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 30, 2013

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

CMP: Custom Measure Protocol

Cv: Coefficient of Variation

DLC: Direct Load Control

DR: Demand Response

DSM: Demand Side Management

EC: Evaluation Contractor

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

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

PA PUC: Pennsylvania Public Utility Commission

PEG: Program Evaluation Group

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

TRM: Technical Reference Manual

TWG: Technical Working Group

UMPSP: Uniform Methods Project Sampling Protocols

VFD: Variable Frequency Drive

VOI: Value of Information

Please see Appendix A for 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 (EDC) whose energy efficiency and conservation (EE&C) program plans were approved by the Pennsylvania Public Utility Commission (PA 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, in March 2013, the PA PUC contracted with a Statewide Evaluation Team (SWE Team), led by GDS Associates, Inc. (the Statewide Evaluator, or SWE) 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 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 PA 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 PA 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 it was developed.

In accordance with the RFP, 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 (EC) 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 levels
    - Statewide Market Potential Study for additional energy and load reductions

In order to prepare for the program year beginning June 1, 2013, the PA 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 is able to develop plans and prepare for its responsibilities that begin on June 1, 2013. This timing will allow 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 PA 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 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 PA 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 requirements
- Required reporting formats
- Data management and quality control guidelines and requirements
- Guidance and requirements 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 PA 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 PA 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 in 2016. 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 PA 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 PA PUC entered an Implementation Order at Docket No. M-2012-2289411. As part of the Implementation Order and Act 129, the PA 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 PA 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	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[4]		XX	
Review and get approval of Statewide Baseline, Market Potential, and metering studies			XX
Initiate and coordinate annual updates to TRM and interim updates (new protocols)		XX	
Get approval of 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	EC		

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

Coordinate the development of and approve the methodologies for EDC NTG studies consistent with the NTG white paper		XX	
Update NTG white paper as needed.		XX	

Table 1-2: Roles and Responsibilities – Audit and Assess 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 gain approval of EDC evaluation plans		XX	XX
Review and update existing Phase I Audit Plan (now known as the Evaluation Framework)		XX	
Update the Evaluation Framework as needed		XX	
Gain approval of revised statewide Evaluation Framework			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[5]	ICSP		
Establish and implement quality control of EDC program tracking database(s)[6]	EC	XX	
Oversee statewide data management and quality control, including: design, implementation, and maintenance of statewide database of program, portfolio, EDC and statewide energy and demand savings and cost-effectiveness reporting		XX	
Develop and maintain secure SharePoint site for maintenance and exchange of confidential data and information with EDCs		XX	

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

[6] 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.

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		
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 PA 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 PA 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 EE program and portfolio net and gross impacts, as applicable, and cost-effectiveness, and EDC progress in reaching targets	EC		
Develop the statewide semi-annual and annual report templates. Review EDC reports and advise the PA 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 PA PUC.		XX	
Review and approve SWE semi-annual and annual reports			XX
Review EDC's 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

Table 1-7: Roles and Responsibilities - Best Practices

Task and/or Deliverable	EDC	SWE	PUC
Participate in impact evaluation process review and improvement workshops as needed	EC		XX
Prepare best practices recommendations for improvements to impact 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 PA 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, PA 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 the EDC EE&C plans implemented in a cost-effective manner in accordance with the TRC test? • Did the EDC EE&C plans produce real energy and demand savings?	• Which programs were the most successful and why? • Which programs were the most cost-effective and why? • How can EDCs improve programs in order to meet statutory targets?

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 sectors have the greatest potential for additional energy and demand savings?
Pennsylvania EDCs	• What factors contributed to differences between planning estimates and reported gross savings? • What factors contributed to differences between *reported* gross savings and *verified* gross savings? • What factors contributed to differences between planned cost-effectiveness and actual cost-effectiveness? • 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 changes can the EDCs adopt to improve cost-effectiveness of the programs and portfolios?

2 Policy Requirements

2.1 Requirements From the Phase II Implementation Order

Act 129 requires the PA PUC to:

Adopt an "energy efficiency and conservation program to require electric distribution companies[7] 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"[8]

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"[9]

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."[10]

Based on findings from the Phase I Market Potential Study dated May 10, 2012, the PA 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 PA 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.[11]

2.1.1 Phase II Energy Reduction Targets for Each EDC

The PA 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.

[7] This Act 129 requirement does not apply to an electric distribution company with fewer than 100,000 customers.
[8] 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.
[9] 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.
[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 51.
[11] 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

The PA PUC's framework also encourages EDCs to exceed the targets if additional, cost-effective savings can be obtained within the programs' budgets. The PA PUC's goal is not to just require achievement of minimum targets. The PA PUC stipulates that the EDCs are to continue to acquire cost-effective savings until they reach the end of a phase. At that point they may stop spending their approved budgets and seek reconciliation of revenues collected compared to actual costs incurred.

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

The PA 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.
3. EDCs should determine the initial mix and proportion of energy efficiency programs, subject to PA PUC approval. The PA 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 PA 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 achievement of the prescribed reduction targets.[12]

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 demand response (DR) programs. Instead, the PA PUC established a three-year term for Phase II for several reasons. One primary reason was to enable the PA 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 PA 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 PA 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 PA 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 PA 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 PA PUC directed that NTG research be used to direct Act 129 program design and implementation, but not for compliance purposes.[13] The PA 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 PA PUC directed in its Phase II Order that EDCs 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.

[12] "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.

[13] 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.

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 PA PUC to use a TRC Test to analyze the benefits and costs of the EE&C plans that certain EDCs must file.[14] The PA 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 PA 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 Order 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 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. Finally, the Order addresses proper protocols for determining measure incremental costs, as discussed 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 efficiency costs. For 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 replacement of a functioning device, incremental cost is defined as the full cost of the new efficient measure (including installation costs) that is being bought to replace the existing, installed measure. 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 cost is the difference between the full cost of the efficient device (including installation) minus the present value of the standard device (plus installation costs). "Present Value" indicates that the early replacement costs should be discounted to reflect the time value of money associated with the installation of the efficient device compared to the installation of the standard device that would have occurred at a later date[15]. If an EDC uses this

[14] The Pennsylvania TRC Test for Phase I was adopted by PA 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.

[15] National Action Plan for Energy Efficiency, Understanding Cost-Effectiveness of Energy Efficiency Programs, November 2008, page 4-16. EDCs should refer to this NAPEE report for information on how to calculate early retirement or replace on burnout measure costs.

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, adjusted values from the DEER database, 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.

2.2.4 TRC Order Schedule

The SWE will coordinate and participate in a periodic review of the TRC Test the PA PUC and EDCs use to determine cost-effectiveness. The PA 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 PA PUC in updating the TRC Test during the first eight months of 2015 if the PA 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 PA 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.[16] The PA PUC also directed the Bureau of Conservation, Economics and Energy Planning (CEEP)[17] to oversee the implementation, maintenance, and periodic updating of the TRM.[18] On January 16, 2009, in the Energy Efficiency and Conservation Program Implementation Order for Phase I of Act 129's EE&C Program,[19] the PA PUC adopted the TRM as a component of the EE&C Program evaluation process. In the Phase I Implementation Order, the PA 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."[20] Soon after the adoption of the EE&C Program Phase I Implementation Order, PA 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.[21] In adopting the 2009 TRM, the PA PUC

[16] Order entered on October 3, 2005, at Docket No. M-00051865 (October 3, 2005 Order).
[17] 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.
18 See October 3, 2005 Order at page 13.
[19] 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.
[20] Ibid.
[21] 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.

recognized the importance of updating the TRM annually.[22] A program evaluation group (PEG)[23] 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 PA PUC convened a Technical Working Group (TWG)[24] meeting to discuss the proposed TRM updates.[25]

During the Phase I of Act 129, the PA 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 PA PUC website.[26]

The approval date of the TRM is the date the TRM Order was entered after the PA 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.

Table 2-2: Phase I Approved TRMs[27]

TRM Version	Program Year	Approval Date	Effective Date[28]
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[29]
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

[22] *Ibid*., pages 17 and 18

[23] The PEG is chaired by PA 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.

[24] The TWG is chaired by PA 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.

[25] The PA 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.

[26] See link:
http://www.puc.state.pa.us/filing_resources/issues_laws_regulations/act_129_information/technical_reference_manual.aspx

[27] See Table 2-3 for Phase II TRMs

[28] 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.

[29] 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 PA PUC, which effectively overrode all 2009 TRM protocols. This was an exceptional case; normally, such a decision does not occur unless explicitly stated through PA PUC directives. See 2010 TRM Order, pages 16-17.

For Phase II of the Act 129 EE&C program, the PA PUC again adopted the TRM as a component of the EE&C Program evaluation process.[30] The Phase II Implementation Order also recognized the importance of the continued use of an annual updating process for the TRM for Phase II.[31] The PA 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.

The TRM Order represents the PA PUC's continuing efforts to establish a comprehensive TRM with a purpose of supporting the EE&C Program provisions of Act 129. The PA 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 PA 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.

2.3.1 Purposes of the TRM

The TRM serves a variety of purposes for Act 129. In addition to providing measure 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 PA 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 also use the 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 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

[30] *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.
[31] *Ibid*. at page 75

2.3.2 TRM Update Process

In order for the TRM to be an effective tool for Act 129, the PA PUC ordered a regular annual update to the TRM through the 2009 Implementation Order.[32] The PA PUC intends to update and expand the TRM annually to fulfill the requirements of the EE&C provisions of Act 129.[33] 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 PA 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).

Figure 2-1: TRM Update Process (Annual)

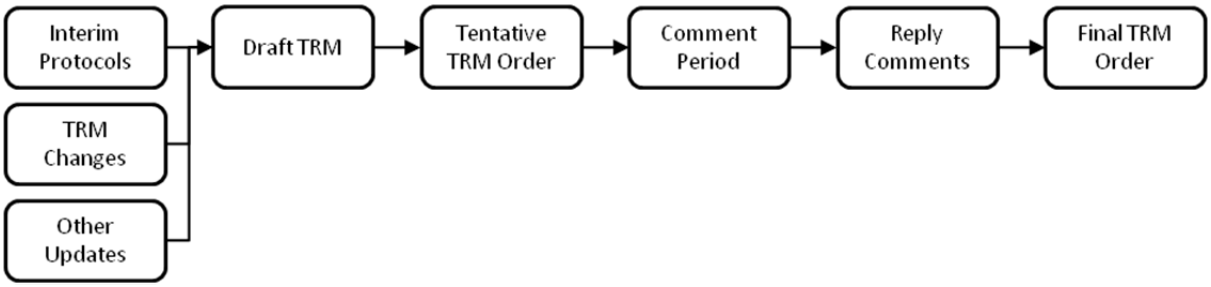


Table 2-3 displays the estimated schedule for approval of the TRM and effective dates. While the actual dates of the deadlines leading to the approval of each TRM may change, the PA PUC targets an approval date of December 31, which provides a five-month period between the approval date and the effective 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 less than five months.

Table 2-3: Timeline for TRM Updates

TRM Version	Program Year	Approval Date	Effective Date[34]
2013 TRM	PY5	December 20, 2012	June 1, 2013 – May 31, 2014
2014 TRM	PY6	December 31, 2013*	June 1, 2014 – May 31, 2015
2015 TRM	PY7	December 31, 2014*	June 1, 2015 – May 31, 2016

***Estimated date**

As stated before, the *approval date* of the TRM is the date the TRM Order was entered after approval by the PA PUC during a public meeting. This differs from the *effective date* of the TRM, which specifies when

[32] 2009 Implementation Order reference
[33] Phase II Implementation Order, pp. 71-72
[34] 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.

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, comprised of 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 PA PUC regarding the quantification of energy and demand savings. 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

Much of the guidance provided in this Framework is intended to align with the 2014 TRM and its future trajectory. Because the effective date of this Framework falls before the effective date of the 2014 TRM, the SWE recognizes that there is potential for conflicting expectations between the 2013 TRM and the guidance within this Framework. For example, Sections 2.3.4, 2.3.5, and 3.3 discuss thresholds above which customer-specific data collection is expected to be conducted in the non-residential sectors by ICSPs or evaluation contractors. The 2013 TRM makes no reference to thresholds for non-residential sectors. In the 2014 TRM update, the addition of thresholds and increased options for "EDC data gathering" will be points of emphasis. When Framework guidance is dependent upon prospective TRM requirements, the effective date of that particular Framework guidance is considered to take effect coincident with the corresponding TRM revision. The following content of this Framework should be considered effective June 1, 2014:

- Thresholds (for non-residential measures) by end-use categories above which customer-specific data collection is required (2.3.4, 2.3.5, and 3.3)
- Grouping of measures into end-use categories (2.3.5.1)
- Guidance on how ICSPs should use the TRM to calculate *ex ante* impacts (2.3.5.1)
- Guidance on how evaluation should use the TRM to calculate *ex post* impacts (2.3.5.2)
- Adoption of the PJM peak demand window (3.3.2.2.7, and 3.3.2.2.8)

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.

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 PA 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 PA 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 protocols for high-impact measures that should be reviewed annually and provide any needed clarifications or modifications based on evaluation findings, statewide studies, and the most reliable, recent secondary research through the PEG.
- 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 on an annual basis based on the above listed activities.

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
- Horsepower of a fan or pump motor

Recommendations of the SWE to the PA 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, 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 non-residential 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 C&I lighting metering studies and eQUEST modeling for C&I HVAC, motors, and VFD measures) to update key assumptions for high-impact measures and provide load shapes for each measure variant.[35]
- 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 and a consistent framework for EDC ICSPs to estimate *ex ante* (claimed) savings and for EDC evaluation contractors to estimate *ex post* (verified) savings achieved for the Pennsylvania energy efficiency portfolios.

[35] The schedule for developing load shapes for measures in the TRM will be determined by the SWE in collaboration with the EDCs and their evaluation contractors.

2.3.5.1 Using the TRM to Determine Ex Ante Savings

This section outlines how EDC ICSPs should calculate *ex ante* savings.[36]

For replacements and retrofits, the applicable date for determining which TRM version to use 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,"[37] and when savings actually start to occur. This is analogous to when a commercial customer's meter "sees" the savings under expected and designed-for operation. For most projects, this is obvious. For projects with commissioning, the CDO occurs after the commissioning is completed. For incented measures that have been installed, but are not being used because there is no occupant, or will not be used until another, 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 begin to accrue at 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. Values for any open variables in the algorithm will come either from customer-specific information or default values provided in the TRM. EDC ICSPs should attempt to collect these values for each rebated measure, ideally through the measure application form or application process. Only variables specifically identified as open variables[38] may be adjusted using customer-specific information. Fully deemed assumptions not indicated as an open variable may not be adjusted by implementers using customer-specific information.[39]

The determination of whether or not to use default values for open variables in the TRM in the savings calculations is a function of the impact (savings) and uncertainty associated with the measure. For example, default TRM parameter values may be used for low-impact and low-uncertainty measures such as lighting retrofits in a small business facility. On the other hand, customer-specific values determined through interviews with facility contacts, metering, or EMS data collection are required for high-impact

[36] In some cases, an EDC may choose to implement a program "in-house" rather than engaging an implementation CSP. In this case EDC staff is acting in the capacity of the implementation CSP.
[37] 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.
[38] Open variables are listed with a default value and an option for "EDC Data Gathering" in the TRM.
[39] The SWE will collaborate with the EDCs and their evaluators to identify any deemed/stipulated assumptions for measures in the TRM that should be changed to "open variables" during the annual TRM update process.

and high-uncertainty measures, such as HVAC or lighting retrofits in universities or hospitals that have diverse facilities, and where those types of projects represent a significant share of program savings for the year.

The determination of when customer-specific data are required involves the following steps:

Identification of the appropriate evaluation sampling stratum
Determination of the expected total ex ante savings for each measure in the project

For non-residential measures, the TRM provides end-use specific thresholds[40] that should be used to determine when customer-specific data at the measure level are required for *ex ante* savings. Section 3.3.2.3 provides information on evaluation sample stratification of projects for the *ex post* savings analysis.

If potential savings for individual end-use categories (i.e. lighting, motors) within projects fall below the threshold specified in the TRM, the ICSPs may gather customer-specific data, or ICSPs may use the default deemed value for any open variable below the threshold specified in the TRM. It is important 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. Evaluation contractors 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.

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. If an EDC does not wish to use the TRM protocols for a measure, they may use a custom method to calculate and report savings in their quarterly and/or annual reports, as long as they also calculate savings using the TRM protocols and report them appropriately.

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 for which following a deemed or partially deemed approach is appropriate, 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

[40] The threshold kWh/yr is stipulated in the 2014 TRM and will vary depending on the type of measure.

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	Deemed kW and kWh/year	Furnace Whistle
TRM partially deemed measures	Savings algorithms with open variables	C&I lighting, Residential lighting (CFLs & LEDs), C&I motor
Custom measures	All other, unspecified	Non-TRM VFD, Non-TRM chiller, EMS

For **deemed measures**, the TRM provides per-unit savings allowances to be used by both the ICSPs and evaluators; 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 that will 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[41] values in the TRM algorithm as possible with customer-specific or program-specific information gathered through evaluation efforts. Note that 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. Using more accurate customer-specific values for a partially deemed measure is encouraged for high-value non-residential projects above a threshold

[41] Open variables are signified by the term "EDC data gathering" in the TRM

kWh/yr[42] when doing so is allowed by the TRM. 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 non-residential 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 *ex ante* savings and/or ask their evaluation contractor to use a custom method to verify *ex post* 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 *ex ante* and *ex post* protocols in the quarterly and/or annual reports in which 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 PA PUC may challenge their reported 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 PA 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 and their evaluators are encouraged to use customer-specific or programs-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 savings 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 PA 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 PA PUC recognized that it was

[42] The threshold kWh/yr is stipulated in the TRM and will vary depending on the type of measure.

necessary to expand the TRM by adding more savings protocols from the early stages of Act 129.[43] 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.[44] 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* 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.

2.3.6.1 Interim Protocol Approval Process

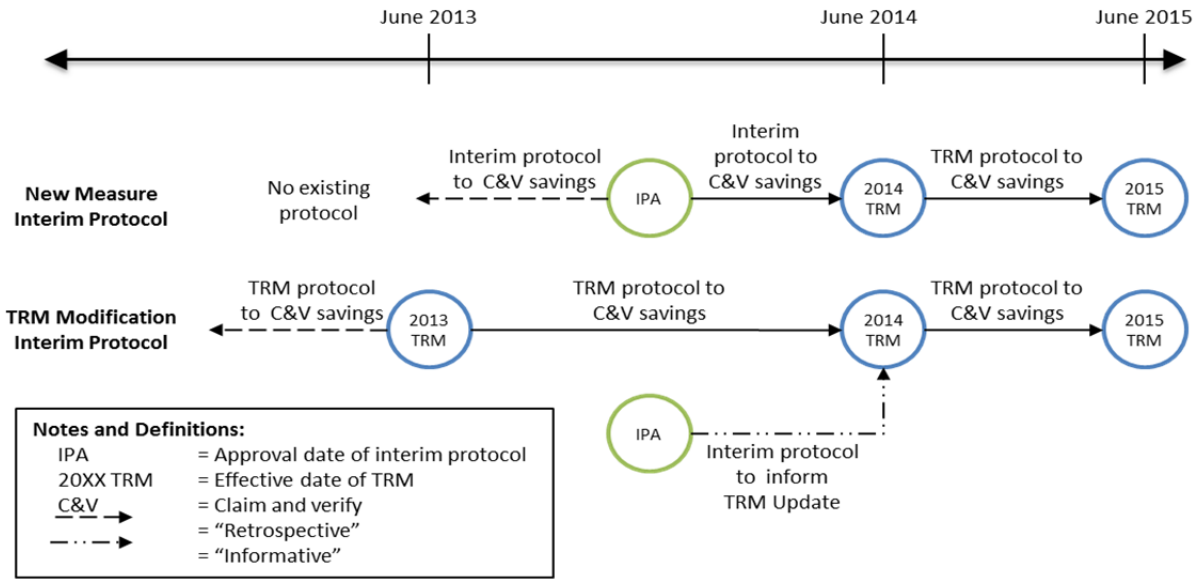
The IMP approval process is informal, and is intended to minimize risk for EDCs planning to offer measures that do not have a TRM protocol by developing savings protocols through a collaborative review process in the PEG. The IMP review and approval process includes the following steps:

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.

[43] 2009 TRM Order, Page 10.
[44] Ibid. Page 17-18.

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 PA PUC approval. The PA PUC-approved TRM Protocol, whether changed or unchanged, will apply prospectively; a PA PUC-approved TRM Protocol will not apply retrospectively, unless the PA 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, PA PUC-approved TRM Protocol, it cannot override that associated PA PUC-approved TRM Protocol. The TRM Modification Interim Protocol will be used to inform the next TRM update. When the PA PUC approves the proposed Modification Interim Protocol, the protocol will be effective prospectively, although EDCs may submit a formal request to the PA PUC for retrospective application of the protocol.

2.3.7 Custom Measures

While TRM measures are reviewed and approved by the PA 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 implementers 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.[45] 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 should recommend the IPMVP option and M&V protocols to be used by the ICSP.

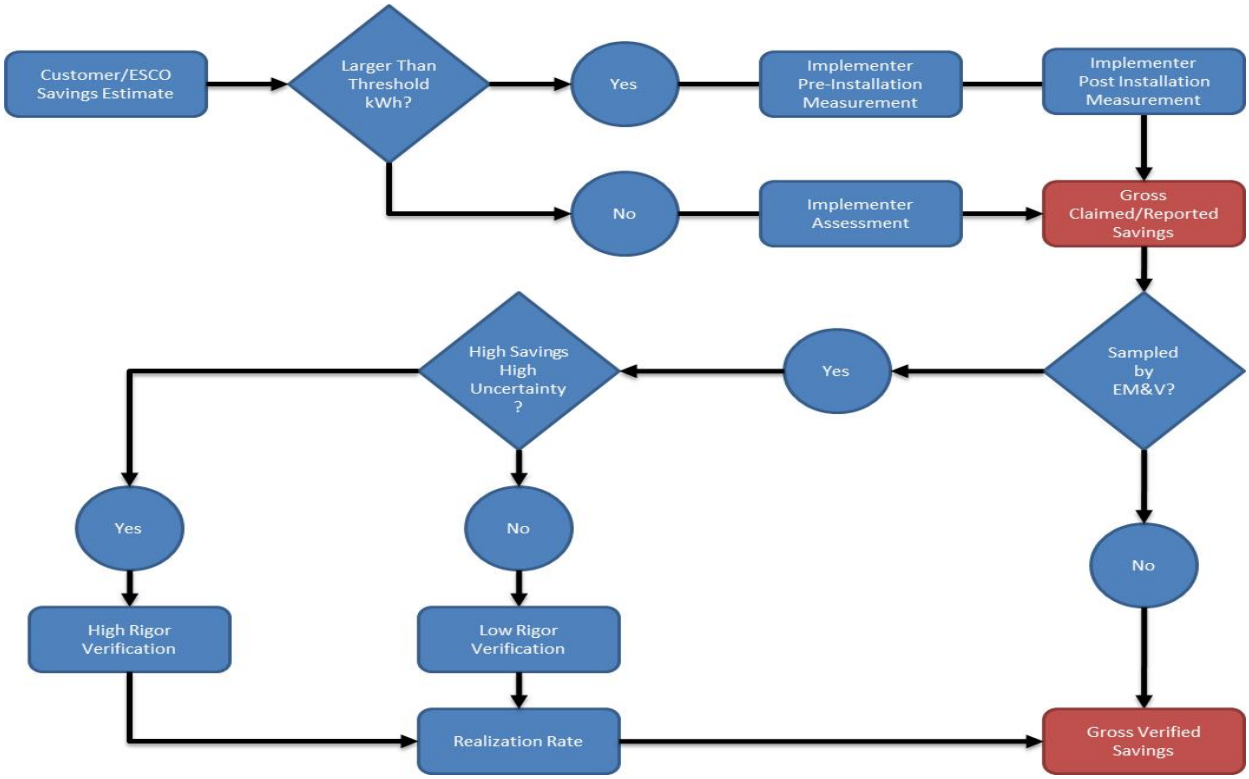
The PA PUC will not determine M&V protocols for custom measures to improve the EDCs' ability to support energy services that meet the EDCs' energy savings goals. EDC evaluation contractors are permitted to determine the appropriate M&V protocols for each project. EDC evaluation contractors must verify 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.[46] 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, but not required, that evaluators' baseline verifications also include site inspections for large custom measures with high savings and high uncertainty according to the evaluators' best practice. Ideally, these would be independent of the EDC ICSP's site inspection in order to minimize the influence of the evaluator on the *ex ante* savings, although that is not always possible and may unnecessarily burden the customer.

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, and include them in the program's realization rate calculations at the discretion of the EDC evaluation contractor.

[45] 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.
[46] Ibid.

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

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, which 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.

2.4 Guidance Memos

This Evaluation Framework is developed to provide an overarching framework for Act 129 programs and may not address all nuances discovered through the actual implementation process. For such issues, the SWE will develop guidance memos to clarify and memorialize decisions through an iterative review process

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

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 PA 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 PA PUC and, therefore, are not binding on the PA PUC.

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 usually are 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 are 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 evaluation plan for each evaluation and submit it to the SWE SharePoint site for review within 90 days from the beginning of the program year (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. Changes to program delivery and evaluation approach are typically limited within a program phase, so 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.

Each EDC and its evaluation contractor will choose the optimal structure and design for their evaluation plans. At the least, the evaluation plan should reflect a shared understanding of the program delivery mechanisms, research objectives and methodology, data collection techniques, 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. 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 EECM 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

In order 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 value to the corresponding value in the EDC quarterly and annual reports. 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[48].

3.2.2 Installed Dates, In-Service 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.
- **Reported Date**: Also referred to as the Recorded Date. The date the measure is entered into the program system of record. 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.
- **Rebate Date**: The date the program administrator issues a rebate to the participant for implementing an energy efficiency measure.

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 is reported in a later quarter/year than the quarter/year the measure went in-service. For example, a measure goes in-service in PY5 but is not 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.[49] *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 PA PUC. EDCs are directed to file preliminary annual reports on July 15 and final annual reports on

[48] Some worksheets used to calculation of individual customer impacts will not be embedded in the tracking system.
[49] 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.

November 15 following the end of the program year[50] using the existing reporting structure, which accounts for 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 which 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.[51] EDCs should include another table representing kW savings.

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 ISD/CDO 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 savings actually start to be realized by the customer 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 measures, the date which governs the version of the TRM an EDC should use to determine savings is the date the application is received by the EDC from the customer. This is to account for the long life cycle of new construction projects which 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 have been submitted. 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 in the next quarterly data submission after the correction has been made. This will allow the SWE to update its records and track program progress using the corrected values.

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. 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 2012 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[52]
- Seasonal Energy Efficiency Ratio of the qualifying unit being installed

[50] Secretarial Letter entered on May 25, 2011, at Docket No. M-2008-2069887.
[51] 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.
[52] This assumes that an "early replacement" savings protocol is followed.

- 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 induced benefits, which include reductions in electric energy usage, electric demand, and avoided air emissions[53] that can be attributed directly to an energy efficiency program. There are many stages to an impact evaluation. At each stage of the evaluation process, decisions must be made 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.

Impact evaluation contractors use data collected during program implementation and conduct independent data-gathering activities. If the data collected by the ICSP is unreliable, end-use equipment operating conditions have changed post-installation, or 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. 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. 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 assessment of key variables (rigor) at a sampled site or among program participants depends on the value of that information in confirming the claimed savings.

There is a hierarchy in 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 framework. Open communication between ICSPs and evaluation contractors helps reduce or eliminate redundant data collection efforts when appropriate (Section 3.3.2).The SWE follows the SWE audit protocols (Table 3-1) to audit and confirm the evaluation contractor's verified impacts.

[53] Air emissions are not included in EE&C program impact evaluations, although estimates of emission reductions can be estimated easily using Energy Information Administration and Federal Energy Regulatory Commission reports and gross energy savings.

Table 3-1: Impact Evaluation Hierarchy

Level	Description	Requirements/ Discussion Points
1. Program Implementation • Deemed savings • Partially deemed savings • Custom/unspecified	M&V protocols and Site-Specific M&V plans used to calculate *ex ante* savings • TRM protocols • Interim TRM protocols • Custom measure protocols	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* savings are claimed using the appropriate protocol (TRM, IMP, or CMP).
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, or CMP).	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*. For programs 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 2.3.3, 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 what 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 verify the *ex ante* savings claimed by the EDCs. This framework establishes a minimum level of detail in order 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.

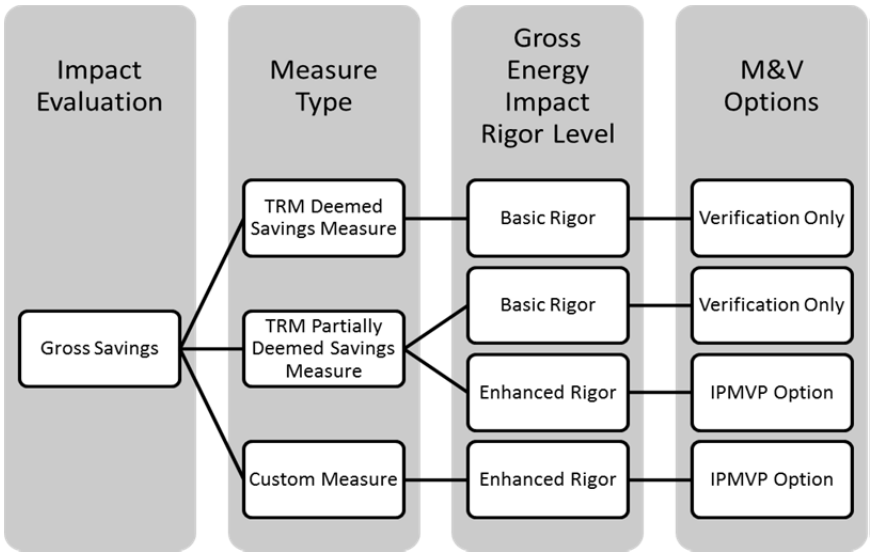
The Evaluation Framework defines two levels of rigor: *basic* and *enhanced*. Each level of rigor provides a class of minimum allowable EM&V methods, which are based on standard evaluation practices, in order to offer flexibility for the evaluation contractors to assess and propose the most accurate and cost-effective methods to verify gross savings, by 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 3-3. 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 3-2 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 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.[54] 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.

Figure 3-1: Required Protocols for Impact Evaluations



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

Table 3-2: Required Protocols for Impact Evaluations

Rigor Level	Minimum Allowable Methods for Gross Impact Evaluation
Basic	1. Verification-only analysis for TRM fully deemed measures. 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 on-site.
Enhanced	1. Simple engineering model with EM&V equal to IPMVP Option A for TRM partially deemed measures. 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) 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 required 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.

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 applies mainly to the TRM fully deemed measures, but also may be used for TRM partially deemed measures with impacts 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. 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 on-site 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 measures producing savings above the threshold values[55] identified in the TRM as requiring customer-specific data collection.

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 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

[55] Thresholds will only apply to non-residential measures

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 generally 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 impact 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 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 allocation of gross energy savings through the use of allocation factors, coincidence factors, or end-use load shapes during the hours of 2:00 pm to 6:00 pm 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 3-3. Custom measures should follow an enhanced rigor approach.

The SWE encourages evaluation contractors to recommend improved coincidence factors values using a load shape from metered or vetted sources, when applicable. 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. This 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 pm to 6:00 pm 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. However, the actual number of strata used will be at the evaluation contractor's discretion and the guidance in 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 3-3: Definitions of Program Strata and Their Associated Levels of Rigor for Impact Evaluation

Strata Level	Minimum Allowable Methods for Gross Impact Evaluation
Stratum 1 – High-Impact and/or High-Uncertainty Measures [56]	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, that are not already taken into account in the TRM, exist between measures
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.

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.

[56] The EDC and evaluation contractor may determine the appropriate level of impact and uncertainty when stratifying measures.

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. 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.

If the evaluation contractor finds that a measure was not installed or is not operating in the manner specified in the project's design documents, 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 necessary 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. Efficiencies can be gained by requiring participants to supply this information on program application forms. The first objective is to verify that the existing equipment is applicable to the program under which it is being replaced and that the installed equipment will reduce energy consumption within the home or business. 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 on-site inspection, 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. At an enhanced level of rigor, measure installation should be verified through on-site 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, and 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 On-Site Sampling of Installations

This section provides guidance in determining the number of installations to measure and verify during the on-site 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 on-site. This 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. 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 monitoring 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 sample unit may be the individual measure, a particular circuit or point of control, as designated by the evaluation plan. Systematic sampling with a random starting point is acceptable. The recommended relative precision for sampling on-site installations is ± 20% at the 90% confidence level at the facility level. The sampling unit (measure, circuit, control point) 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 assumption for determining the minimum sample size should be set at 50% as this will maximize $p*(1 – p)$ and guarantee that precision targets are met. Increased verification proportions may be used for projects completed by contractors with a proven record of performance in the program.

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 AHU 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 on-site counting exercise, an error band 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 connected kW load. If the evaluation-adjusted kW (connected load) for any usage group in the sample is within +/- 5% of the claimed kW, the project savings should be accepted at the claimed value, or the EDC evaluators should revise and recalculate the calculations. For example, if the program tracking record for a project claims that 240 fixtures were retrofitted in the hallways of an office building and the evaluation contractor confirms the fixture wattage, but counts only 238 fixtures, it is not necessary to adjust the savings estimate. However, if the evaluation contractor verifies only 210 fixtures in the facility hallways or observes a 10% higher or lower fixture wattage than what was assumed by the ICSP, *ex post* savings values should be recalculated based on the evaluator's observations.

3.3.3.2.4 Site Specific Measurement and Verification Plan

A Site Specific M&V 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. 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 includes specifying data to be gathered and stored for field measurements that document the project procedures and rationale. 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 high priority 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. It should be noted that the IPMVP is written to allow 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 a 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 custom measures must 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. The site characteristics description should include:

- General 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;
- Building occupant information, such as number of occupants, occupancy schedule, and building activities;
- Internal loads, such as lighting power density, appliances, and plug and process loads;
- Type, quantity, and nominal efficiency of heating and cooling systems;
- Important HVAC system control set points;
- 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 occur and while drafting the SSMVP. However, evaluators should include as much attainable descriptions as feasible in the SSMVP and include any remaining descriptions in the final on-site report.

EM&V Method: The EM&V method chosen for the project should be specified. EM&V methods generally should adhere to the applicable IPMVP protocol for the defined level of rigor. The evaluation contractors have considerable latitude regarding the development of a 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. Site-specific data collection activities should be conducted as permitted by the TRM. Equations and documentation supporting baseline assumptions should be provided. This 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 framework should be specified.

3.3.3.3 Additional Guidance for Ancillary Programs

In addition to the common guidelines provided in Section 3.3.3, certain ancillary programs require special treatment. The MMPs for each of these programs is discussed in detail in this section.

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 should 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.

Act 129 low-income programs are very similar in scope to the Low Income Usage Reduction Program (LIURP) efforts the EDCs also administer. In many cases, the two programs use the same ICSPs. The following guidelines are established in an effort to standardize savings calculations between LIURP and Act 129 and between EDCs within Act 129.

1) **Weather Normalization** – Savings calculations based on a comparison of pre-retrofit and post-retrofit consumption in a home must be weather-normalized to ensure that savings estimates are not unfairly impacted by differences in the observed weather patterns between the two periods. The relationship between consumption and weather should be determined via a regression analysis that uses cooling degree days (CDD) and heating degree days[57] (HDD) with base 65 degrees (F). These relationships should be applied to the 20-year normal values for the appropriate region from 1990 to 2009.
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:
    - Twelve (12) months[58] of pre-retrofit billing history and 12 months of post-retrofit billing history are 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, the site should be excluded.
    - 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 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.

If an EDC changes the measure mix offered, the EDC may not include homes prior to the change in the savings analysis for homes evaluated after the change occurred. For example, if an EDC provides additional attic insulation to participants in PY5, but discontinues this practice in PY6, participating homes from PY5 cannot be included 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.

[57] A HDD term typically is not necessary for homes with natural gas or oil (non-electric) heat.
[58] 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.

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 a custom protocol (MMP) that was developed by the evaluation contractor and approved by the SWE. As a guide, the SWE recommends a billing analysis be performed that is similar in scope to IPMVP Option C. In addition, the SWE recommends properly specifying control and treatment groups. The treatment group's consumption pre-intervention could serve as a control group, provided results are adequately normalized to account for weather or other exogenous changes. The control group also could be composed of a sample of homes or buildings similar in size, location, or demographics to the treatment group. Absent a basis for program distinction, if a pre/post analysis is conducted only on program participants, evaluators should assume participation within the treatment group is similar to that of the entire service territory.

It is important to note that a billing analysis will capture the effects of all changes made within a home, so there is potential for double-counting of savings if participants are engaged in other programs within an EDC's portfolio. One example is a participant in a behavioral program who also receives a rebate for replacing their 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 excluding such customers from the *ex post* analysis, or subtracting the savings associated with other rebate offerings from the impact estimate for the program.

Upstream programs (such as discounted CFLs) complicate the situation, because customer participation is not tracked and evaluators will be unable to identify upstream participants and remove them from a billing analysis. If a pre/post analysis is conducted only on program participants, evaluators should assume participation within the treatment group is similar to that of the entire service territory. For example, if an EDC's average residential customer purchases 1.3 upstream CFLs per program year, the savings impact from 1.3 CFLs should be subtracted from the estimates from the behavioral program.

If a matched control group is used, evaluation contractors should assume that upstream participation in the test group is proportional to upstream savings in the residential portfolio as a whole. Consider an EDC where 60% of the residential energy savings come from upstream CFLs and 40% from rebated (tracked) measures.[59] If an analysis of the behavioral program participants determines that 2% of that program's savings actually came from measures rebated in other programs, the following calculation will determine the proportion of the behavioral program savings that should be attributed to upstream CFLs.

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

$$CFL\ Share = 3\%$$

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

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 provide all of the data necessary to determine the true impact values of 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.

Equation 1: Coefficient of Variation

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

Equation 2 shows the formula used to calculate the required sample size for an evaluation sample, 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 2: Required Sample Size

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

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%)

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.

The sample size formula shown in Equation 2 assumes that the population of the program of the population is infinite and that the sample being drawn is reasonably large. In practice, these assumptions are not always met. The provisions for addressing these situations are discussed below:

1) 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 2 by the FPC formula shown in Equation 3 will produce the required sample size for a finite population.

Equation 3: Finite Population Correction Factor

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

Where:

N = Size of the population

n = Sample size

2) 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 replace the z-statistic shown in Equation 2.

3.4.1 Annual Evaluation Precision Requirements for Verified Savings Estimates

Table 3-4 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 3-4 assume a two-tailed design and must be met or exceeded annually. A gross verified energy savings estimate with 10% relative precision at the 90% confidence implies that there is a 90% chance that the true annual kWh impact of the program is within +/- 10% of the savings estimate. 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.

Table 3-4: 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[60] 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 3-4.
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

[60] 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.

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

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 3-4 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 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. Evaluators should use their professional judgment to develop size thresholds 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 3.4.3 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 3-4 for programs which use a census regression approach. The resulting precision of a regression model generally 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.

If the 85% confidence interval around the savings estimates for a program that uses a billing regression analysis includes 0 kWh, an EDC should not claim savings for the program until a more precise estimate of program impacts can be 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 Sampling Techniques

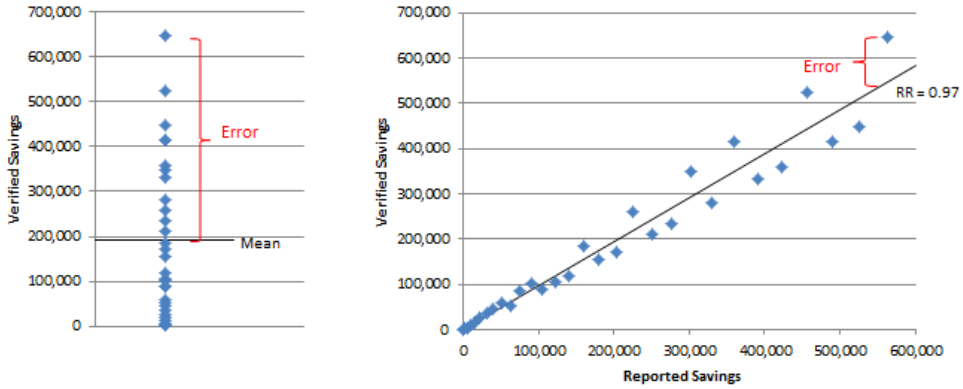
Evaluators may choose to employ two broad classes of probability sampling techniques in the impact evaluation of EE&C programs:

1) **Simple Random Sampling** (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) **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 the 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 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:

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



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 PA 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 3-4 are met, the relative precision values of the total Phase II savings will meet or exceed the annual requirements at the same level 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 always should represent the mean estimate of savings, or the midpoint of the confidence interval. In addition to the verified savings estimates for energy and demand, EDCs should report the error bound and the relative precision of the savings estimate such that:

Equation 4: 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, etc.) (This formula will differ according to the sampling technique utilized)

$z - statistic$[61] = Calculated based on the desired confidence level and the standard normal distribution

Table 3-5 provides the appropriate z-statistic to use for several commonly used confidence levels. Each value assumes a two-tailed design.

Table 3-5: Z-statistics Associated with Common Confidence Levels

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

The parameter of interest (proportion, realization rate) 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 5: 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[62] for each program should be reported, as well as the confidence level at which it was calculated. This formula is shown in Equation 6:

Equation 6: Relative Precision of the Savings Estimate

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

[61] As described in Section 3.4, the t-distribution and associated t-statistic are the proper parameters to use when sample sizes are smaller than 30.

[62] 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.

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.

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 for 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 which expire prior to May 31, 2016[63]. 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 7 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-level 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 4 and Equation 5 using the standard error of the parameter estimate and the z-statistic associated with the 90% confidence interval (1.645).

Equation 7: 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 8: Relative Precision of Phase II Savings Estimate

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

Equation 7 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.

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. Systematic uncertainty is error that is introduced into evaluation findings consistently (not randomly) through the manner in which parameters are measured, collected, or described. Systematic uncertainty is more challenging to quantify and mitigate than sampling uncertainty because sources of systematic

[63] 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.

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. On-site 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, 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 estimate for a residential CFL. In the 2013 Pennsylvania TRM, the value for CFL hours of use 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.

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:

$$\Delta\text{kWh/yr} = (\text{Watts}_{\text{base}} - \text{Watts}_{\text{CFL}}) * \text{CFL}_{\text{hours}} * 365 / 1000 * \text{ISR}^{64}_{\text{CFL}}$$

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

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

The program savings estimate is equal to:

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

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

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

These upper and lower bounds of the deemed CFL_{hours} term is 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.

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

[64] 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.

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.

Given the number of deemed or stipulated values in the TRM, inclusion of 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[65] (UMPSP) identifies six areas, which may be examined to determine how rigorously and effectively an evaluator has attempted to mitigate sources of systematic error. A summary of the six areas is given below:

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?

[65] UMPSP were in draft form when the SWE wrote this Evaluation Framework. Draft protocols can be found at http://www1.eere.energy.gov/wip/pdfs/ump_draft_protocols_01-24-2013.pdf

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?
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 is warranted.

3.6 Net Impact Evaluation

The PA PUC stated in the Phase II Implementation Order that NTG adjustments be treated the same way for Phase II as were during Phase I. "Specifically, the Commission [PA PUC] proposed that NTG research be used to direct Act 129 program design and implementation, but not for compliance purposes."[66] Their reasons for this 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'."[67] The PA PUC believes that, based on published studies, these two effects often come close to offsetting each other and result in a NTG ratio close to 1.0.[68] Due to the substantial additional costs to calculate "free-riders" and "spillover," the PA 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 PA PUC will continue to mandate that the EDCs calculate the NTG ratio as they did for Phase I.[69]

Efforts to measure net savings are valuable to inform program design and program planning, as well as future phases of Act 129. The EDCs' Evaluation Contractors should conduct NTG research and consider

[66] 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.

[67] Ibid, page 83.

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

[69] 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.

conducting additional research to assess market conditions and market effects to determine net savings. Market effects research is discussed in Section 3.6.2.3.

3.6.1 Purpose of Net-to-Gross Research

NTG research traditionally has two primary purposes: 1) adjust gross savings to reflect actual program influence on savings, and 2) explicate 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 if a program should be modified, expanded, or eliminated based on its net-to-gross ratio.

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 highly 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 on-site confirmation is included, it becomes very costly. [70]

Social science research suggests that using self-reports to measure free-riders and spillover is not a valid measure of free-ridership. However, years of research have shown that NTG self-report questions are reliable. That is, when these questions are asked across different populations or at different points in time, the results are similar and tend to vary over time in the expected direction. Thus, the SWE believes it is reasonable to conclude that NTG free-rider and spillover questions result in measurement of something that is positively correlated with true free-ridership, and thus can be useful in assessing changes over time or differences across programs. However, it is very unlikely that this approach yields an accurate quantitative point estimate of free-ridership or spillover and thus, supports the PA PUC's decision that self-report-based NTG should not be used to calculate net savings estimates for compliance purposes.

The primary concern of the SWE is whether the EDCs' evaluations are helping the EDCs fully understand the effects of their programs on the markets in their service territory. Further, the SWE must ensure that the EDCs are maximizing the influence the program can have to both grow the market for energy efficiency in their service territory and to ensure that ratepayer funds primarily support customers who need that support in order to invest in energy-efficient solutions.

3.6.2 Acceptable Approaches

Program evaluators traditionally use one of several methods to assess a program's net savings, including self-report surveys, econometric methods, and market effects approaches. For Phase II of Act 129 the NTG ratio will not be used to determine whether the EDC met their energy and demand reduction targets in

[70] 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 Efficiency Economy.

Phase II of Act 129. The SWE recommends that EDC evaluation contractors opt for a market share approach if possible and use a survey-based approach for determining free-ridership and spillover rates when a market-share approach cannot be used. The PA PUC has approved the following three approaches for assessing program effects on the market.

- Self-Reports for Estimating Free-Ridership and Spillover
- Econometric Approaches
- Market Share and Market Effects Studies

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

- **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 NTG ratio removes free-riders from the savings calculation and adds program spillover. The NTG is defined below in Equation 9:

Equation 9: 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.

3.6.2.1.1 Free-Rider Measurement

The SWE recommends that EDCs use standard sampling techniques, data collection approaches, survey questions, survey instruments, and analysis methodology for NTG studies. Standardization – at a minimum within the EDCs' measurement activities and ideally across all EDCs – can provide consistency in

explications of the programs' effects. The SWE will develop a standard set of questions and a reporting approach for the EDCs to use during Phase II[71]. A timeline for these questions will be developed in the fall 2013.

Care must be taken when developing the battery of questions used to measure free-ridership and spillover. The research approaches used in Massachusetts, Wisconsin, and by the Energy Trust of Oregon (Energy Trust) describe some of the best practices for free-ridership and spillover calculation and can serve as models for the Pennsylvania EDCs. The SWE recommends an approach similar to that chosen by the Energy Trust, which uses a short battery of questions and additional satisfaction questions. This battery is implemented within a month after participants have participated in an Energy Trust program.[72]

The responses to a sequence of free-ridership questions are used to compute an overall free-ridership score for each measure. It is very important that more than one question be used to determine the level of free-ridership. Free-ridership questions should include at least two components:

- Intention
- Program influence

Each component provides a portion of the variance in the estimate. This is also how partial free-riders emerge. If the respondent had specified more than 0% and less than 100% of the measures, they could be considered partial free-riders. Free-ridership estimates should be consistent with gross savings estimate requirements for 85% confidence with ±15% in precision at the program level, and 90% confidence with ±10% precision at the sector level. However, it is important to note that because of the number of questions used and the variable response rates, the final confidence limits may be broader than 85/15. It is likely that error compounds across multiple batteries and question sets, where possible this should be addressed by the EDC evaluator.

3.6.2.1.2 Spillover Measurement

It is more challenging and generally more costly to estimate spillover than free-ridership because spillover claims often require on-site measurement in addition to self-reports to verify whether claimed installed measures in fact qualify for the program. Net savings claims that include spillover studies are more robust that those that just include free ridership estimates.

Spillover may occur when a participant or nonparticipant reports any of the following:

(1) Installing a measure that is above code;
(2) Installing a measure that did not receive any rebate (through any channel); or

[71] The EDCs are allowed to use their existing approaches until a standard set of questions are developed and the schedule for adoption is defined by the SWE. The SWE plans to develop this standard set of questions during Phase II of Act 129.
[72] http://energytrust.org/library/reports/101231_Fast_Feedback_Rollout.pdf

(3) A sufficient level of program influence to indicate they would not have installed the measure (or the same number of measures) in the absence of the program.

Examples of spillover include:

- Program participants adopting additional measures without an incentive
- Consumers acting on the program influencing available energy-using equipment in the marketplace
- Impacts of architects and engineers using more-efficient practices due to the program, thereby forcing consumers into desired behaviors
- Changes in nonparticipants' behaviors resulting from program direct marketing or changes in stocking practices

Trade ally surveys and reports also can help evaluators assess spillover. The following responses on trade ally reports should be considered when determining the nonparticipant portion of a program's spillover impact:

- They installed measures in nonparticipating projects that were the same as those installed in participating projects, or above code
- They installed qualifying measures without a program incentive
- The program exerted a sufficient degree of influence on the decision to indicate they would not have installed the additional measures if they had not participated in the program

3.6.2.2 Econometric Approaches

Econometric approaches may be used to estimate net savings. Typically, these use historical billing data and require a nonparticipant group of similar buildings for which the owner has invested in similar equipment without program support.

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.[73] An 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.

Another approach sometimes used for opt-in or voluntary commercial-sector programs requires that the evaluator conduct on-site verification of the energy efficiency level of the equipment and a survey of both participants and nonparticipants. A discrete choice model typically estimates the net savings and the program effects.

Another econometric approach with billing data commonly used for opt-in or voluntary residential programs is to include participants and nonparticipants with similar buildings. A second-stage model using

[73] 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.

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.

The primary disadvantages of these two approaches are: 1) the difficulty in identifying comparison groups of similar buildings, or those in which similar equipment has been installed, and 2) the additional cost. Further, it is not possible to disaggregate free-riders or to identify spillover, so this approach inadvertently may embed spillover in the net savings estimates, as is the case for the self-report approach.

3.6.2.3 Market Share and Market Effects Studies

Studies of market share and market effects have value as a means to estimate program effects and to provide information on the market needs and changes in the market for energy efficiency in an EDC service territory. The purpose of measuring market effects is to make appropriate strategic decisions about program offerings and timing so that the market for energy-efficient products and services grows more readily than it would have without the program.

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."[74]

Market share and market effects studies 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.[75]

(1) Baseline measurements are very important; these form the basis of comparison. They 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 the 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, etc.
(4) Program cost-effectiveness that includes market assumptions lends itself better to estimating program effects as the assumptions can be tested and refined throughout program implementation.

[74] 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.
[75] Hoefgen, L., Angela Li, and Shel 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.

In summary, NTG ratios 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 useful to determine if program design changes are needed relating to such features 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 in order 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 respond to the program goals, rather than an ideal. 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 allows the necessary flexibility and control for program administration and management, and ensures the PA PUC that the EDCs are managing 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, research activities, and specific issues to be addressed, along with a schedule of milestones and deliverables[76].

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 evaluation. 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 take time to implement. The New York State Process Evaluation Protocols[77] 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 aid program staff in understanding how the market is structured and operating (characterization) and how the market is 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 and useful 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 in these cases, data collection 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 compare the effectiveness of changes to previous program years. These early feedback evaluations should be

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

[77] 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/Process_Evaluation_Protocols.pdf

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 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 in real time regarding opportunities for improvement. Real-time evaluations typically last for one to two years, with ongoing data collection, as well as quarterly to bi-annual reporting that targets the type of information program staff need to gauge their program's progress and effectiveness.

Most programs do not need a process evaluation every year of their implementation cycle. For instance, some aspects of a program may be assessed most successfully during its second or third year. On the other hand, it may be beneficial to conduct a process evaluation of new programs during their first year and to involve the program evaluation staff early in the design process. The value of process evaluation is in *the doing*. Process evaluations are considered a formative evaluation; by conducting the evaluation, the evaluator actually has an effect on the program. For this reason, it is important to consider the reason and timing for the process evaluation carefully, and address them in the research plan.

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. An example is usability studies of websites; or production engineers in industry. These situations are uncommon.
- 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 used
- 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 typically can glean a wealth of information about the program from data that the program maintains, including: the tracking system; program communications documents (usually electronic but possibly print); and the materials used for marketing, outreach, and publicity. There also may be process flow diagrams, program theory and logic documents, program 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 sample 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 program-related individuals include: communications within the program and with customers and stakeholders as well as staff 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 can 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 PA 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 often can be 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 experience of the customers who are participating in or have participated in a program, in order to design 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 the program and their participation experience, and 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 level of savings 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 commonly 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, and not to rely on single or minority points of view.

Evaluators must construct the data collection instruments carefully, to ensure that similar questions are posed across groups. This will facilitate the analysis process. It also is 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).
- The SWE will follow up with the EDC staff to determine how the EDC plans to address the recommendations.
- Through these conversations, the EDC will tell the SWE what, if any, action 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 utilized as inputs to 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 as appropriate. In accordance with the PA 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 Commission's latest TRC Order.

3.8.1 TRC Method

The PA 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."[78]

According to the California manual:

> *The Total Resource Cost Test measures the net costs of a demand-side management 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.[79],[80]

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).

[78] *California Standard Practice Manual: Economic Analysis of Demand-Side Program and Projects*: October 2001
[79] *Ibid*: October 2001; P. 18.
[80] Pennsylvania Public Utility Commission, *2013 Total Resource Cost Test Order*, Docket No. M-2012-2300653, August 30, 2012.

3.8.3 Aligning Measure Savings with Incremental Measure Costs

In order 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 made, 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 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 3-6, along with a summary of the definition of incremental measure costs and savings for each of the decision types.

Table 3-6: Measure Decision Types

Type of Measure	Incremental Measure Cost ($/Unit)	Impact Measurement (kWh/yr/Unit)
New Construction	Cost of efficient device minus cost of standard device	Consumption of standard device minus consumption of efficient device
Replace on Burnout (ROB)	Cost of efficient device minus cost of standard device	Consumption of standard device minus consumption of efficient device
Early Replacement - Simple (Example: Replacement Windows, Additional Insulation, Air Sealing)	Cost of efficient device plus installation costs	Consumption of old device minus consumption of efficient device
Early Replacement - Advanced (Example: HVAC equipment, Appliances)	Cost of efficient device minus cost of standard device plus remaining present value	*During remaining life of old device:* Consumption of old device minus consumption of efficient device *After remaining life of old device:* Consumption of standard device minus consumption of efficient device
Early Retirement (No Replacement)	Cost of removing old device	Consumption of old device[81]

[81] The advanced retrofit 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. "Present Value" indicates that the early replacement costs should be discounted to reflect the time value of money associated with the installation of the efficient device compared to the installation of the standard device that would have occurred at a later date.

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.

3.8.4 Data Requirements

In order 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[82] or the National Action Plan for Energy Efficiency's "Understanding Cost-Effectiveness of Energy Efficiency Programs" report.[83] Below is a brief list of these data requirements:

- General Modeling Assumptions
  - Avoided energy costs
  - Avoided capacity costs
  - Avoided transmission and distribution costs
  - Energy and peak demand line losses
  - Utility Discount Rate
- 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[84] (non-incentive costs)
  - Program/measure load shapes
  - Measure-specific peak coincidence factor

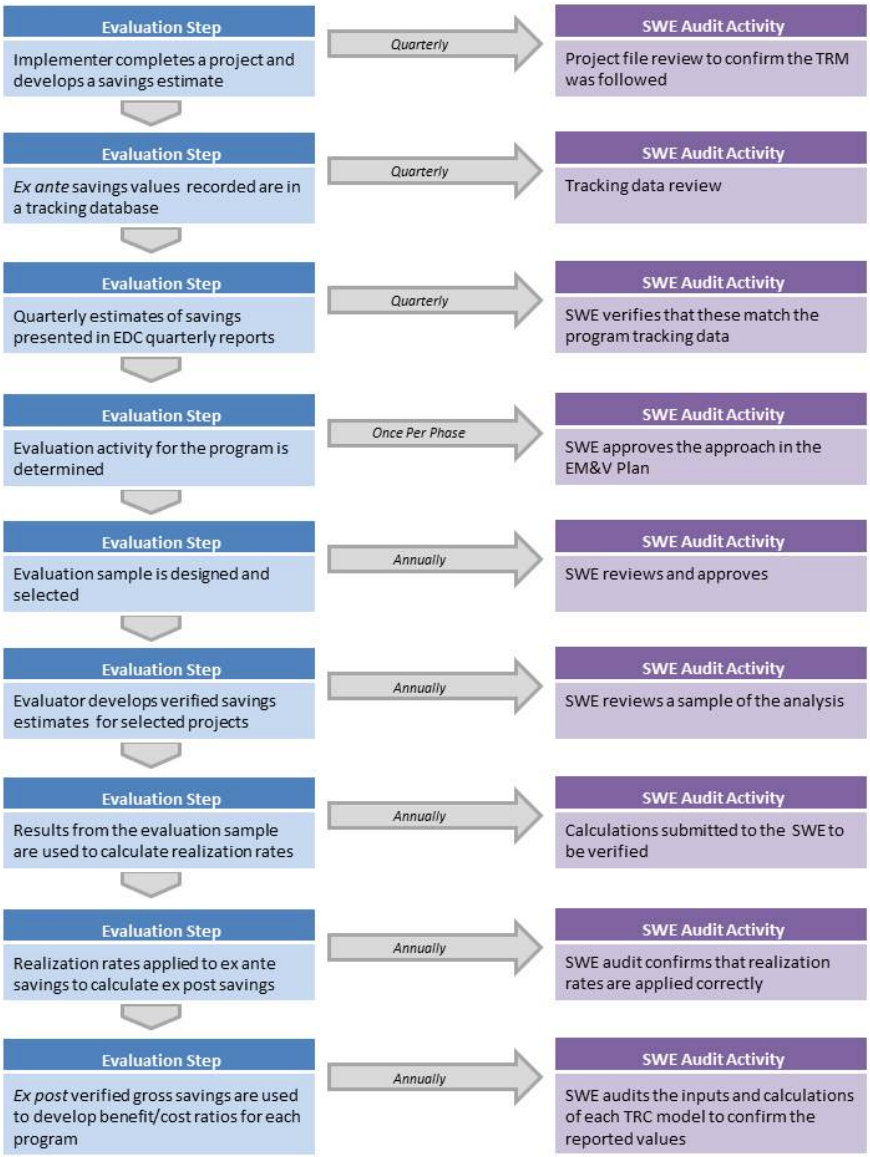
4 Statewide Evaluator Audit Activities

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 review/audit of 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

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

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.

Figure 4-1: SWE Audit Activities



4.1 EDC Report and SWE Report Schedule

The quarterly and annual reports defined by the PA PUC are one of the ways by which stakeholders are informed about the spending and 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 PA 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 PA PUC requires.[85]

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: Incremental and To-Date • Preliminary Verified Savings
Program Year X, Quarter 2	January 15	• Quarter 2 Report • Implementation and Evaluation Updates • Gross Reported Savings as of November: Incremental and To-Date • Preliminary Verified Savings
Program Year X, Quarter 3	April 15	• Quarter 3 Report • Implementation and Evaluation Updates • Gross Reported Savings as of February: Incremental and To-Date • Preliminary Verified Savings
Program Year X – Preliminary	July 15	• Quarter 4/Preliminary Annual Report • Gross Reported Savings as of May: Incremental & To-Date • Preliminary Verified Savings
Program Year X – Final	November 15	• Final Annual Report • Gross Reported Savings as of May 31 • Gross Verified Savings as of May 31

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

4.1.2 Statewide Evaluator Report Schedule

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

- Summarize program and portfolio progress to date for each EDC

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

- 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 EDC's 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 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 • 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 PA 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, so it is important that this foundation is solid.

4.2.1 Quarterly Data Request – Ex Ante

In an updated quarterly data request issued December 7, 2011,[86] 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.[87] All information provided in response to the 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 SharePoint site page.

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, as well as a sample of participant invoices, work orders, and receipts.

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 on the data request document.

For the *commercial programs*, the SWE requests five separate pieces of data. As for 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 separately from the sample drawn by the EDC evaluation contractor for the gross impact evaluation. 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.

[86] The data requests are available for the EDCs on the PA Act 129 SharePoint Site.
[87] In this context, the EDC preliminary annual report due July 15 serves as the Q4 report.

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 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 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 (5) 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 (5) 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 and will produce accurate and unbiased estimates of program impacts. The following list includes some of the issues the SWE will consider as it reviews survey instruments:

- Are the skip patterns adequately prescribed? 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

Fifteen (15) days after the submittal of an EDC's final annual report for a program year, the SWE must submit its annual data request. 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[88], EDCs 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?

[88] 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

EDCs 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 energy and capacity
- 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, and project files. 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,[89] EDCs 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
    - Savings calculations sheets
    - Other pertinent information

[89] Ibid

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 non-residential program, EDCs 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
- 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 should provide the following information for non-residential 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
    - 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 PA 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. 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 PA PUC to approve or reject the protocol. The SWE's recommendation should not be construed as PA PUC approval, as the PA 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** – Following the close of the third quarter (Q3) of each program year, the SWE will request 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. 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. 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, clearly showing the data collection process and how it is utilized in savings analysis
- Site-inspection findings
- 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.

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 non-residential programs[90]. 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.

2) **Independent Site Inspections (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. 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.

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

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 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:** 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.

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, whether it leads to a savings claim or is used primarily to assess market progress, explicate market engagement, or monitor a program's market engagement and transformation effects.

4.4.1 Research Design

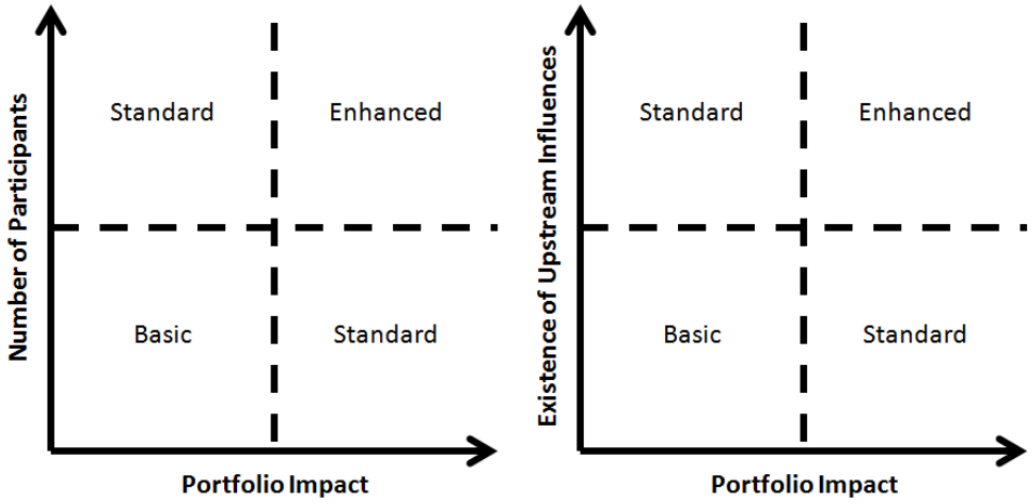
The audit will assess whether the approach used is consistent with the results, based on SWE-defined levels of rigor of analysis in the SWE Net-to-Gross Study Methods guidance document distributed to the EDCs and their evaluation contractors on February 27, 2012. The levels of rigor (basic, standard, 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 mid-stream 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.

Selection of method 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 if the sampling were appropriate. Probability sampling should be used for any net savings or market share/market effects studies.

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. In the past, information regarding the methodology, data collection, sampling, survey design, algorithm design, or analysis often was lacking. In certain instances, overall free-ridership or NTG ratios were presented with little 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 following process evaluation components.

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. See Section 3.4.2 for 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 to the population. 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 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 (5) business days per the guidelines below.

4.5.3.1 General Instrument Characteristics

The reviewers will ensure that the instrument includes:

- 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 reviewer 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).[91] 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 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.

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

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.
- 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 with TRC calculations have been performed according to the Commission's latest TRC Order. The SWE review of B/C ratios for these various groupings will help to identify the most cost-effective programs across the state. When possible, the SWE also will assess these B/C ratios by comparing them to cost-effectiveness calculations from other, similar best programs throughout the United States.

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 (PA PUC) Website

The SWE will provide documents for sharing on the PA PUC's public website[92], 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[93] to improve communication and coordination of activities among the SWE Team, the BTUS, the EDCs and their evaluator contractors, and the Energy

[92] 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
[93] The URL for the SWE SharePoint site is: https://sp.gdsassociates.com/sites/39701/SitePages/Home.aspx

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 PA 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[94]; RFP-2009-1).

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, 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.

[94] The PA PUC Bureau of Conservation, Economics, and Energy Planning (CEEP) is now known as the Bureau of Technical Utility Services (BTUS).

5.3 Program Evaluation Group (PEG) Meetings

The SWE will chair and set the agenda for monthly 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 PA 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 PA 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: 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 PA PUC regarding fixed and transportation utility regulatory matters, as well as serves 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: 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's 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 (MEASURES/GROUPS): Refers to a broad or sometimes narrower category upon which the program is concentrating. Examples of end-uses include refrigeration, food service, HVAC, appliances, building envelope, and lighting.

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: 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.

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.

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 that 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 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 PA 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 PA 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, who is conducting the evaluations of the Phase II Act 129 programs. Team members are: Nexant, Inc., Research Into Action, Inc., and Apex Analytics, LLC.

TECHNICAL REFERENCE MANUAL: 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 PA 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 create a standardized lighting audit and application template 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 6-1: 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, 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, as built
Line item quantity	ICSP inspection, acceptance of as-built
Line item hours of use	ICSP inspection, 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 large 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 6-1, 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 equal to or better than the NEMA Premium efficiency standard

Program managers should ensure that the motor end-user or motor installation contractor record 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 6-2: 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 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 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

HVAC equipment will be provided through participating installation contractors and dealers who also will install the HVAC equipment. The utilities should require the installation contractors to collect and record information about each project that receives an incentive through the program. This information is necessary to provide accountability and more accurately assess the programs' gross energy and demand savings.

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 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 6-3: 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
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 PA 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 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 6-4: 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, application stage Logged kW or proxy data Modeled savings
Line item quantity	Motor inventory form, application stage
Line item hours of use (not FLH or EFLH)	Motor inventory form, application stage
New Equipment Information	
Proposed motor nameplate data Drive nameplate data and efficiency, Post kW over representative loading period	Motor inventory form, application stage Logged kW or proxy data
Line item quantity	Motor inventory form, application stage
Line item hours of use (not FLH or EFLH)	Motor inventory form, 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 6-5: 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	Photograph and nameplate Information on application form (applicable to early retirement scenario only)
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
New Equipment Information	
Manufacturer and model Efficiency: SEER and EER (A/C), HSPF (ASHP), or EER and COP (GSHP) Capacity	Cut sheets
Quantity	Invoice
Location/EFLH	Address
Type, quantity, nominal efficiency, and set points of other	Statement on application form

heating and cooling systems	
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 participating 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 6-6: Residential New Construction Program Information to Be Collected and Verified

Site Information	Required Documentation[95]
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, PTAC, 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

[95] 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.

	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.

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 require the retailers to 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 6-7: 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 hot water heater

Clothes washer: Gas or electric hot 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 collect 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 6-8: 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 Pennsylvania TRM assumes that 84% of bulbs ultimately are installed in residential sockets. 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. 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 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 6-9 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 6-9: 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, 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 6-10 and Table 6-11 summarize the data that the program ICSP should collect, and that the EDC evaluation contractor should verify (for projects in the evaluation sample).

Table 6-10: 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 6-11: 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.