure. If an EDC provides low-income program participants with a kit which includes 4 CFLs, a furnace whistle and an LED nightlight this should be counted as three measures (CFL, furnace whistle and LED nightlight) when calculating the proportion of measures offered to the low-income sector.

ENERGY SAVINGS (4.5%) CRITERIA

A separate issue involves non-low-income measures and programs that low-income customers can choose to participate in. These savings can be counted toward the Phase II 4.5% compliance criteria, once verified via approved surveying methods. *See*, 2013 TRC Test Order, entered August 30, 2012, pg 49. Since these programs will not be targeted to low-income customers, they will not meet the definition of specific low-income measures. However, the savings will still apply toward the 4.5% savings compliance criteria.

Please note that there does not have to be consistency between the two compliance requirements. The first requirement; "proportionate measures," is taken directly from the Act. This was determined by the legislation to ensure that there were specific measures available for and provided to low-income customers. The second compliance requirement; "4.5% savings," was developed by the Commission to gauge the impact on the Low-Income sector. It is recognized that low-income customers have and will continue to participate in non-low-income specific programs, and capturing these savings provides a means of determining to what extent that is occurring. Further, it is understood that the low-income

population is comprised of "known participants (ex. CAP, LIURP, LIHEAP, etc.)" and an almost equally large segment of the population that is "unknown (meets 150% of FPIG, but is not identified in any EDC database)". Those "unknown" low income customers would likely not be serviced under Act 129, even though they are eligible for the measures at no cost, as are the "known" low-income participants.

CONCLUSION

EDCs should use the foregoing information as guidance for examining compliance with regard to the low-income programs included in their EE&C plans. EDCs can consider the definition of low-income measure to satisfy the proportionate number of measures requirement for Phase II to be a measure that is targeted to low-income customers and is available at no cost to low-income customers. And then to satisfy the 4.5% savings requirement, EDCs will be provided the flexibility to include savings from general residential measures that low-income customers may have participated in (determined by surveying as outlined in the 2013 TRC Test Order) in addition to savings attributed to specific low-income measures.

It is important to note that the proportionate number of measures will be examined when compliance is assessed for Phase II. If it an EDC's Annual Report shows that there are not enough measures available specifically to the low-income sector, then EDCs will likely be directed to expand their offerings.