Regional Technical Forum

ROADMAP FOR THE ASSESSMENT OF ENERGY EFFICIENCY MEASURES

REGIONAL TECHNICAL FORUM

Release Date: December 8, 2015

TABLE OF CONTENTS

1. Introduction

In 1996, Congress directed the Bonneville Power Administration and the Northwest Power Planning Council to convene a regional technical forum to develop standardized protocols for verifying and evaluating conservation savings and to ensure that the region continues to meet the Council's targets for securing cost-effective conservation. In April 1999, the Council voted to charter the Regional Technical Forum (RTF) as an advisory committee to the Council. The document provides a roadmap for how the RTF operates to fulfill its mandate.

1.1. Purpose

The purpose of this document and the Guidelines it references is to describe how the RTF selects, develops and maintains methods for assessing the lifetime costs and benefits of energy efficiency measures. This document provides definitions of key concepts and describes the basic operating assumptions of the RTF. RTF members and all parties interested in conforming to RTF Guidelines should have a clear understanding of this roadmap before attempting to apply the referenced Guidelines and other supporting documents and tools.

1.2. Structure

This document provides an overview of the RTF Guidelines, various supporting documents and tools used in creating and managing those documents. The following three Guidelines are referenced throughout this document:

- Guidelines for the Estimation of Energy Savings
- Guidelines for the Estimation of Incremental Measure Costs and Benefits
- Guidelines for the Estimation of Measure Lifetime

Each of these Guidelines is referred to, respectively, as the Savings Guidelines, Cost Guidelines, and Lifetime Guidelines. They are collectively referred to as the Guidelines. Section 6 further describes the RTF documents and tools that support these Guidelines and the documents that are created in conformance with these Guidelines.

The relationship of this roadmap to the Guidelines and supporting documents and tools is illustrated in Figure 1.

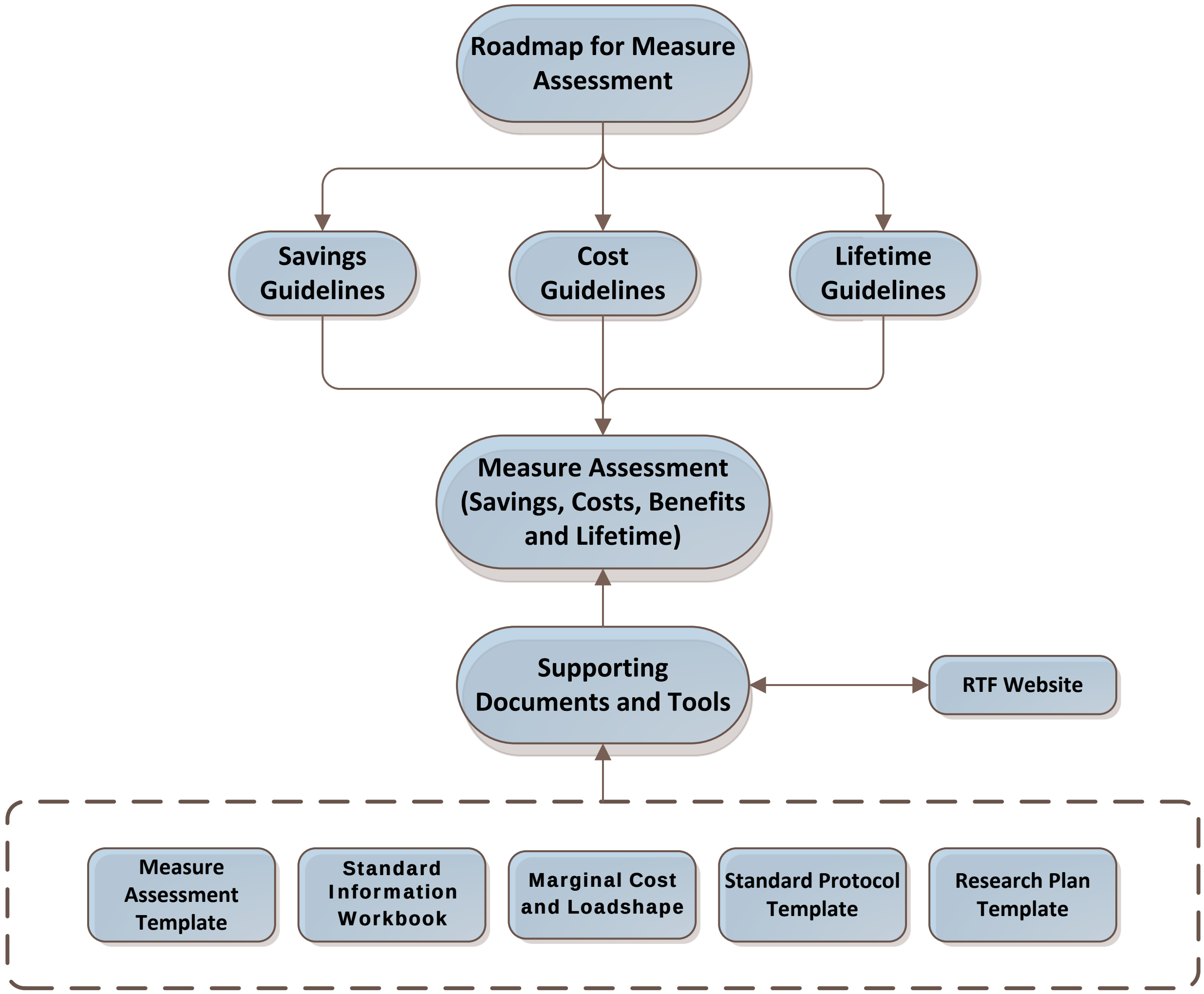

Figure 1: Structure and Relationship of RTF Documents and Tools

1.3. Key Concepts

The following key concepts are referenced throughout this document and the Guidelines. Their definitions are critical to the correct interpretation of the RTF's intent. In addition, each of the Guidelines contains its own unique concepts.

1.3.1. Measure

A measure is one or more changes in system configuration, equipment specifications or operating practices that reduces electric power consumption as a result of increases in the efficiency of energy use, production, or distribution. Measures may be further defined by their specific application. Specific measure applications may be defined by characteristics of the affected building, end use, system, equipment or location. For example, wall insulation may be applied to single-family residences, with basements, located in the climate zone west of the Cascade Mountains. Specific measure applications may be further distinguished by the method of a measure's delivery. For example, efficient showerheads for single-family residences may be delivered via mail-by-request, retail outlets or direct installation.

1.3.2. Savings

Savings is defined as the difference in annual energy use between the baseline (see section 3.2) and post (after measure delivery) periods, which is caused by the delivery of a measure. The terms "net" or "gross" are intentionally not used to modify the term "savings," as they may conflict with the definition of "baseline," provided in section 3.2. The current practice baseline defines typical choices of eligible end users, as dictated by codes and standards and the current practices of the market. The most important conflict would arise if savings were estimated against a current practice baseline and then those savings were further adjusted by a net-to-gross ratio, where the net-to-gross ratio was the probability that the measure would have been delivered in the absence of program influence.

1.3.3. Lifetime Savings

Savings may vary over the lifetime (see Lifetime guideline) of a measure. For measures with a current practice baseline the savings estimate should apply throughout the period between measure delivery and the end of the measure lifetime. If the remaining useful life (RUL) of a pre-conditions measure is expected to be ten years or less, then two baselines must be used in estimating lifetime savings. The first baseline applies between measure-delivery and when the RUL of the pre-condition expires. The second baseline applies between expiration of the RUL and the end of the measure lifetime. For example, an air compressor might be scheduled for replacement in three years, but is replaced sooner with a more efficient model. The lifetime of the efficient air compressor might be twenty years; however, the RUL would be three. The first baseline applies to years one through three. A second baseline is applied in years four through twenty. Further, description of these baselines is provided in section 3.2.

1.3.4. Measure Interactions

The savings from one measure may in part be determined by whether another measure has already been delivered to the end user, for example, lighting wattage reductions can change the savings from lighting controls or the installation of an electronic thermostat can change the savings from a ductless heat pump. A measure interaction is significant if the RTF determines that it could change a measure's savings estimate by more than ± 10%.

1.3.5. Measure Assessment

Measure assessment is the analysis of an energy efficiency measure by which estimates of savings, costs, benefits, and lifetime are determined and proposed to the RTF for approval and determination of measure cost-effectiveness.

1.3.6. ProCost Model

ProCost is a spreadsheet tool, developed by the NW Power and Conservation Council, which computes Regional measure lifecycle cost-effectiveness. ProCost uses Regional economic and power system assumptions that are updated with each Council power plan.

1.3.7. Sunset Date

The RTF will establish a sunset date for certain types of measures. By this date, the RTF will either approve an update for a measure (and a revised sunset date) or deactivate the measure.

1.3.8. Savings Reliability

An estimate of savings from a measure is reliable if the errors associated with sampling, data collection or modelling is sufficiently small and unbiased. Sufficiency is determined by the collective opinion of the RTF.

1.3.9. Diligent Review

The RTF uses estimates of parameters, e.g., average length of a residential shower or heat/cool interaction factors, from studies performed by other agencies in estimating measure savings. The RTF must diligently review a study before approving the use of these values in the estimation of measure savings. A diligent review will include, but is not limited to understanding the characteristics of the sample studied, the study's data collection methods and analysis methods, and the variability of the parameter estimates across the study sample. A diligent review will consider whether the sample is applicable to measures delivered in this region and if not, whether it is feasible to normalize the results for application to this region.

1.3.10. Best Practice Savings Estimate

A best practice savings estimate is one that relies on the best practical and reliable data collection and estimation methods. Practical means that the required data collection and estimation can be carried out with proven techniques and resources deemed reasonable by the RTF. Best practice savings estimates may rely on parameter values, e.g., average length of a residential shower or heat/cool interaction factors, from studies performed by other agencies, if the RTF has determined, following a diligent review, they are sufficiently reliable.

1.3.11. Program

A program is a collection of strategies designed to cause delivery of one or more measures to end users in one or more eligible market segments. Strategies may target end users, mid-market actors, or up-stream market actors or some combination of these. Program strategies may include education, training, technical advice, financial incentives and assistance with the design, installation, operation or maintenance of measures.

1.3.12. Program Operator

Any agency that delivers energy efficiency measures in the Pacific Northwest, including individual utilities, Bonneville Power Administration (BPA), the Energy Trust of Oregon (ETO) and the Northwest Energy Efficiency Alliance (NEEA) or third-parties operating programs under contract to these agencies.

1.3.13. Delivery Verification

Delivery verification is the process by which a program operator confirms that a delivered measure conforms to the measure's specification. The RTF provides delivery verification requirements for each measure. These requirements enumerate what information should be gathered to confirm delivery. The RTF generally does not specify how this data is to be gathered, but it may specify how to gather this data if such guidance is needed to assure the reliability of measure savings. Otherwise, it is the responsibility of program operators to determine how to gather this information.

1.3.14. Research Strategy

A research strategy outlines a possible approach to the data collection and analysis needed to support an RTF decision on whether to approve as Proven a savings estimate or a savings estimation method. The goal of all research strategies is to describe least cost methods of obtaining the required results. A strategy must include estimates of the required sizes of any new data collection efforts. All strategies must include an estimate of the likely range of costs for the research. A Planning measure (see Section 2) requires an approved research strategy.

1.3.15. Research Plan

A research plan describes in detail the data collection and analysis that will be performed to support an RTF decision on whether to approve as Proven a savings estimate or a savings estimation method. A research plan is required prior to approval of a Provisional measure (see Section 2). The plan will include the sample design (sampling frame, domains of study, stratification, quotas, selection method, replacement procedure and expected sampling precision) to be used for any new data collection efforts. The plan will also describe how new and existing data will be used to develop sufficiently reliable estimates of measure savings or to prove the reliability of a savings estimation method. A research plan only describes funded data collection and analyses.

2. Measure Classification

Measures are classified by the RTF according to the method used in estimating savings. Four savings estimation methods are defined: Unit Energy Savings (UES), Standard Protocol, Custom Protocol and Program Impact Evaluation. It is the RTF's intention that each method will produce savings estimates of comparable reliability sufficient to meet the needs of regional energy planners. These methods are also expected to support regulatory processes related to the adoption and planning of energy efficiency programs.

- **Unit Energy Savings (UES)**. The UES method is appropriate for measures whose unitized savings, e.g., savings per lamp or motor, is stable (both the mean and variance) and can be reliably estimated throughout the period from RTF approval to the measure's sunset date (section 3.5).
- **Standard Protocol**. A Standard Protocol method is appropriate when savings from a measure are widely varying, e.g., variable frequency drives and transformer de-energizing, but can be determined by a standardized procedure for data collection and analysis that is applicable to many different end use sites.
- **Custom Protocol**. Custom protocols are appropriate for measures that require site-specific savings estimation planning, data collection and analysis in order to develop a reliable estimate of savings, e.g., tankless electric on-demand water heater for hospital laundry.
- **Program Impact Evaluation**. Program impact evaluations estimate savings from a period of program operation. Program impact evaluations involve the analysis of a reliable sample of program participants (and possibly non-participants) to determine the savings.

For UES measures, the RTF approves the method for estimating savings and the savings values for each measure application. For the Standard Protocol method, the RTF only approves the savings estimation method. Both may be approved within any of the following four categories. Before a measure is approved for one of these categories[1] it must meet the savings estimation quality standard for the category as described in the Savings Guidelines.

- **Proven**. Proven UES savings and Standard Protocol methods are those that the RTF considers reliable. As they are considered reliable, the UES savings values may be multiplied by the verified number of delivered units to obtain an estimate of program savings in an impact evaluation (see Savings Guideline). Impact evaluations apply a faithful application standard in determining the savings from Proven Standard Protocol measures.
- **Provisional**. Provisional UES savings and Standard Protocol methods requires a funded research plan, approved by the RTFs.

[1] An existing measure may be approved in any of these categories when its savings are updated. For example, a Planning or Provisional measure may be become a Proven measure. It is also possible for a Proven measure to become Provisional, Planning or Small Saver due to changes in market conditions or product standards.

- **Planning**. Planning UES savings and Standard Protocol methods require an RTF approved research strategy.
- **Small Saver**. The RTF may determine that the regional technical potential savings from a measure are too small to warrant the resources needed to meet the quality standards defined for Planning, Provisional or Proven measures. Such measures are categorized as Small Saver. These guidelines do not restrict the scope of a measure. For example, all types of insulation applicable to single- family buildings can be included in a weatherization measure. Alternatively, each type of insulation, e.g., wall or ceiling could be considered a separate measure. How broadly or narrowly a measure is defined can be determined by a number of factors, including but not limited to scope of supporting data, common estimation methods or applicable codes and standards. However, the RTF will consider the scope of each measure to ensure that the scope has not been arbitrarily restricted in order to qualify for the less restrictive quality standards applied to measures in the Small Saver category.

The Savings Guidelines also define two other savings estimation methods: Custom Protocol and Program Impact Evaluation. For these methods, the RTF provides guidance on how to estimate savings, but it does not approve savings estimation methods. Rather it leaves this to the discretion of program operators.

The Savings Guidelines describe a savings estimation quality standard that applies to all Custom Protocol measures, thus no additional categorization of these measures is required. Guidance is also provided on how to conduct program impact evaluations for all classifications of measures and the appropriate treatment for other UES measures that are delivered by program operators but which have not been approved by the RTF.

Measure status is the final level of classification. It applies to all UES and Standard Protocol measures. Possible statuses are Active, Under Review and Deactivated.

- **Active.** The RTF approves an Active status for a measure if it currently conforms to the Savings Guidelines quality standard for its category (Proven, Provisional, Small Saver or Planning).
- **Under Review**. The RTF may place an Active measure Under Review if there are indications that the savings estimation method will need to be updated in order to continue to be in compliance with the Savings Guidelines, or if the sunset date is approaching.
- **Deactivated.** The RTF may deactivate a measure if it is out of compliance with the Savings Guidelines and the RTF does not believe that it is of sufficient value to the region to warrant the resources needed to bring it into compliance.

The classification system used by the RTF is defined by the Savings Guidelines. The Cost and Lifetime Guidelines apply to all measures regardless of their Savings classification.

The RTF’s system for classifying measures is illustrated in Figure 2.

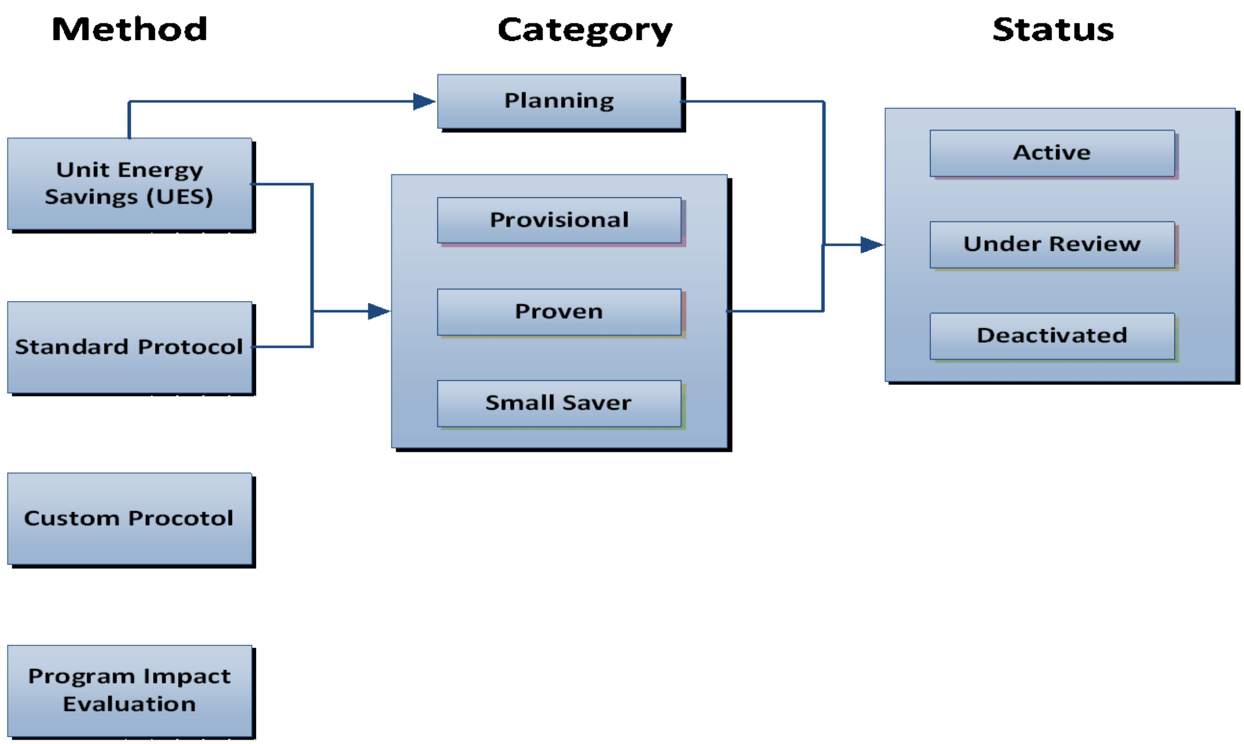

Figure 2: System for Measure Classification

3. MEASURE SPECIFICATION

The measure specification describes all eligible applications of the measure, which may be defined by characteristics of the affected building, end use, system, equipment, location and method of delivery. In addition, the measure specification defines the appropriate baseline conditions for each measure application, along with relevant implementation and product standards. Finally, the specification defines the measure sunset date.

These specifications must be provided so that an informed decision can be made about the most appropriate savings estimation method. This information also helps the RTF select the best strategies for developing the information required for savings estimation.

Measure specifications should be prepared for all measures regardless of the method used in estimating savings. For UES and Standard Protocol Measures, the RTF must have this information before approving the measure. For Custom Protocol measures, the RTF expects program operators to provide this information as part of each measure’s savings estimation report. The RTF also expects program operators to require this information for all Program Impact Evaluation measures prior to approving their delivery under any program, including other UES measures not approved by the RTF.

The information needed to fully specify each measure is as follows.

3.1. Measure Identifiers

Measure identifiers are the characteristics that uniquely identify each eligible measure application for Unit Energy Savings measures[2]. Measure identifiers must be clearly described and limited to those characteristics data that can be reliably obtained by the programs so that the correct savings are estimated for each delivered measure. Separate savings values may be estimated for specific applications of a measure. For example, the clothes washer measure may have separate UES for water heating fuel type and dryer fuel type. In addition, the specific applications may be identified by end user, equipment, and program design characteristics. These may include, but are not limited to, heating or cooling climate zones, heating and cooling system types, delivery method, size range, and efficiency category.

Measure identifiers define relatively homogenous applications of the measure with respect to savings or the method used in estimating savings. In addition, separate applications of a measure may be defined in order to achieve homogeneity of measure costs or measure lifetime or the methods used to estimate cost or lifetime. The RTF will try to minimize the number of measure applications, focusing on those identifiers that cause significant differences in savings (± 10%), but maintaining separate applications as needed in determining measure eligibility or reliably representing differences in cost, benefits or lifetime.

[2] Eligibility criteria and key determinants play a similar role for Standard Protocol and Custom Protocol measures, respectively.

To the extent practical, measure identifiers should be used to account for significant measure interaction. For example, the savings from ceiling or floor insulation will vary significantly depending on whether the home is heated by heat pump or electric resistance. These measures should include an identifier of heating system type so that separate savings are estimated for electric resistance and heat pump systems.

3.2. Savings Baseline

Measure savings must be determined against clearly defined baseline conditions. Each measure application must be associated with one of the following definitions of baseline conditions.

3.2.1. Current Practice[3]

A current practice baseline is used if the measure affects systems, equipment or practices that are at the end of their useful life or for measures delivering new systems, equipment or practices, e.g., ENERGY STAR® specifications for new homes. There are a number of possible indicators that current practice is the appropriate baseline:

- Measure is delivered as part of a new construction project or is subject to the requirements of current state and local building codes or federal standards, including major renovations that are covered by codes and standards.
- Relevant equipment is no longer operable and must be replaced
- Equipment is old and due to increasing frequency and difficulty of repairs and maintenance the end user has firm plans to replace the equipment
- Equipment must be replaced due to regulatory requirements, such as those promulgated by the US EPA (Environmental Protection Agency)
- Existing equipment cannot serve the end user's likely near-term loads

For these measures, the baseline is defined by the typical choices[4] of eligible end users in purchasing new equipment and services at the time of RTF approval. The RTF estimates this baseline based on recent choices of eligible end users in purchasing new equipment and services. These choices may be inferred from data on shipments, purchases (equipment or services) or selected design / construction features. For example, the baseline for more efficient televisions is the average efficiency of recent television shipments. The period between RTF approval and the sunset date should be shortened as needed to reliably estimate savings for a measure whose baseline is rapidly changing.

[3] The RTF also uses the term "new" to indicate the measure has a current practice. Other terms for current practice baseline used outside the RTF include new construction, normal, replace on burnout, or natural replacement baseline.

[4] For Custom Protocol measures, energy savings estimates are specific to a single end user. Therefore, the choices of individual end users define typical current practice conditions for Custom Protocol measures.

The RTF may determine that current state and local building codes or federal standards provide a reliable definition of the baseline for these measures. As a general rule the RTF will use a baseline that is characterized by current market practice or the minimum requirements of applicable codes or standards, whichever is more efficient. The RTF may decide to use an alternative current practice based on other factors.

3.2.2. Pre-Conditions[5]

A pre-conditions baseline is used when the measure-affected system, equipment or practice still has remaining useful life (RUL). The baseline is defined by the typical conditions of the affected system, equipment or practice at the time of RTF approval. The RTF estimates this baseline based on data from recent adopters, or if there has been no significant adoption, it uses data from the typical conditions found among eligible end users[6]. For example, the baseline for agricultural motor replacements is defined by the average efficiency and operating hours of in-service agricultural motors.

If the estimated RUL is longer than ten years, then the RTF will assume that RUL equals measure lifetime. If RUL is ten years or less, savings, costs, and benefits should be estimated for the following time periods: a) between the time of measure delivery and when the RUL expires, and b) between the time when the RUL expires and the measure lifetime expires. The Measure Assessment Template contains a separate sheet (SummaryRUL) where the assumptions made and analysis performed to forecast savings, costs, and benefits after the RUL expires should be documented. The RTF expects careful consideration of savings, costs, and benefits during this period, but does not impose any specific quality standards on these estimates beyond the use of best available data and professional judgment.

3.3. Implementation Standards

Measures may involve equipment, practices or both. Whatever the nature of the measure, there must, be standards that govern its implementation. These standards must specify procedures for measure implementation and delivery verification and may include provisions for independent third party quality assurance. Training required for staff that performs any of these functions should be specified. These standards should also address, as appropriate, installation procedures, equipment sizing and ratings, maintenance procedures and expected operating conditions and practices. These standards must be clearly documented so that they can be correctly accounted for in the estimation of savings, costs, benefits and lifetime.

---

[5] The RTF's use of the terms upgrade, replacement and conversion to indicate a measure has a pre-conditions baseline. Other terms for pre-conditions baseline used outside the RTF include early replacement and retrofit baseline.

[6] For Custom Protocol measures, energy savings estimates are specific to a single end user. Therefore, existing conditions of individual end users define typical existing conditions for custom measures. This is also true for Standard Protocol measures that rely on site-specific measurements of baseline equipment performance.

3.4. Product Standards

Some measures involve equipment or building components that must meet or exceed certain performance specifications. These performance specifications may be substantiated by standardized test procedures. These specifications must be clearly documented so that they can be correctly accounted for in the estimation of savings. Some measures may encompass a range of specifications, such as the seasonal energy efficiency rating (SEER) for cooling equipment, with each level of the rating defining a specific measure application. Separate savings values may be required for each variation on the specification.

3.5. Sunset Date

As part of approving a UES or Standard Protocol measure, the RTF will establish a sunset date for the measure. This date should not be more than five years past the measure's approval date. By this date, the RTF must change the sunset date or change the sunset date and approve a change in the measure's classification. The RTF will record the factors that justify each sunset date such as expected revisions to energy codes, federal standards, completion of RTF approved research plans, or shifts in current practices of consumers.

Research plans are approved for each Provisional measure. These plans must be approved by the RTF and must include a date by which the research is to be complete. The sunset date for these measures should be no later than the completion date of the research. The RTF may set the sunset date to a major milestone of the research, expected before the completion date, and then choose to extend the sunset date only if that milestone was achieved and the research appears to be proceeding toward a successful conclusion.

Sunset date is not appropriate for Custom Protocol measures as by definition they are delivered only once to a specific end use site. However, program operators need to consider similar factors, such as energy codes and federal standards in determining whether a proposed Custom Protocol measure is eligible for their programs.

3.6. Delivery Verification Guidance

As part of approving a UES and Standard Protocol measure, the RTF will provide guidance on delivery verification. The guidance should describe key data that needs to be collected as part of delivery verification. This guidance will depend on factors included in the savings estimation method. For example, for retail delivery of lighting measures, where the storage rate and installation rates are embedded in the savings estimate, the needed data would be sales of the product by the retailers. For other measures, delivery verification may require data collected by visual inspection or physical measurement. For example, measures involving commissioning controls and equipment sizing where the savings estimate does not assume a rate of specification compliance. For these measures, a site inspection would be needed to determine the rate of compliance with the specification. The delivery verification guidance should be documented as part of the measure specification for UES measures. For Standard Protocol measures, the guidance is provided by the protocol document.

4. Measure Assessment

Measure assessment is the process by which data are collected and analyzed to estimate energy savings, costs, benefits and lifetime and how these estimates are used in estimating cost-effectiveness. The process also encompasses diligent review of relevant data sources and estimation methods, as described in the Guidelines, and the preparation of complete and transparent documentation of methods and data sources. As needed, this process may also include research planning. Research will be required to create or update savings estimation methods or input parameter values for measures in the Provisional category. Research planning will also be required for program impact evaluations and market studies that provide data on current market conditions, e.g., the Residential Building Stock Assessment, or that provide data on the characteristics of recently adopted measures needed to account for significant measure interactions.

Documentation, transparency and access are key attributes of a measure assessment that conforms to RTF Guidelines. All methods and supporting data sources must be clearly documented so that results can be reproduced by members of the RTF and other agencies. The RTF documentation expectations for each type of estimate (savings, costs, benefits and lifetime) are described in the Guidelines. For energy savings, the documentation must be adequate to support RTF approval of a measure for a specific category and status. For measure costs, benefits and lifetime, the documentation of data sources must be sufficient to achieve the transparency and accessibility expectations, but does not have to meet specific quality standards.

Measure assessment may be conducted separately by the RTF or in cooperation with program operators. Program operators may also conduct measure assessments, such as for other UES measures not approved by the RTF but delivered by the program operator and included in program impact evaluations.

For UES and Standard Protocol measures, the RTF estimates Regional measure lifecycle cost-effectiveness using the ProCost lifecycle cost-effectiveness calculator (ProCost). ProCost uses Regional economic and power system assumptions that are updated with each Council power plan. ProCost measure-specific input requirements define the estimates of energy savings, capital and O&M (operations and maintenance) costs, benefits, RUL, and measure lifetime that must be developed by the RTF for each UES and Standard Protocol measure. Program operators, following the RTF Guidelines, estimate values for these same parameters for Custom Protocol and Program Impact Evaluations measures and, using calculations similar to those in ProCost, estimate lifecycle cost-effectiveness with economic and power system assumptions tailored to their own service territories.

Figure 3 illustrates the measure assessment process.

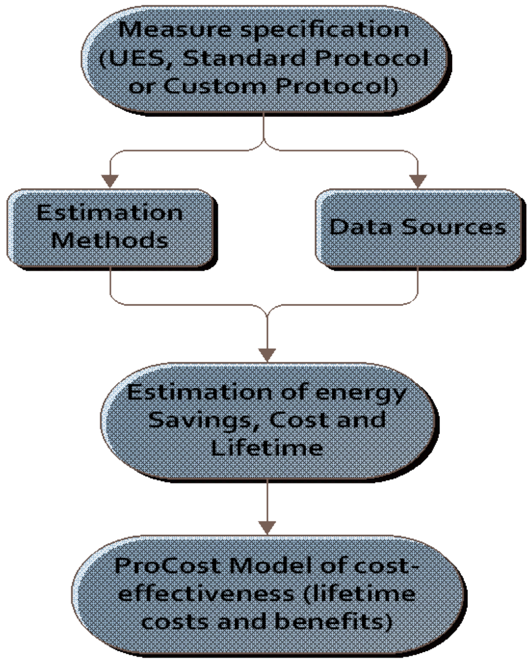

Figure 3: Measure Assessment Process

5. Measure Lifecycle Cost-Effectiveness

A measure is cost-effective if its delivery results in benefits that exceed costs when these benefits and costs are counting throughout the measure's lifetime. The RTF uses the ProCost model to account for all benefits and costs for each measure. Inputs to this model include electricity savings, measure lifetime, impacts on other fuels, capital costs, operations and maintenance costs, and potentially many other costs or benefits as described in the Cost and Benefit Guideline. In addition, there are a number of inputs that derive from the most recent NW Power and Conservation Council power plan, such as the marginal cost of new energy supplies and the discount rate.

The NW Power and Conservation Council develops a new power plan every five years. Each plan documents the methods used to develop all inputs to ProCost and the methodology of conducting the cost-effectiveness analysis performed on the measures included in the plan. Once each plan is finalized the RTF and council staff collaborate on revising the document entitled *RTF Implementation of Council Cost-Effectiveness Methodology*. This document explains how the many possible futures analyzed for the power plan are reduced to a single future that is used to formulate avoided costs and risk factors, and other parameters used in ProCost. Shortly after the Council adopts each plan, the RTF approves the new version of the *RTF Implementation of Council Cost-Effectiveness Methodology* for use in conducting measure assessments.

6. RTF OPERATIONS

The RTF consists of voting members, corresponding members and staff. As needed, the RTF forms subcommittees to explore technical topics in greater detail, such as the development of a new Standard Protocol Measure or updates to one of the Guidelines. In addition to the subcommittees, the staff collaborates with many parties in developing presentations and proposals for consideration by the RTF. On a monthly basis, the RTF convenes to consider, and take action as needed, on the products of the subcommittee and staff deliberations. The work of the RTF, its subcommittees, staff and collaborating parties falls into the following four categories:

- Development and maintenance of Guidelines
- Development and maintenance of efficiency measures
- Research planning
- Work plan development

6.1. Guidelines

The RTF develops and maintains this document and the series of Guidelines (Savings, Cost and Lifetime) that it references. As part of this effort, the RTF develops and maintains supporting documents and tools, such as the RTF Measure Assessment Template. These supporting documents and tools are further described in section 6.

6.2. Measures

The RTF conducts measure assessments for UES and Standard Protocol measures. For Custom Protocol and Program Impact Evaluation measures, the RTF provides guidance that is more general and relies on program operators to implement this guidance in conducting measure assessments.

6.2.1. New Measures

Any party may propose a new UES or Standard Protocol measure for consideration by the RTF. Measures specifications (section 3) and other documentation, as described in the Guidelines, must be provided to support classification of the measure, determine what work is required to complete the measure’s assessment in the conformance with the Guidelines, and determine the measure’s priority and treatment in the RTF work plan. Work needed to complete the measure’s assessment may be conducted by RTF staff, the proposing party, program operators or others. Once complete it will be presented for approval by the RTF voting members. As part of the RTF’s approval the measure is assigned a sunset date, along with a description of the factors that justify that date, which will determine when the first review and update of the measure will occur. On approval, the measure status is Active.

6.2.2. Existing Measures

The RTF commences review and update of measures in advance of their sunset date. The review determines what work is needed to maintain the measure in compliance with the Guidelines. Based on recommendations from the review, the RTF may decide to change the measure's status to Deactivated or Under Review, or maintain Active status and set a new sunset date. For measures under review, the objective is to update the measure and return it to Active status prior to the sunset date. Once the update process is complete, the measure is presented to the RTF for approval and a new sunset date is set. This process continues until the RTF determines that update is not feasible or the measure is no longer relevant to the region's program operators, and the RTF changes the measure's status to Deactivated.

6.3. Research Planning

The RTF plays a role in regional research planning which supports the development of new measures, updates to existing measures and the conduct of program impact evaluations.

6.3.1. Research Strategies

The RTF develops research strategies (see section 1.3.14) which outlines a possible approach to the data collection and analysis required to support a decision on whether to approve a measure (UES or Standard Protocol) in the Proven category. A research strategy is required prior to RTF approval of a Planning measure.

6.3.2. Research Plans

The RTF assists other parties in developing research plans (see section 1.3.15) that detail how research strategies will be implemented, including how they are to be funded. A research plan is required prior to RTF approval of a Provisional measure (UES or Standard Protocol).

6.3.3. Market Studies

RTF measure assessment work relies to a substantial degree on the availability of studies that characterize the region's energy consumers and the businesses that provide products and services to these consumers. Examples include the residential and commercial stock assessments and non-residential new construction baseline studies. The Residential Building Stock Assessment, for example, collected data on housing size, envelope construction, appliances and other features available from the Residential Building Stock Assessment. This information was then used to develop calibrated SEEM prototypes. These prototypes were then used in developing savings estimates for many residential measures. Research may be needed to provide data on the characteristics of recently delivered measures. Such studies may also provide data needed to account for significant measure interactions.

The RTF may play a role in identifying the need for market studies. In addition, the RTF may provide input to the parties conducting market studies to ensure that the data needed to support measure assessments will be gathered.

6.3.4. Program Impact Evaluation

The RTF provides guidance to the region's program operators on how to conduct program impact evaluation. This research supports measure assessments for all classes of measures. The results of this research are frequently used by the RTF in developing new or updating existing RTF approved UES and Standard Protocol measures. This research may also provide Guidelines-compliant saving estimates for other UES measures not approved by the RTF. By request, the RTF may provide review and comments on research plans for program impact evaluations.

6.4. Work Plan Development

The RTF adopts a work plan each year that allocates resources among its various operations. The work plan also sets expectations concerning the Guidelines, measure and research planning work that will be completed and the associated decision making. The RTF considers the following requirements when forming each year's work plan:

- Work in progress on existing measures.
- Measure review and update work that will be required by sunset dates.
- Backlog of new measure proposals, and information from parties about their intent to propose other new measures.
- Work required to develop research strategies for Planning measures and research plans for Provisional measures.
- Assistance that other parties need in formulating research plans for Provisional measures.
- Assistance that program operators need in formulating plans for market studies and program impact evaluations.

The RTF sets priorities and allocates budget to support these various requirements based on the value of the associated measures in meeting short- and long-term conservation goals of the region's program operators.

Figure 4 illustrates the RTF operations and how they contribute to forming each year's work plan.

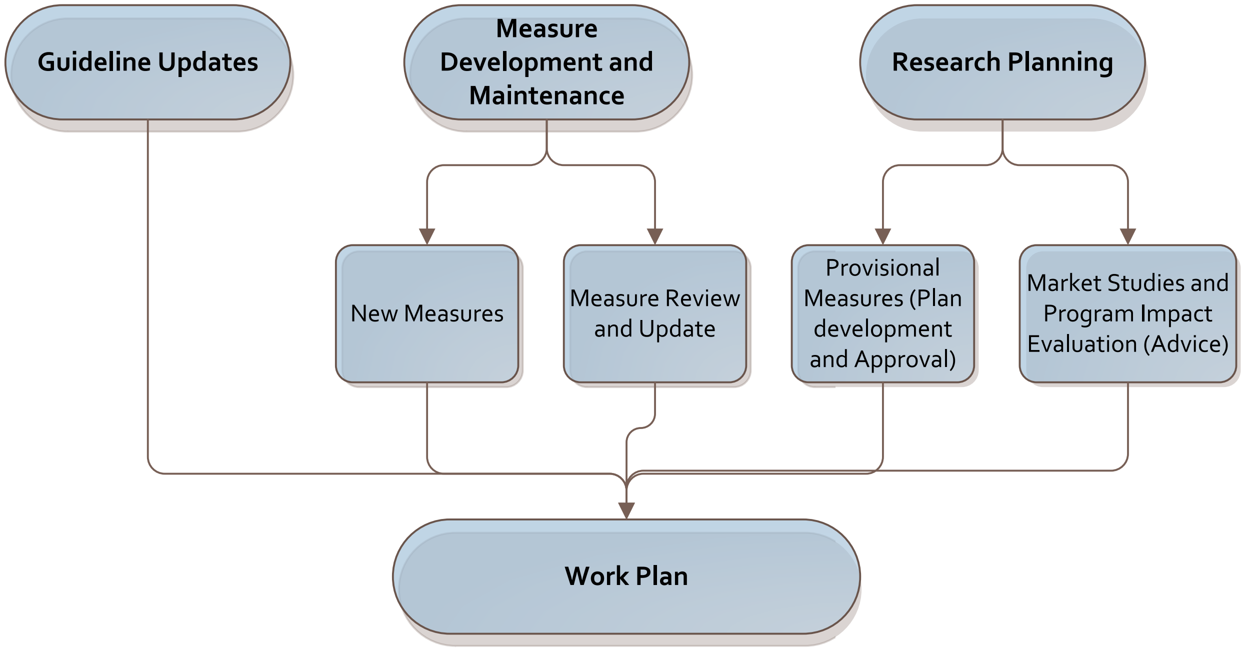

Figure 4: RTF Operations and Work Plan Development

7. RTF Documents and Tools

The RTF maintains a collection of documents and supporting tools that define its measure assessment process and support the completion of specific measure assessments. This document and the referenced Guidelines (Savings, Cost and Lifetime) describe how measure assessments are conducted. The RTF also maintains specifications, e.g., Weatherization, Performance Tested Comfort Systems and Network PC Power Management, which have been reviewed and approved by the RTF. These may be referenced when specifying a measure.

The RTF maintains a set of tools that are used in completing measure assessments in conformance with the Guidelines.

- **Measure Assessment Template**. This Excel workbook is used for all classes of measures. It contains a Summary sheet that documents the measure specifications, estimation methods (savings, costs, benefits and lifetime) and the data sources that support these estimates. An example Summary sheet is included in the template for UES, Standard Protocol and Custom Protocol measures. Also included is the Checklist sheet, which documents the major milestones that are relevant to the development and maintenance of measures in conformance with the Guidelines. This template includes ProCost and sheets needed to supply and document input and output values for this cost-effectiveness model. It includes a SummaryRUL sheet where the assumptions made and analysis performed to forecast savings, costs and benefits after the RUL expires (measures with Pre-Conditions baseline) are to be documented.
- **Standard Information Workbook**. This Excel workbook provides RTF approved generic[7] values and supporting documentation for various parameters needed in completing measure assessments. For example:
    - Hourly costs for various labor categories that can be used in estimating capital and O&M costs for a measure.
    - Lifetime Reference Table, which may be used as a starting point in estimating measure lifetime.
    - Commonly used input parameters for determining UES savings, including weightings used for heating/cooling zones, house sizes, standard SEEM inputs, lighting hours of operation, and water-heater temperature set points.

  Typically, the Standard Information Workbook will be updated after a Power Plan is released. The RTF will make additions, corrections, and more frequent updates to the workbook as needed. The RTF maintains these values and posts the current version of the Standard Information Workbook on the RTF website.

[7] Where data or methods that are more appropriate to the measure are available, they should be used instead of those specified in the Standard Information Workbook and their use documented.

- **Marginal Cost and Load Shape (MC and Loadshape)**. This workbook contains marginal energy costs and measure load shapes. These are used by ProCost in estimating measure cost-effectiveness.
- **Standard Protocol Template**. This Word document provides a template, by example, for describing a Standard Protocol measure. A Standard Protocol must also have a completed Summary and Checklist sheets (in the Measure Assessment Template) and a calculator (as described in the Savings Guidelines). In addition, a research plan is required for Provisional Standard Protocol measures.
- **Research Strategy and Plan Templates**. These Word documents provide templates for documenting either a Research Strategy or Research Plan. Such strategies or plans are required for the approval of Provisional or Planning UES measure. Research Plans are required for approval of Provisional Standard Protocol measures.

The RTF completes assessments for UES and Standard Protocol measures, using the Guidelines and the supporting documents and tools described above. Microsoft Word®, Excel®, PowerPoint® and other files associated with each measure's assessment are maintained on the RTF website. The website provides public access to current and prior versions of each file associated with a measure assessment. It also tracks the RTF meetings where a measure is presented and discussed and provides access to meeting minutes that summarize the discussion and decision-making. The website provides many examples of documents that can be modeled in preparing various documents required by the RTF's decision-making process such as:

- UES review recommendation memos
- UES review/update presentations
- Provisional UES measure research plans
- Standard Protocol calculators
- Research strategies and plans for planning and provisional UES measures
- Research plans for provisional Standard Protocol measures

8. Developing an RTF Approved Measure

Four methods for estimating measure savings were described in section 2:

- Unit Energy Savings (UES)
- Standard Protocol
- Custom Protocol
- Program Impact Evaluation

The RTF approves measures that use the first two of these methods. For the last two methods, the RTF provides guidance on how measures assessments are conducted by program operators but does not approve those methods.

Approval for the first two methods occurs once the measure assessment is complete and the RTF has reviewed the resultant analysis and documentation of savings, costs, benefits, RUL, lifetime and cost-effectiveness. Approval occurs first when the measure is created. Approval may occur one or more additional times as the measure's estimation method goes through periodic reviews and updates in accordance with sunset dates set by the RTF.

8.1. Creating a New Measure

New measures may be proposed by any party for consideration by the RTF, using the RTF Proposal Submission Form. The measure's proponents must provide documentation of the measures specification, as described in section 3. The RTF determines whether the measure merits consideration, and the method and category for which its approval will be sought. Measure methods and categories are described in section 2.

The RTF decision on the appropriate method and category is critical. It determines how the measure assessment is conducted and the quality standards that must be met to achieve RTF approval. The quality standards associated with each method and category are described in the Savings Guidelines. A measure may be proposed as a Small Saver (UES or Standard Protocol), in which case, the quality standards are relatively low. However, the RTF must be convinced that the regional technical potential savings from the measure are, in fact, small. Approval in the Provisional and Proven categories may also be sought for either method, but they are associated with progressively more stringent quality standards.

The measure assessment can begin once the measure has been specified and assigned to a method and category. The assessment must provide the following products for review and approval by the RTF.

- **UES Measures**. For UES measures both estimation procedures and values are approved (savings, costs, benefits, RUL, lifetime and cost-effectiveness).
    - **Small Saver**. The proponents must provide the RTF with a completed Measure Assessment Template that fully documents the measure specification and the estimation procedures for savings, costs, benefits, RUL, lifetime and cost-effectiveness.

Presentation materials must also be prepared and used in presenting the case for the measure's approval. The assessment workbook and presentation will be reviewed by an RTF subcommittee or by RTF staff prior to RTF consideration, to ensure the analysis and documentation are acceptable and the measure satisfies the quality standards for this category. The quality standards for this category only require sound engineering and statistical analysis. They do not require any empirical evidence.

- **Planning**. An assessment workbook, presentation and research strategy are required for this category of UES measure. Similar to Small Saver, the assessment need only demonstrate sound engineering and statistical analysis.
- **Provisional**. An assessment workbook and presentation are also required for this category of UES measure. In addition, a **funded** research **plan** must be presented. Similar to Small Saver, the assessment need only demonstrate sound engineering and statistical analysis.
- **Proven**. This category has the highest quality standard. Data are needed to substantiate both baseline and efficient-case conditions for the measure. No research plan is required as the research must be complete before RTF approval can be achieved. The results of that research must be fully documented in the assessment workbook and presentation to the RTF.

- **Standard Protocol Measures**. The RTF approves estimation procedures for savings, costs, benefits, RUL, lifetime and cost-effectiveness for Standard Protocol measures. However, it is left to program operators to use these procedures in the assessment of all or sample of delivered measures. The main challenges in creating a Standard Protocol measure are developing a best practice savings estimation procedure and then reducing it to the simplest reliable form.
  - **Small Saver**. Similar to the UES measures, proponents must provide the RTF with a completed Measure Assessment Template that fully documents the measure specification and the estimation procedures for savings, costs, benefits, RUL, lifetime and cost-effectiveness. In addition, a document is prepared that follows the Standard Protocol Template, describing step-by-step what data is collected and how it is used in estimating savings. The protocol document is accompanied by a protocol calculator that can be used to carry out the estimation procedure for any delivered measure. Presentation materials must also be prepared and used in presenting the case for the measure's approval. The assessment workbook, protocol document, protocol calculator and presentation will be reviewed by an RTF subcommittee or by RTF staff prior to RTF consideration, to ensure the analysis and documentation are acceptable and the measure satisfies the quality standards for this category. The quality standards for this category only require sound engineering and statistical analysis. They do not require any empirical evidence.
  - **Planning**. This category is for measures whose savings potential is substantial, but the RTF lacks best practice data needed to test the reliability of simplified methods. As with all categories, a complete assessment workbook, protocol document, protocol

calculator and presentation are required. In addition, a research **strategy** must be presented.

- **Provisional**. This category is for measures whose savings potential is substantial, but the RTF lacks sufficient best practice data needed to test the reliability of simplified methods. As with all categories, a complete assessment workbook, protocol document, protocol calculator and presentation are required. In addition, a **funded** research **plan** must be presented.
- **Proven**. This category has the highest quality standard. Sufficient best practice data must be available to prove the reliability of the simplified method, i.e., the Provisional data collection requirements can be removed from the protocol document. No research plan is required as the research must be complete before RTF approval can be achieved. The results of that research must be fully documented in the assessment workbook, protocol document, protocol calculator and presentation to the RTF.

8.2. Updates to Existing Measures

A sunset date and the factors supporting that date are set when a measure of either method (UES or Standard Protocol) or any category (Small Saver, Planning, Provisional or Proven) is initially approved. In addition, the measure status is set to Active. As the sunset date approaches, the RTF can extend the date, change the status to Under Review or Deactivated the measure. It is also possible for a Proven measure to become Provisional, Planning or Small Saver due to changes in market conditions or product standards. At any time, any party may propose to the RTF that an existing measure be updated using the RTF Proposal Submission Form.

If a measure's status is changed to Under Review, the RTF staff or other parties will undertake a comprehensive review of the methods for estimating savings, costs, benefits, RUL and lifetime. As part of the review, new sources of data will be investigated, along with any relevant changes in codes, standards and other indicators of current market and product characteristics. The review may conclude that it is not practical to update the measure to maintain compliance with the Guidelines. In that case, a recommendation to deactivate the measure will be brought to the RTF.

If continued compliance with the Guidelines is practical, the review will identify the needed updates. These will be presented to the RTF for approval. The approved updates will be made and the measure brought back to the RTF, with a new sunset date.

GUIDELINES FOR THE ESTIMATION OF ENERGY SAVINGS

REGIONAL TECHNICAL FORUM

Release Date: December 8, 2015

TABLE OF CONTENTS

1. Introduction

The Regional Technical Forum (RTF) develops and maintains a series of documents that provide guidance on how to assess energy efficiency measures. This document (the Savings Guidelines) provides guidance on how to estimate energy savings.

1.1. Purpose

The purpose of these Savings Guidelines is to describe how the Regional Technical Forum (RTF) selects, develops and maintains methods for estimating savings from the delivery of energy efficiency measures. Four savings estimation methods are defined: Unit Energy Savings (UES), Standard Protocol, Custom Protocol and Program Impact Evaluation. It is the RTF's intention that each method will produce savings estimates of comparable reliability sufficient to meet the needs of regional energy planners. These methods are also expected to support regulatory processes related to the adoption and planning of energy efficiency programs.

The RTF plays two roles in savings estimation. For the UES and Standard Protocol methods, it approves detailed estimation procedures that can be directly applied to estimating savings for specific instances of delivered measures. For the custom protocol and program-impact evaluation methods, the RTF provides guidance that is more general and relies on program operators to implement this guidance in estimating savings.

1.2. Key Concepts

The following key concepts are referenced throughout this document. Their definitions are critical to the correct interpretation of the RTF's intent for each of the savings estimation methods.

1.2.1. Roadmap Concepts

A number of key concepts such as "Measure" and "Savings" are defined in the *Roadmap for the Assessment of Energy Efficiency Measures* (the Roadmap). These Savings Guidelines assume the reader has a thorough understanding of the Roadmap. Additional concepts that specifically apply to these Savings Guidelines are described below.

1.2.2. Delivery Verification Rate

The delivery verification rate is the ratio between the program impact evaluator's estimate of measure quantity delivered to the program operator's claim of delivered quantity. The units used to count the measure quantity will vary, for example, number of CFLs or area of insulation. Such rates can be computed for a single measure delivery, a representative sample of measure deliveries, an entire program or a portfolio of programs.

1.2.3. Savings Realization Rate

The savings realization rate is the ratio between the program impact evaluator's estimate of savings to the program operator's claim of savings. Such rates can be computed for a single measure delivery, a representative sample of measure deliveries, an entire program or a portfolio of programs.

1.3. Savings Estimation Methods

Four savings estimation methods are defined by these guidelines. For the first two methods, Unit Energy Savings and Standard Protocol, the RTF approves detailed estimation procedures. For the third and fourth methods, Custom Protocol and Program Impact Evaluation, the RTF provides more general guidance.

1.3.1. Unit Energy Savings (UES)

The UES method is appropriate for measures whose unitized savings, e.g., savings per lamp or motor, is stable (both the mean and variance) and can be reliably estimated throughout the period defined by the measure's sunset date. In addition, the data available on key estimation parameters and the estimation procedure used in establishing the unit energy savings meet the minimum quality requirements described in section 2.4.

The UES method reduces program delivery cost by simplifying the data that must be collected. Programs are only required to collect a verified count of delivered units, and the information needed to assign each delivered unit the correct UES value, e.g., savings for attic insulation in a single-family residence with an electric forced air furnace west of the Cascades. Total savings is the UES multiplied by the number of delivered units.

A measure unit is considered delivered if it is conforms to the measure's specifications.

1.3.2. Standard Protocol

A Standard Protocol method is appropriate when savings from a measure are widely varying but can be determined by a standardized procedure for data collection and analysis that is applicable to many different end use sites. Quality standards and other guidance related to standard protocols are found in section 3. Standardization of data collection reduces cost by eliminating or minimizing the need for site-specific data collection planning. Standardization of the analysis procedure also reduces the planning burden and ensures uniform quality in the analysis product. In addition, standardization reduces the skill level needed to reliably estimate savings and perform quality assurance activities.

1.3.3. Custom Protocol

Custom protocols are appropriate for measures that require site-specific savings estimation planning, data collection and analysis in order to develop a reliable estimate of savings.

Guidance concerning skill and documentation requirements for custom protocols is found in section 4. Custom protocols require site-specific documentation of the data collected and how those data are used in estimating savings. Highly skilled and experienced practitioners are required to estimate savings for Custom Protocol measures.

1.3.4. Program Impact Evaluation

Program impact evaluations estimate savings from a period of program operation. Program impact evaluations involve the analysis of a reliable sample of program participants (and possibly non-participants) to determine the savings. Guidance on how to design and conduct program impact evaluations is presented in section 5. The research designs for impact evaluations vary widely and yield program-level savings estimates covering the delivery of one or a group of measures during a period of program operation.

1.4. Treatment of Remaining Useful Life (RUL)

In some cases, the expected lifetime (see Lifetime Guidelines) of a measure with a pre-conditions baseline will exceed the remaining useful life (RUL) of the pre-condition system, equipment or practice. If RUL is ten years or less, savings, costs, and benefits (see Cost and Benefit Guidelines) should be estimated for the following time periods: a) RUL: time between measure delivery and when the baseline conditions would have changed if the measure had not been delivered, and b) Balance of Measure Lifetime (BML): time between RUL expiration and the end of the measure lifetime. The RTF expects careful consideration of savings, costs, and benefits during this period, but does not impose any specific quality standards on these estimates beyond the use of best available data and professional judgment.

The assumptions made and analysis performed to estimate savings for the first and second periods should be documented, respectively, in the Summary and SummaryRUL sheets of the Measure Assessment Workbook for UES measures. For Standard Protocol measures, these assumption and analysis should be documented in the appropriate sections of the protocol. For Custom Protocol measures, the assumptions made and analysis performed should be documented in the site-specific savings report.

If the estimated RUL is longer than ten years, then the RTF will assume that RUL equals measure lifetime.

1.5. IPMVP Adherence

These Guidelines do not address the issue of adherence to the International Performance Measurement and Verification Protocol (IPMVP). IPMVP adherence is not required for RTF deliberations. If need be, it is left up to agencies that operate programs to determine whether the planning, data collection, savings estimation and reporting that they perform adheres to one or more features of the IPMVP. For more information on the IPMVP see www.evo-world.org.

1.6. Development of RTF Savings Estimation Methods

There are many steps in the development process for savings estimation methods, starting with the specification of a measure and concluding with a reliable method for estimating savings for that measure. The process begins (see Figure 1) when regional research and development and program planning activities supply information on feasible measures. The RTF prioritizes its work on these measures and, for each, determines the most appropriate type of savings estimation method: Unit Energy Savings (UES), Standard Protocol, Custom Protocol or Program Impact Evaluation.

UES measures are approved in one of four categories (Proven, Provisional, Small Saver or Planning) depending upon the quality of data available to support estimates of savings. Maintenance and updates to UES measures occur according to each measure's sunset date. Research subject to RTF peer review supports this maintenance process. These research activities include data collection needed to advance measures from the Planning to Provisional or Proven categories, or from the Provisional to the Proven category.

Standard Protocol measures are placed in the Provisional category if the data available are not sufficient to prove the reliability of the savings estimation method. As with UES measures, the Provisional category and the associated data collection effort play an important part in advancing measures from the Provisional to the Proven category.

For Custom Protocol measures, the RTF develops and approves general guidelines, but relies on program operators to implement this guidance in estimating savings.

RTF Approved UES savings values, standard protocols and guidance on the treatment of Custom Protocol measures are all utilized by program impact evaluations. Program impact evaluations produce reliable savings estimates for all measures, including other UES measures not approved by the RTF.

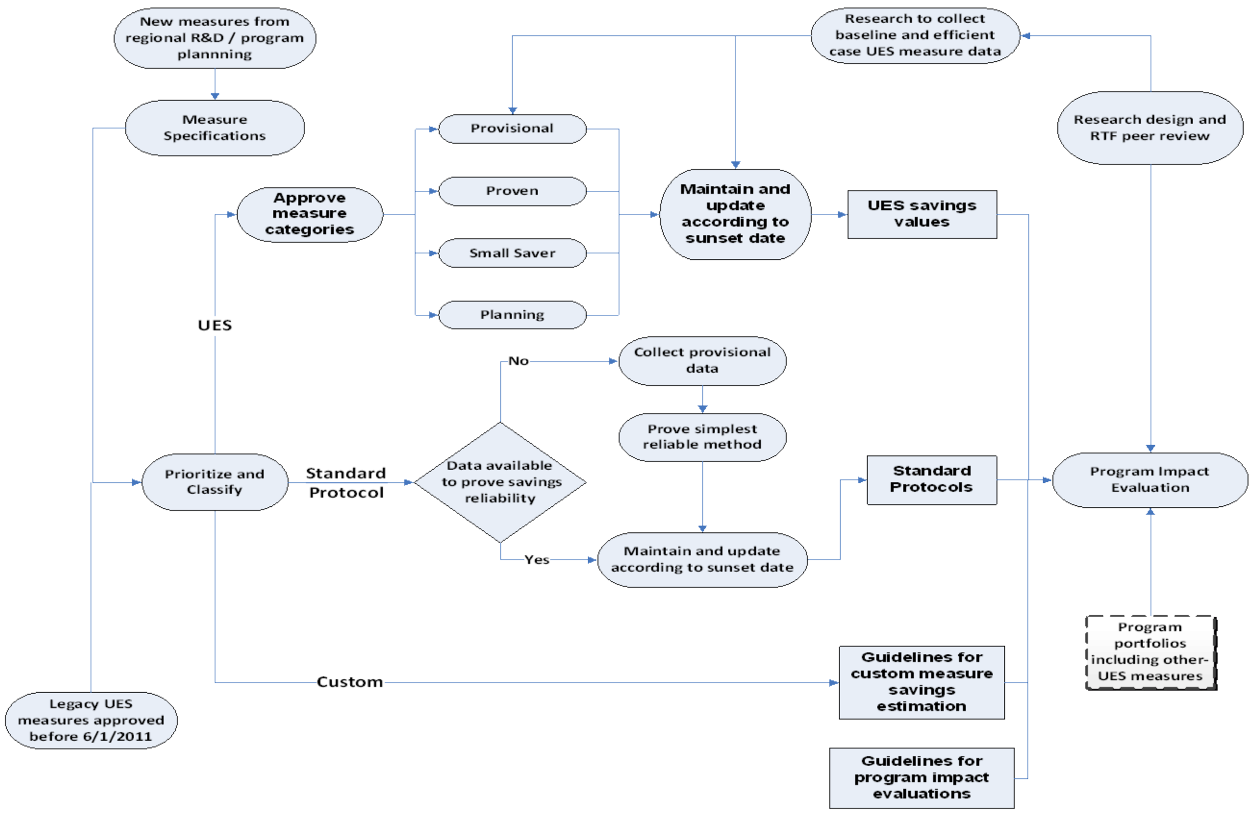

Figure 1: Development and Application of RTF Savings Estimation Methods

1.7. Supporting Documents

- **Measure Assessment Template**. The Excel® Measure Assessment Template contains the ProCost model, ProCost inputs and outputs, energy and costs and benefits analyses, and summary tables. It is used to fully document the data and methods used in assessing savings, costs, benefits, RUL, lifetime and cost-effectiveness for a measure. The template also contains a checklist that can be followed in conducting this assessment. A completed measure assessment workbook will either contain or reference the measure specification and will list or link to other files used in the assessment, such as SEEM workbooks.
- **Standard Information Workbook**. The RTF is responsible for maintaining the data provided in the Standard Information Workbook. This workbook includes information, which may be used in the assessment of measures, such as generic data values, data sources, and methods. For example, it contains commonly used input parameters for determining UES savings, including: weightings used for heating/cooling zones, house sizes, standard SEEM inputs, lighting hours of operation, and water-heater temperature set points.
- **Standard Protocol Template**. This Word® document provides a template, by example, for describing a Standard Protocol measure. A Standard Protocol must also have completed Summary and Checklist sheets (in the Measure Assessment Template) and a calculator (as described in these Guidelines).
- **Research Strategy Template**. This Word® document provides a template for documenting a research strategy. Such strategies are required for the approval of a Planning UES or Standard Protocol measure.
- **Research Plan Template**. This Word® document provides a template for documenting a research plan. Such plans are required for the approval of a Provisional UES or Standard Protocol measure.

2. Unit Energy Savings (UES)

This section describes the requirements for obtaining RTF approval of savings values for UES measures. The UES savings estimation method reduces program delivery costs by simplifying the data that must be collected. Programs are only required to collect a verified count of delivered units, and the information needed to assign each delivered unit to the correct UES value, e.g., savings for attic insulation in a single-family residence with an electric forced air furnace west of the Cascades. A measure unit is considered delivered if it conforms to the measures specifications. Total savings is the UES multiplied by the number of delivered units. Section 5 provides further guidance on the estimation of program-level savings from UES measures.

2.1. Eligible Measures

The UES method is appropriate for measures whose unitized savings, e.g., savings per lamp or motor, is stable (both the mean and variance) and can be reliably estimated throughout the period from RTF approval to the measure's sunset date. Each measure must adhere to its specifications. A measure may have multiple applications as defined by measure identifiers (see the Roadmap). Each application can be associated with a different baseline condition. Program operators must collect data needed to identify the specific measure application. This includes data needed for pre-conditions baselines to prove the relevant pre-condition was present at the time of measure delivery.

2.2. Required Knowledge and Skills

The analyst or the team supporting the analyst responsible for developing a UES measure for the RTF must have an understanding of the following:

- The measure's specification, including each of the specific applications of the measure
- Factors that determine the energy use of the measure-affected system(s) and equipment, e.g., the impact of outside air temperature on the performance of a residential heat pump
- The portions of these guidelines related to UES measures
- Analysis tools used to implement statistical or engineering models for estimating savings, including appropriate diagnostics to determine if a model is reliable
- Operation of affected systems and equipment, including the control settings and sequences that are relevant to obtaining the desired efficiency

The analyst or the team supporting the analyst must also be able to perform the following tasks:

- Develop and critique the savings estimation method
- Find and apply relevant primary data sources, including the definition of typical baseline and efficient-case conditions for each applicable measure baseline

- Calibrate models to base and efficient-case energy use
- Develop transparent documentation of the estimation methods and data sources
- Assist the RTF in determining the reliability of the savings estimate

2.3. Savings Estimation Procedures

The RTF must develop a best practice savings (Roadmap Key Concepts) estimation procedure for each UES measure. Documentation of the procedure (section 2.5) must include the methods used to derive all key input parameters and model(s) used to derive the UES values. The sources used in estimating input parameters or defining model(s) must be cited and accessible. Sources may be used in estimating savings if, after a diligent review, the RTF determines that they are sufficiently reliable. Three types of estimation procedures are allowed:

2.3.1. Statistical

The UES estimate may be derived from statistical analysis of baseline and efficient-case energy use for reliable samples of relevant end users or end uses. Such measurements of energy use can be used to estimate typical savings for a population that are representative of the likely future participants. Statistical designs can include comparison of randomly selected treatment and control groups or pre/post data collection for a treatment group. The quality of these estimates is judged primarily by the relative error of the mean savings estimate. It is also critical to determine whether there are systematic errors (biases) associated with sampling or data collection procedures that reduce or increase savings for all or a portion of the sample studied.

Statistical estimation procedures are not recommended, but may be feasible for UES measures whose savings significantly interact with other measures, e.g., lighting wattage reductions can change the savings from lighting controls or the installation of an electronic thermostat can change the savings from a ductless heat pump. Interaction is significant if the RTF determines that it could change a measure's savings estimate by more than ± 10%. Studies large enough to estimate savings for all combinations of interactive measures are likely to be cost prohibitive. However, statistical estimation may be appropriate for a highly interactive measure if the measure has a short lifetime, during which the interactions with other measures are likely to be insignificant.

2.3.2. Meta-Statistical

In some cases, a number of different agencies conduct relatively small statistical studies. None of these studies alone provides sufficient confidence in the UES estimate. However, the RTF may determine that a value in the range of savings demonstrated by these studies constitutes a sufficiently reliable estimate. Meta-statistical estimation is not recommended for measures whose savings significantly interact with other measures.

2.3.3. Calibrated Engineering

There are many instances where statistical or meta-statistical procedures are not practical or are not the best choice. In those instances, UES may be estimated with calibrated engineering procedures. These procedures rely on reliable data for key parameters of an engineering model. The model may be a simple equation, such as multiplying a change in efficiency rating by average capacity by average hours of operation. In other cases, more complex bin, regression, or simulation models may be used. Models may be calibrated to individual cases or to the average characteristics and consumption of groups. In some cases, the estimation may be carried out via a series of models. For example, a meta-statistical result for percent savings from single-family electronic thermostats is multiplied by typical heating consumption for such homes with the heating consumption derived from calibrated hourly simulation of typical homes.

Calibrated engineering estimates should be based on measurement and modeling of savings for end users representative of the measure's baseline.

For savings estimates derived from calibrated engineering procedures there are five factors that must be considered in judging the quality of the savings estimation procedure:

- Input parameters
- Model calibration
- UES components
- Interaction between measures
- Heating and cooling interactions

2.3.3.1. Input Parameters

The data supporting each baseline and efficient-case input parameter must be documented and determined to be reliable. A key consideration is whether the values are expected to be different for the RTF region based on factors such as existing building stock characteristics, demographics, climate, market characteristics and energy prices. In some cases, appropriate normalization of national or other-regional data may overcome the need for region-specific data. This may require that critical normalization parameters, e.g., efficiency rating of televisions, be gathered during program delivery. If normalization of national or other-regional data is not possible, or not sufficient, then region-specific data sources are required. The RTF may consider the relative importance of each parameter in determining what data collection is needed to establish reliable values. For some parameters, which are not primary determinants of measure savings, the RTF may rely on consensus opinion from a panel of experts in lieu of primary data collection.

2.3.3.2. Model Calibration

In most cases, calibrated engineering procedures will involve at least one stage of modeling in which baseline and efficient-case energy consumption are estimated for the measure-affected end use. For example, the heating load for single-family homes is estimated as part of the derivation of UES for ductless heat pump conversion. A simulation model is used to derive the heating end use for typical homes in different climate zones. Ideally, the model would be calibrated to measured heating end use for a sample of homes. If end use data are not available, the model should at least be calibrated to metered total use for the sample. Calibration should also be performed for samples that have adopted the measure, i.e., the efficient-case. For measures that have a current practice baseline, such as those delivered as part of new building construction, the calibration may be limited to the efficient-case or to comparable buildings of recent vintage.

2.3.3.3. UES Components

Often savings are estimated for separate UES components. For example, this is common for measures that reduce water and energy use. One portion of the estimation procedure will derive the UES component associated with energy savings at the site where the measure is delivered. Another UES component will be separately estimated based on the reduction in energy used to deliver water to the end use site. A reliable estimation procedure for each UES component must be separately described.

2.3.3.4. Interactions between Measures

In many cases, the savings of one measure depends on whether another measure is present. For example, the savings for residential weatherization measures depend on the efficiency of the heating system, in particular, whether a heat pump is used for space heating. The UES for each measure should account for all significant interactions with any of the other measures that are currently RTF approved. Interaction is significant if the RTF determines that it could change a measure's savings estimate by more than ± 10%.

To the extent practical, identifiers should be used to account for significant measure interactions. For example, the identifier 'heating system type' could be used to account for the interactions between various envelope improvement measures and a heat pump measure.

When measure identifiers do not account for all significant interactions, the savings can be estimated using the following steps:

1. Estimate baseline energy use of the measure-affected systems for typical adopters at the time of RTF approval.
2. Estimate efficient-case energy use assuming the delivery of the measure and full adoption of all cost-effective current RTF approved interactive measures. This is referred to as the full package. To avoid multiple iterations of the analysis to determine the full package, cost-effectiveness of the interactive measures may be estimated.

3. Compute full-package savings as the difference between the baseline energy use and efficient-case energy use of the full package.
4. Marginal savings for any measure is defined as the difference between the full-package energy use and the energy use of the full package without that measure. Compute marginal savings of the measure and each measure that interacts with it in the full package.
5. Compute the total marginal savings as the sum of the marginal savings for the measure and the marginal savings for each measure that interacts with it.
6. To estimate the measure's savings, compute the ratio of full-package savings to total marginal savings and multiply this ratio by the measure's marginal savings.

Other methods may be used if the RTF determines that they more reliably account for measure interactions.

The methods used to account for measure interaction must be documented in the Summary sheet of the Measure Assessment Workbook.

2.3.3.5. Heating/Cooling Interactions for a Measure

Some measures reduce the waste heat rejected to conditioned spaces or distribution systems. Reduction in waste heat results in increased primary heating consumption and decreased primary cooling consumption. The amount of the interaction varies based on heating/cooling system type, envelope and distribution system characteristics, climatic conditions, and other variables. The savings estimation procedure must include appropriate adjustments to UES savings to account for significant heating/cooling interactions or provide an explanation as to why the interactions are not significant. Interaction is significant if the RTF determines that it could change a measure's savings estimate by more than ± 10%.

2.3.3.6. Other Interactions

Interactions other than those described above are possible. For example, a more efficient clothes washer may also reduce the energy needed for clothes drying. The savings estimation procedure must include appropriate adjustments to UES savings to account for other significant interactions or provide an explanation as to why the interactions are not significant. Interaction is significant if the RTF determines that it could change a measure's savings estimate by more than ± 10%.

2.4. Quality Standards

The following standards will be applied by the RTF to determine the quality and reliability of UES values. Separate quality standards are defined in this section for each of the four categories of UES measures. In addition, this section provides guidance on how the RTF determines the reliability of the savings from UES measures.

All of these quality standards are subject to RTF judgment concerning the limitation of best practice savings estimation (Roadmap Key Concepts). The RTF shall determine which inputs are needed for savings estimation, taking into account data availability, practicality, and impact. For example, it may not be practical to measure heating and cooling interactions with internal loads such as lighting. The RTF may determine certain inputs to savings estimation via professional judgment when it determines that there are limits to best practice savings estimation.

2.4.1. Proven

Proven savings estimation methods are those that meet the highest-quality standard. Reliable estimates of savings from Proven UES measures only require delivery verification as defined in section 5.4.6. A sunset date must be set for Proven measures at the time they are approved. Proven UES values may be approved if the following criteria are met.

For statistical or meta-statistical estimation procedures:

- The statistical findings are determined to be sufficiently reliable.

For calibrated engineering estimation procedures:

- Reliable data are available for estimating the baseline[1] and efficient-case[2] measure-affected end use energy consumption.
- Sound engineering and statistical analyses are performed to develop the UES estimate.
- Any models used in the estimation procedure have been calibrated at a minimum[3] to baseline and efficient-case energy consumption.

2.4.2. Provisional

Provisional savings estimation methods are those that the RTF approves with special conditions requiring the collection of data from all or a sample of measure deliveries. For measure in this category, it must be possible to obtain the data and analyses needed to meet the Proven

[1] As defined in the Roadmap, current practices (e.g., product specifications, construction practices and operations) define the baseline for some measures. As these conditions cannot be directly observed or measured, reliable baseline consumption must be estimated, for example by the use of engineering modeling or measurement of control groups.

[2] The RTF may determine that efficient-case consumption can be estimated from measurements of baseline conditions and reliable data on critical efficient-case performance parameters, e.g., using baseline hours of residential lighting operation in combination with Energy Star test results for LED bulb efficacy.

[3] Calibration to the affected end use consumption is preferable.

quality standard. The data and analyses may come from programs delivering the measure or from other studies. The plan for completing the necessary data collection and analyses must be approved along with the Provisional UES values. In addition, a sunset date must be adopted that is consistent with the plan. For measures that require long periods to collect and analyze baseline and efficient-case data, the plan should include staged analyses so that early experience with deployment, baseline conditions, and measure performance can be used to adjust the sunset date.

Provisional UES values may be approved if the following criteria are met.

For statistical or meta-statistical estimation procedures:

- The statistical findings are determined to be sufficiently reliable, even if the level of reliability is less than that required for Proven UES values.
- A research plan has been developed and funded that is likely to result in sufficient data on baseline and efficient-case energy consumption or other required variables to support an RTF decision on whether to approve Proven UES values.

For calibrated engineering estimation procedures:

- Sound engineering and statistical analyses are performed to develop the UES estimate.
- A research plan has been developed and funded that is likely to result in sufficient data on baseline and efficient-case energy consumption or other required variables to support an RTF decision on whether to approve Proven UES values.

2.4.3. Planning

The RTF may determine that UES values are needed, even though they do not meet the quality standards of the Provisional or Proven categories. These may be needed by program operators for planning purposes, such as the design and operation of pilot programs or regional coordination. Planning UES values may be approved if the following criteria are met.

For any type of estimation procedure:

- A research strategy (see Roadmap) has been developed that is likely to result in sufficient data on baseline or efficient-case energy consumption or other required variables to support an RTF decision on whether to approve Proven UES values.
- The measure has sufficient usefulness and applicability in the region.
- Sound engineering and statistical analyses are performed to develop the UES estimate.

2.4.4. Small Saver

The RTF may determine that the likely savings from a measure are too small to warrant the resources needed to meet the quality standards for Provisional or Proven category. In making this determination, the RTF will consider the size of the regional technical potential savings. Measures with small savings may be RTF Approved in the Provisional or Proven categories if data needed to meet the quality standards of those categories are readily available. Measure specifications (see the Roadmap) are required for small savers and must be provided before the RTF can approve a measure as a Small Saver.

Small Saver UES values may be approved if the following criteria are met.

For any type of estimation procedure:

- The measure is applicable in the region.
- Sound engineering and statistical analyses are performed to develop the UES estimate.

2.4.5. Savings Reliability Standard

The RTF, after reviewing full documentation of the estimation procedure and the supporting data, will determine whether any UES measure savings estimate is sufficiently reliable. This determination will be based on many factors, including sample size, sample design, validity of the model specification, model calibration and measurement errors. Measure cost-effectiveness will not be a consideration in determining whether the UES estimate is sufficiently reliable. For example, if a measure had a benefit to cost ratio of ten there might be a tendency to allow for less reliable primary data by applying a large discount factor to the UES. By resisting this tendency, the RTF enhances the reliability of the savings estimate.

2.5. Documentation Standard

Throughout the life cycle of a UES measure, the primary vehicle for the estimation of measure savings and documenting the data sources and estimation procedures is an Excel® workbook. A primary workbook is created at the time a measure is proposed based on the Measure Assessment Template.

The Summary sheet in the workbook (see the Measure Assessment Template for an example) contains four sections that pertain to savings estimation:

- **Measure Classification and Properties**. This section includes: market sector, market segment, measure category (Proven, Provisional, Planning, and Small Saver), measure description (including references to important specifications and eligibility requirements), sunset date, sunset factors, primary workbook, linked workbooks, and a description of the number of measures and UES components for which UES values are estimated.
- **Measure Identifiers**. This section documents the identifiers of the specific measure applications for which UES values are estimated. The table lists the possible values for each identifier, e.g., heating zones 1, 2 and 3, and provides further explanation and relevant sources that support each identifier. Collectively, these identifiers define all the possible measure applications. Measure Type is the only required identifier. For some measures, there is only one measure type, e.g., high efficiency televisions. For others there can be many measure types, such as each type of residential single-family weatherization treatment, e.g., wall and ceiling insulation.

- **Constant Parameters**. This section lists all of the key input parameters whose values do not vary between baseline and efficient-case for a measure. If these constants have different values across the specific measure applications, the constants for each of those applications are described. The table also contains further explanation and sources for each of these constants.
- **Energy Savings Estimation Method, Parameters and Data Sources**. This section documents the analysis used to derive the UES for each of the UES components for each measure type. Primary parameters or adjustment factors are listed for each UES component, along with their baseline and efficient-case values. The sources for each of these values are also listed.

When a measure is initially proposed, the Summary sheet can be used to document the expected estimation procedure and data sources. The Summary sheet along with preliminary analysis of savings would be the basis for an RTF decision on whether to consider the measure.

2.6. Method Development and Maintenance

Figure 2 illustrates the development and maintenance process for UES measures. Each measure is approved in one of four categories, based on the measure specification and documentation provided in the measure workbook. The sunset date triggers review of the measure. All available primary data is considered in this review, including those provided by the data collection plans associated with Planning and Provisional measures. The reviews provide recommendations concerning the appropriate status of the measure: Active, Under Review or Deactivated. In addition, a special status, Out-of-Compliance, is defined for legacy measures (those approved by the RTF prior to June 1, 2011).

The RTF assigns the sunset date whenever the status of a UES measure changes. The RTF records the factors that justify each sunset date such as expected revisions to energy codes, federal standards, completion of RTF approved research plans, or shifts in current practices of consumers. The sunset date determines when the UES savings values may be used. Beyond that date, the UES values are no longer RTF approved.

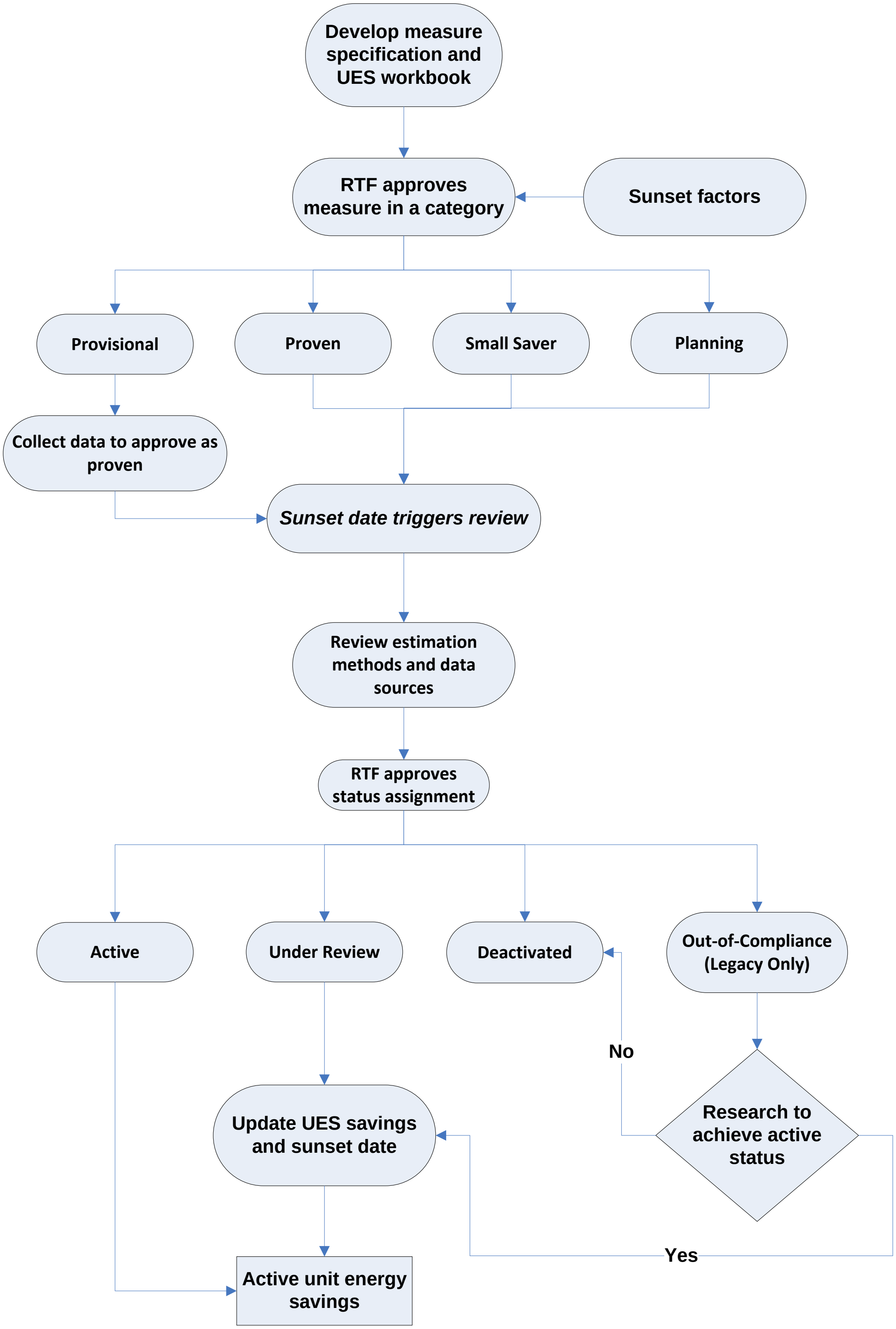

Figure 2: UES Measure Development and Maintenance

2.6.1. Measure Status

A UES measure may be assigned to any of the following statuses. The RTF assigns a status to a measure when it is initially approved. It may change the status of a measure based on the sunset date or at other times. The status indicates how the measure's savings values are used and whether the RTF is engaged in efforts to review and possibly revise these savings estimates.

2.6.1.1. Active

Active status means that the measure's specification is current and that the measure workbook contains and documents reliable saving values for each of the measure applications. When the RTF initially approves a measure its status is Active.

2.6.1.2. Under Review

At any time prior to the sunset date, the RTF may decide to place the measure under review. This may be the result of a review of the UES savings estimation procedure, the availability of new sources of data for baseline or efficient-case consumption or a change in the measure's specification. The UES values will remain RTF approved while the measure is under review. As a result of the review, the UES values may be re-estimated. Under-review status is generally not appropriate for Proven measures if the RTF expects the review will change measure savings by more than ± 10%. Instead, the RTF may choose to modify the measure category to Provisional or Planning, or to deactivate the measure.

2.6.1.3. Deactivated

If the sunset date has passed and the new or revised UES estimates have not been approved, the RTF will decide whether to deactivate the measure. This means that the UES estimates are no longer approved by the RTF. The RTF may decide to deactivate a measure before the sunset date based on unanticipated factors, such as the adoption of new energy codes or the release of study results with findings that invalidate the UES values or the procedures for estimating those values.

2.6.1.4. Out-Of-Compliance

For UES measures approved prior to June 1, 2011 the RTF may determine that the UES values do not comply with one or more requirements of these guidelines. The UES estimates for these measures will continue to be RTF approved if a plan for bringing the measure into compliance is approved by the RTF within one year following the RTF determination that the measure is out-of-compliance. If no plan is approved within one year, the RTF will decide whether the measure should be deactivated. The RTF intends that all Out-of-Compliance measures will be deactivated, reclassified as a Standard Protocol measure, or meet the quality standards for Small Saver, Provisional or Proven UES measures, as soon as possible, but no later than June 1, 2016.

2.6.2. Research on Measures

A Research strategy is developed by the RTF for each Planning measure. Following discussions amongst sponsors a research plan is developed and funded for Provisional measures. Templates for the research strategy and plan are provided (see Supporting Documents) and address the following topics:

- Sample design(sampling frame, domains of study, stratification, quotas, selection method, replacement procedure, and expected sampling precision) .
- Data collection requirements and methods.
- Data analysis and savings estimation methods.

Comparable plans may also be needed to support the update of Proven measures, such as data needed to update baseline characteristics to correctly model measure interactions.

3. Standard Protocols

Standard protocols support estimation of savings for a measure at specific end user sites. Standard protocols provide guidance on what data are required to establish the key determinants of savings and the simplified methods to be used to reliably estimate those savings. The extent of data collection and analysis required by the protocol is the minimum level needed for reliable savings estimation. Standardization of data collection reduces cost by eliminating or minimizing the need for site-specific data collection planning. Standardization of the analysis procedure also reduces the planning burden and ensures uniform quality in the analysis product. Standardization reduces the skill level needed to reliably estimate savings.

Standard protocols may be used on a census or a sample of sites where the measure is delivered. Application of standard protocols to a sample of sites may be warranted under two conditions:

- **Collecting Data Needed to Develop the Proven Protocol**. If insufficient best practice data are available from national or regional sources, the development of a Standard Protocol will require new data collection. It may be necessary for program operators to collect data from a sample of delivered measures to provide the information needed for a Proven protocol.
- **Using the Protocol for Program Impact Evaluation**. Standard protocol savings estimates are only required for a sample of delivered measures that is sufficient to reliably estimate total program savings. The treatment of these Standard Protocol samples for program impact evaluation is further discussed in section 5.

3.1. Eligibile Measures

A Standard Protocol describes reliable data collection and savings estimation procedures for a measure. The protocol must contain all measure specification information required in by the Roadmap, including clear description of the measure applications for which the protocol can be used to reliably estimate savings. Lists of either eligible or in-eligible measure applications may define eligibility. For each application, the protocol must specify the appropriate baseline to be used in estimating savings: current practice or pre-conditions.

3.2. Required Knowledge and Skills

The analyst or the team supporting the analyst responsible for developing a Standard Protocol measure for the RTF must have an understanding of the following:

- Measure specification, including each of the specific applications of the measure.
- Factors that determine the energy use of the measure-affected system(s) and equipment, e.g., the impact of outside air temperature on the performance of unitary HVAC equipment.
- The section of these guidelines related to Standard Protocol measures.

- Operation of affected systems and equipment, including the control settings and sequences that are relevant to obtaining the desired efficiency.

The analyst or the team supporting the analyst responsible for developing a Standard Protocol measure for the RTF must also be able to perform the following tasks:

- Describe a best practice savings estimation procedure.
- Identify and prove the simplest reliable savings estimation procedure.
- If needed, identify Provisional data collection needed to prove the reliability of the savings estimation procedure.
- Implement both the best practice and simplest reliable savings estimation procedure.
- Develop transparent documentation of the estimation methods and data collection procedures.
- Assist the RTF in determining the reliability of the savings estimation procedure.

3.3. Savings Estimation Procedures

Standard protocols define how data is collected and analyzed to estimate savings from the delivery of a measure to an end user site. The data to be collected and the analyses to be performed are determined by the characteristics of the measure and the end use application. The Standard Protocol consists of both a document that describes how to collect and analyze the data and a calculator that performs the required analyses with the specified data. Input parameter values for models used to estimate savings may be derived from sources other than data collected at an end user site. These sources must be cited and accessible. Such sources may be used in estimating savings, if after a diligent review; the RTF determines that they are sufficiently reliable.

All standard protocols approved by the RTF must meet the following requirements:

- Precise specification of the measure including eligibility rules.
- Simplest reliable savings estimation procedure.
- Entirely prescriptive data collection and analysis procedure that will work for all eligible measures.
- Independent of program design and delivery method.
- Skills required are common among the region's program implementation workforce.
- Savings estimates account for interactions with other measures previously implemented at the same site.
- Documentation that complies with section 3.5.

For measures with pre-conditions baselines and RUL of ten years or less, the protocol must describe the assumptions and analyses used to estimate savings between the time when the RUL expires and the measure lifetime expires.

3.4. Quality Standards

The following standards will be applied by the RTF to determine the quality and reliability of standard protocols. There are four categories of standard protocols, each with their own quality standards.

3.4.1. Planning

The development of the protocol for a measure begins by assembling site-specific data that support a best practices savings estimate. In some cases, the number and quality of best practice examples are not adequate to support the definition of the simplest reliable savings estimation method. In those cases, the protocol can be assigned to the Planning category. The best practice data collection requirements should include the following:

- True power measurements or reliable methods for estimating true power of affected systems or equipment.
- Measurements of the primary determinants of consumption for the affected systems and equipment, such as flow rates, outside air temperature, building size, baseline equipment efficiency, hours of occupancy, production levels, and temperature settings.
- Measurements of less expensive and less technically challenging surrogates for the primary determinants, such as damper position instead of flow, which will support the development of the simplest reliable savings estimation method.
- If the measure has a pre-conditions baseline, trend logs for baseline period that are of sufficient duration to represent most of the variance in energy use and its determinants. For measure with current practice baseline, sufficient data to model energy consumption for affected systems and equipment under typical current practice conditions.
- Trend logs for post (after measure delivery) periods that are of sufficient duration to represent most of the variance in energy use and its determinants.

A research strategy (see Roadmap) must be developed that describes how the best practice data are used to estimate savings for each site. This estimate should take full advantage of the measurements of true power and the direct determinants of consumption. This best practice savings estimate provides the benchmark against which the reliability of estimates derived with simpler methods can be tested.

The research strategy must also describe one or more simplified savings estimation procedures that will tested against the best practice method. The Standard Protocol should be based on the simplest and least expensive method that provides reliable savings estimates. Following are examples of the types of simplifications that should be considered:

- One-time measurements of principal determinants of usage such as distribution leakage or equipment run time.
- Combining trend logs of current with one-time measurements of true power to estimate a trend log for true power.

- Estimating the primary determinants for consumption, such as flow, by measuring a simple surrogate and applying default performance curves.
- Combining one-time measurements of baseline conditions or default performance curves with post-period trend logs of system/equipment utilization to estimate baseline consumption and performance.
- Reducing the duration of either baseline or post-trend log periods.

In some cases, the best practice method may be exactly the same as the simplest reliable method. In other cases, portions of the best practice method may be the same as the simplest reliable method, such as in the case of non-residential lighting, where fixture wattages from manufacturer specifications are accepted for use in both best practice and simplest reliable methods. No comparison is needed for any aspect of a Standard Protocol that is the same in both best practice and simplest reliable methods. In addition, the best practice method may rely on parameter values obtained from studies performed by other agencies, if after a diligent review of these studies the RTF determines that the values are sufficiently reliable.

Prior to RTF approval each Planning Standard Protocol must have a research strategy that describes:

1. The best practice method for savings estimation and the data that must be collected to implement that method for each measure delivery.
2. One or more simplified methods that will be tested and compared to the best practice method
3. The sample of best practice sites that will be required to prove the reliability of the simplified method.

3.4.2. Provisional

Provisional standard protocols require best practice data collection or data collection on key parameters in addition to the data needed to support one or more simplified methods. Prior to RTF approval each Provisional Standard Protcol must have a funded research plan (see Roadmap). The research plan plan describes in detail the data collection and analysis that will be performed to support the RTF decision on whether the simplified method is sufficiently reliably and can be approved as Proven.

3.4.3. Proven

Proven standard protocols are those whose simplified method is proven reliable, have no Provisional data collection requirements, and have best practice data to support the protocol. One of the requirements for a Standard Protocol is that it be the simplest reliable method for estimating savings for a measure. A precise definition of reliability is difficult to enforce across all standard protocols, but in general, any method that produces savings estimates within ±20 percent of the best practice method (across a representative sample of best practice examples) should be considered sufficiently reliable. Alternatively, a method is sufficiently reliable if the combined sampling and measurement error for the representative sample is less than ±10 % with a confidence interval of 90%. In addition to considering the magnitude of the error, it is necessary to consider whether the simplified procedure is biased. A method that is always 20 percent high or 20 percent low for all tested cases would not be considered reliable.

3.4.4. Small Saver

The RTF may determine that the likely savings from a measure are too small to warrant the resources needed to meet the quality standards for a Proven Standard Protocol. In making this determination, the RTF will consider the size of the regional end use that is affected by the measure or the magnitude of the likely savings. A Standard Protocol in the Small Saver category is reliable if it satisfies the following criteria:

- The measure is applicable in the region.
- Sound engineering and statistical analyses are used in estimating savings

3.5. Documentation Standard

A Standard Protocol consists of three parts. The first part is a measure assessment workbook, based on the RTF Measure Assessment Template (see RTF website). The assessment workbook contains the measure development and maintenance checklist and summary sheets. The summary sheet includes description of the energy savings estimation methods, parameters, data sources, and other supporting information. The second part is a Word® document that describes the protocol. A template for this word document is available on the RTF website. The third part is a transparent and accessible calculator that can be used to compute savings for the measure in a fashion that is consistent with the protocol.

3.5.1. Description of Data Collection and Analysis

The template for the protocol document requires the following information.

- **Purpose**. Defines the measure and key features and objectives of the protocol.
- **Sunset Date and Sunset Factors**. The protocol cannot be used beyond the sunset date. In addition, factors that support this date must be described, such as expected revisions to energy codes, federal standards, completion of RTF approved research plans, or shifts in current practices of consumers.
- **Definition of Key Terms**. Definition of terms that aid in making the language of the protocol concise. Terms are not included if they would be commonly known to the practitioners who have sufficient skills to conduct data collection and analysis in accordance with the protocol.
- **Eligible Measure Applications**. Specifies the types of measure applications that can use this protocol to create an RTF approved estimate of savings. Listing types of applications that are not eligible can be just as important as listing those that are eligible. Must specify which

types are associated with pre-conditions baseline and which are associated with current practice baseline.

- **Required Knowledge and Skills of Practitioner**. Describes the required knowledge and skills for practitioners that will use the protocol to estimate savings.
- **Required Delivery Verification**. Description of the inspection procedure, testing, and documentation review which must be completed by the practitioner to determine whether the measure is fully operational during the post-period.
- **Data Collection Requirements**. Exact description of the data that must be collected during the baseline and post periods. Alternative measurements may be specified that exceed the minimum requirements as in some cases these alternative measurements may be easier to obtain.
- **Savings Estimation Steps**. Describes the computational algorithm that can be found in the accompanying calculator for estimating savings. If any of the eligible measure types have current practice baseline, the protocol must describe how baseline energy use is estimated for current practice conditions.
- **Within-Measure Sampling Procedure**. When applicable, this section describes the allowed method for sampling units, e.g., sampling fixtures at a building where a lighting measure has been delivered. Sampling is applicable when it is common for the measure to be implemented in large quantities of units at a single site.
- *Optional* **- Relationship to Other Protocols and Guidelines**. Discusses relevant relationships to other protocols, such as *International Performance Measurement and Verification Protocol* (Efficiency Valuation Organization - http://www.evo-world.org) and *M&V Guidelines: Measurement and Verification for Federal Energy Projects* (US Department of Energy - http://www1.eere.energy.gov/femp/pdfs/mv_guidelines.pdf), and guidelines such as those maintained by BPA and other regional utilities and agencies.
- **Estimate of Typical Cost**. Provides a listing of the major tasks that must be performed to implement the protocol for a typical measure and estimates the hours and costs (labor and non-labor) associated with this work.
- **Provisional Data Collection**. Describes special data collection that is required during the period of that measure is assigned to the Provisional category. Once the RTF approves assignment of the protocol to the Proven category, these data collection requirements are no longer in force.
- **User's Guide to Savings Calculator**. A step-by-step guide to using the accompanying calculator.

3.5.2. Transparent and Accessible Savings Calculator

Each protocol must provide a transparent and accessible savings calculator, to ensure the correct interpretation of the savings estimation steps. The calculator is a precise rendering of the estimation steps but is not necessarily to be used by program operators. Program operators may develop equivalent calculators for the purpose of estimating Standard Protocol measure savings. The calculator must be capable of estimating savings using data for both the best practice and simplest reliable estimation methods. The calculator may consist of two separate parts, one for the best practice method and one for the simplest reliable method. In some cases, these methods may be identical.

The calculator must accept exactly the data that are required by the protocol. The calculator must be implemented using computer software that is generally accessible and reasonably priced for all practitioners in the region. The software must either be inherently transparent, such as an Excel® workbook that does not rely on extensive macro coding, or it must be fully documented. Fully documented means the exact algorithms for all calculations are completely described in a document accessible to all practitioners or that the analysis method is documented along with the results of a validation process, similar to ASHRAE Standard 140, which demonstrates the comparability of the method to other accepted calculation methods.

3.6. Method Development and Maintenance

The development process begins with the definition of a best practice standard for data collection and analysis. Best practice data are assembled or if necessary collected using the Provisional Standard Protocol for a sample of delivered measures. These best practice data are used to estimate savings for a sample of delivered measures. In addition, the simplest reliable method is used to estimate savings for each element of the sample. The differences are examined and the quality standards in section 3.4.2 are used to determine whether the simplest reliable method has been proven reliable.

Once a simplified method is proven reliable, the Standard Protocol documentation and associated calculator will be modified to be consistent with that method. In addition, any Provisional data collection requirements will be removed.

The RTF will assign a status to a Standard Protocol: Active, Under Review or Deactivated. The measure sunset date is reviewed and possibly revised whenever the measure status is changed.

3.6.1. Measure Status

A Standard Protocol may be assigned to any of the following statuses. The RTF assigns a status to a Standard Protocol when it is initially approved. It may change the status of a Standard Protocol based on the sunset date or at other times. The status indicates whether the Standard Protocol can be used and whether the RTF is engaged in efforts to review and possibly revise the Standard Protocol.

3.6.1.1. Active

Active standard protocols are ready for use by program operators. Documentation describing the protocol is complete and the associated calculator is been shown to correctly estimate savings.

3.6.1.2. Under Review

At any time prior to the sunset date, the RTF may decide to place a Standard Protocol under review. A Standard Protocol may be placed under review for a number of reasons including: concerns about the reliability of the data collection or savings estimation procedures, proposals to change the definition of the measure, or the availability of new sources of best practice data. The Standard Protocol will remain “RTF approved” while it is under review. As a result of the review, the RTF may approve changes to the Standard Protocol and adopt a new sunset date.

3.6.1.3. Deactivated

If the sunset date is passed and the new or revised Standard Protocol has not been approved, the RTF will deactivate the Standard Protocol. This means that the Standard Protocol is no longer approved by the RTF. The RTF may decide to deactivate a Standard Protocol before the sunset date based on unanticipated factors, such as the adoption of new energy codes or the release of study results with findings that invalidate the savings estimation procedure.

3.6.2. Research on Measures

A Research strategy is developed by the RTF for each Planning measure. Following discussions amongst sponsors a research plan is developed and funded for Provisional measures. Templates for the research strategy and plan are provided (see Supporting Documents) and address the following topics:

- Sample design (sampling frame, domains of study, stratification, quotas, selection method, replacement procedure, and expected sampling precision).
- Data collection requirements and methods.
- Data analysis and savings estimation methods.
- Comparable plans may also be needed to support the update of Proven measures.

4. Custom Protocols

This section provides guidance to program operators on how to estimate savings reliably for Custom Protocol measures. Custom Protocol measures are those that require site-specific data collection and analysis to support reliable estimates of savings and for which there is no RTF approved Standard Protocol (see section 3). The guidance provided by this section to program-operators deals only with the treatment of a measure delivered to a single location. Section 5 describes how to apply this guidance in the context of program impact evaluations.

4.1. Eligible Measures

Custom protocols are appropriate for any measure that requires site-specific data collection and analysis in order to develop a reliable estimate of savings.

4.2. Required Knowledge and Skills

The practitioner is the person with lead responsibility for estimating measure savings. This person, or the team supporting this person, must have an understanding of the following:

- Factors that determine the energy use of the measure-affected system(s) and equipment, e.g., the impact of outside air temperature on the performance of a chiller.
- Safety procedures relevant to the end user facility, affected system(s) and the required measurement equipment.
- Section of these guidelines related to Custom Protocol measures and the applicable guidelines enforced by the program operator delivering the measure.
- Sampling techniques as they pertain to selecting a sample of affected equipment that fall within the bounds of the measure.
- Analysis tools used to implement statistical or engineering models for estimating savings, including appropriate diagnostics to determine if the model is reliable.
- Operation of affected systems and equipment, including the control settings and sequences that are relevant to obtaining the desired efficiency.

The practitioner or the team supporting the practitioner must also be able to perform the following tasks:

- Develop a cost-efficient plan for reliably estimating savings from the measure.
- Conduct all required inspections of the affected system(s) to verify they are operating correctly and extract necessary data from related documentation and end user records.
- Supervise any staff taking required measurements.
- Install and operate data collection equipment and obtain necessary trend logs from facility control systems.

4.3. Savings Estimation Procedures

4.3.1. Measure Description

The practitioner responsible for the estimation of savings must understand and be prepared to document the following features of the measure.

4.3.1.1. How the Measure Saves Energy

The physical changes that comprise a measure may cause a reduction in energy use for many different reasons. It is important that the practitioner understand and describe how the measure saves energy. As a general example, consider this description for how installation of a VFD on a fan system would save energy.

Many fans do not need to run at full capacity all of the time. VFDs are more efficient than other throttling methods, such as dampers, at regulating fan flow rates.

For a more detailed example, consider this description of how installation of an electric tankless, on-demand water heater would save energy:

An existing large capacity (400 gallon) electric water heater is dedicated to supplying 160 degrees F hot water to several linen washers in a hospitals laundry department. The linen washers complete an average of two hundred loads per week, using seven gallons of hot water per load. The electric tankless, on-demand water heater will replace the large gallon electric water heater, saving energy by only heating water as needed by the linen washers.

4.3.1.2. Affected Systems or Equipment

The practitioner should identify all the systems and equipment affected by the implementation of the measure. For example, changes to air-handling units may cause changes in a chiller's operation. It is also possible for measure-related changes in one system to affect the energy use of other systems. For example, reduction in the internal loads within conditioned spaces will cause changes in the operation and energy use of the HVAC systems serving those spaces. The impact of a measure on the operation of other equipment at the facility is often referred to as an "interactive effect." The savings estimation model for the measure should quantify the changes (positive or negative) to all the affected systems and equipment if the effects are significant. An effect is significant if, in the judgment of the practitioner, it could change a measure's savings estimate by more than ± 10%.

4.3.1.3. Determinants of Savings

The practitioner must understand the significant determinants of the savings. Determinants may include, but are not limited to, the following:

- Other efficiency measures that may affect the savings from the delivered measure
- Hours of operation

- Equipment efficiency at full and part load operation
- Control sequence and settings
- Outside air temperature or other weather parameters
- Production rate and schedule
- Building occupancy
- Time-of-day.

The savings estimation model for the measure should account for the effects of significant savings determinants. A determinant is significant if, in the judgment of the practitioner, its absence from the model could change a measure's savings estimate by more than ± 10%.

4.3.1.4. Baseline Conditions

The practitioner needs to determine which of the baselines defined in the Roadmap, apply to the measure. The practitioner must determine the correct baseline before selecting an appropriate savings estimation model and before determining what data need to be collected.

4.3.1.4.1. Current Practice

The practitioner should estimate measure savings using a current practice baseline if the affected systems, equipment or practice is at the end of its useful life. There are a number of possible indicators that current practice is the appropriate baseline, as defined in the Roadmap. If any of these conditions prevail, the practitioner should determine current practice for the affected equipment. This can be challenging. The practitioner needs to identify what the end user would typically do without the incentives and services offered by the program operator. The practitioner should seek data on current practices from applicable codes and standards, or one of the following if they constitute a more energy efficient baseline for the measure and the information is practical to obtain and applicable to the delivered measure's location.

- Assumed baseline in the most recent Council Power Plan
- Results of market research provided by the program operator (such research should be done for frequently occurring measures in accordance with guidelines provided in section 5)
- Recent similar purchases by the end user
- Documented end user plans or specifications
- End user or vendor developed alternative designs, considered as part of the measure selection process
- End user description of what was done in similar circumstances elsewhere in the facility or in another facility they operate
- Equipment vendor's description of what they would normally do for this end user

If none of the above is practical or applicable, the baseline should be based on the practitioner's opinion about what would normally be done, based on prior experience with similar projects.

Categories of measures that would typically require current practice baseline include but are not limited to the following:

- Energy efficiency features associated with entirely new buildings, systems or equipment
- Energy efficiency features implemented as part of major renovation of existing buildings, systems or equipment
- Efficient lighting[4]

4.3.1.4.2. Pre-Conditions

The practitioner should estimate measure savings using a pre-conditions baseline if the affected system, equipment, or practice is not at the end of its useful life.

During the pre-conditions period the practitioner should use the data collection methods described in section 4.3.5 to determine the following, as needed, for the affected systems and equipment:

- Loads, e.g., air or water flow, Btu/h, cooling tons, and conveyance delivery rates
- Equipment performance , e.g., sizing, performance curves, part-load efficiency curves
- Control sequence and set points
- Envelope thermal properties, e.g., glazing U-values, ventilation rate, and insulation levels
- Distribution system properties, e.g., leakage, pressure drop and insulation levels
- Operations and maintenance practices
- Weather conditions are often a significant determinant. The selected savings estimation model may require actual weather for periods before and after measure delivery. However, final savings estimates should use typical weather conditions. For typical weather, the practitioner should use the most recent TMY (Typical Meteorological Year) weather file from the station that best represents weather at the end user site.
- Other significant determinants of savings

Categories of measures that would typically require a pre-conditions baseline include but are not limited to the following:

- Improvements to operations and maintenance practices
- Retro-commissioning

[4] Certain lighting efficiency projects may have such compelling benefits that they replace lighting equipment that has substantial remaining useful life and the project does not trigger compliance with local energy codes. Such projects would have a pre-conditions baseline.

- Automated control upgrades for systems under manual control or using mechanical controls

Pre-conditions baseline may be appropriate even if the end user's requirements are changing. For example, a measure can improve the efficiency of the affected systems and allow the existing equipment to serve expanded production levels. If this occurs, the practitioner should normalize the savings estimate to the output changes observed in the period following delivery of the measure.

For measures with pre-conditions baselines and RUL of ten years or less, two baselines must be modeled in estimating savings over the lifetime of the measure. The first applies between the time of implementation and when the RUL expires. The second baseline is between the time when the RUL expires and the measure lifetime expires.

4.3.1.5. Efficient Conditions

The efficient conditions of the affected systems and equipment are those observed after measure has been delivered. This period is also referred to as the post-period. The practitioner should use the data collection methods described in section 4.3.5 to determine relevant efficient-case conditions of the measure. Examples of relevant conditions include, but are not limited to:

- Loads, e.g., air or water flow, Btu/h, cooling tons, and conveyance delivery rates
- Equipment performance , e.g., sizing, performance curves, part-load efficiency curves
- Control sequence and set points
- Envelope thermal properties, e.g., glazing U-values, ventilation rate, and insulation levels
- Distribution system properties, e.g., leakage, pressure drop and insulation levels
- Operations and maintenance practices
- Weather conditions are often a significant determinant. The selected savings estimation model may require actual weather for periods before and after measure delivery. However, final savings estimates should use typical weather conditions. For typical weather, the practitioner should use the most recent TMY (Typical Meteorological Year) weather file from the station that best represents weather at the end user site.
- Other significant determinants of savings

4.3.2. Measure Delivery Verification

The goal of delivery verification is to ensure the measure operates as intended and is capable of achieving energy savings. The practitioner should inspect the measure, discuss its operation with relevant end user staff and vendors, and examine installation documentation and test results. The practitioner may need to request additional documentation or perform additional tests if uncertain about the condition and performance of the measure. The savings estimate should be based on the final as-operated conditions of the affected systems. Documentation providing evidence that the measure was installed, is operational and is capable of achieving energy savings constitutes a critical part of the savings report (section 4.5.2).

4.3.3. Selecting a Reliable Analysis Method

The Savings Guidelines do not require formal documentation of a savings estimation plan. However, whether formal or informal, it is important to think through a plan for each measure. This section provides guidance on how to select a reliable model for estimating savings. Guidance on data collection methods is provided in section 4.3.5.

4.3.3.1. Applicable Energy Use Models

There are two types of energy use models, which may be applicable:

- **Statistical Models**. These models establish the relationship between energy use for affected systems and significant determinants such as outdoor temperature. The models are fit to measurements taken before and after delivery of the measure. The models are then used to estimate energy use for a typical year. The savings from the measure is equal to the difference between these two estimates of annual use. More detailed guidance on how to formulate statistical models and example applications are provided in ASHRAE (2002), BPA M&V Protocols (2011), BPA M&V Guidelines (2011).
- **Engineering Models**. Engineering models rely on thermodynamic, heat transfer and other physical principles to estimate energy usage for systems and equipment. More specific guidance on the selection, development and use of engineering models is provided in ASHRAE (2002), and BPA M&V Protocols (2011). There are two important sub-categories of these models.
  - **General Purpose Software**. DOE 2.1R, eQUEST and AIRMaster+ are all examples of this sub-category. Such software should be well documented regarding the algorithms used and available data input options. The practitioner should understand the strengths and limitations of the software before concluding that it is appropriate for modeling a specific custom measure.
  - **Custom Models**. The practitioner may also consider using an engineering model built specifically for the Custom Protocol measure. These are often developed within a spreadsheet and are often bin models, e.g., loads averaged by 2 degree F outdoor temperature bins.

The key difference between statistical and engineering models is their applicability to different measure baselines.

- **Current Practice Baseline**. Only engineering models can be used for a measure that requires current practice baseline. Statistical models must be fit to trend metering from both the baseline and efficient-case periods. By definition, a current practice baseline for a measure is something that cannot be directly observed or measured at a specific end user location. Only an engineering model can simulate current practice conditions based on the assumed physical properties of that baseline.

- **Pre-Conditions Baseline**. Both statistical and engineering models may be applicable to a measure with a pre-conditions baseline. For both types of models, the primary challenge is to demonstrate that the model can be calibrated (or fit) to the available trend metering representing baseline and efficient conditions and used to reliably estimate annual energy use.

4.3.3.2. Method Selection

Method selection is based on both the applicability of the model (discussed above) and the energy use characteristics of the affected systems and equipment. Measures can be divided into two categories as a first step in selecting the most appropriate savings estimation method.

- **Constant load and timed schedule**. The affected systems operate under reliable automatic control (which may be simply continuous operation or based on a time clock) and for each mode of operation the power that they consume is constant. Further, power consumption during each operational mode and the duration of each mode can be confirmed by inspection (which may include one-time measurements). Examples of this type of measure include, but are not limited to efficiency improvements to:
  - Lighting (in unconditioned spaces) under time-clock control
  - Constant volume air handling units under time-clock control (fan energy savings only)
  - Constant speed and constant head water treatment plant pump operation (24/7)
  - Constant-speed computer room air-handling unit fan operation (24/7)
  - Water park or community pool pumping/filtration operations
- **Variable load or variable schedule**. Custom Protocol measures belong to this category if either or both these criteria are met: (a) the load served by the affected systems varies over time based on the determinants of savings, e.g., outside air temperature or production level; (b) the operating schedule for the affected system or equipment vary over time based on these same or different determinants. Examples of this type of measure include, but are not limited to efficiency improvements to :
  - Constant-speed cooling tower fan operation (operation varies with temperature)
  - Hot water or chilled water pumping, no VFD (operation varies with boiler/chiller operation)
  - Wastewater treatment plant air blowers maintaining constant dissolved oxygen level
  - Industrial 2-speed cooling tower fan operation (speeds controlled by process)
  - Variable air volume air handling unit (AHU) under thermostat control

Figure 3 illustrates the process for selecting a savings estimation method for a Custom Protocol measure. The first step in the process is to determine whether the measure has a pre-conditions or current practice baseline.

- If the baseline is current practice, the decision paths lead only to engineering models. The first step is to define the features of current practice. See section 4.3.1.4.1 for guidance on how to define a current practice baseline. For current practice baseline, only post-period data collection is relevant, as an engineering model must be used to simulate the baseline energy use under the conditions that are determined by current practice. The energy use characteristics of the affected systems determine what data are needed.
  - Constant load and timed schedule measures require data from on-site delivery verification (including spot measurements), and interviews but they do not need trend metering.
  - Variable load or variable schedule measures require similar on-site delivery verification and interview data, but they also require trend metering of time-varying key parameters.
- If the measure has a pre-conditions baseline, both pre and post data collection are needed and decision paths may lead to either statistical or engineering models. The energy use characteristics of the affected systems determine what data are needed.
  - Constant load and timed schedule measures require pre and post data from on-site delivery verification (including spot measurements), and interviews but they do not need trend metering. This path leads to an engineering model for estimating savings.
  - Variable load or variable schedule measures require similar on-site delivery verification and interview data, but they also require pre and post-period trend metering of time-varying key parameters. For these measures, the practitioner must determine whether a statistical model can be fit to the trend metering and whether that model will reliably estimate both pre and post annual energy use. The practitioner should consider prior similar measures they have analyzed along with guidance provided by ASHRAE (2002), and BPA M&V Protocols (2011) in making this determination. The determination needs to be made before data collection commences, as the data requirements of the selected statistical or engineering models are likely to be different.

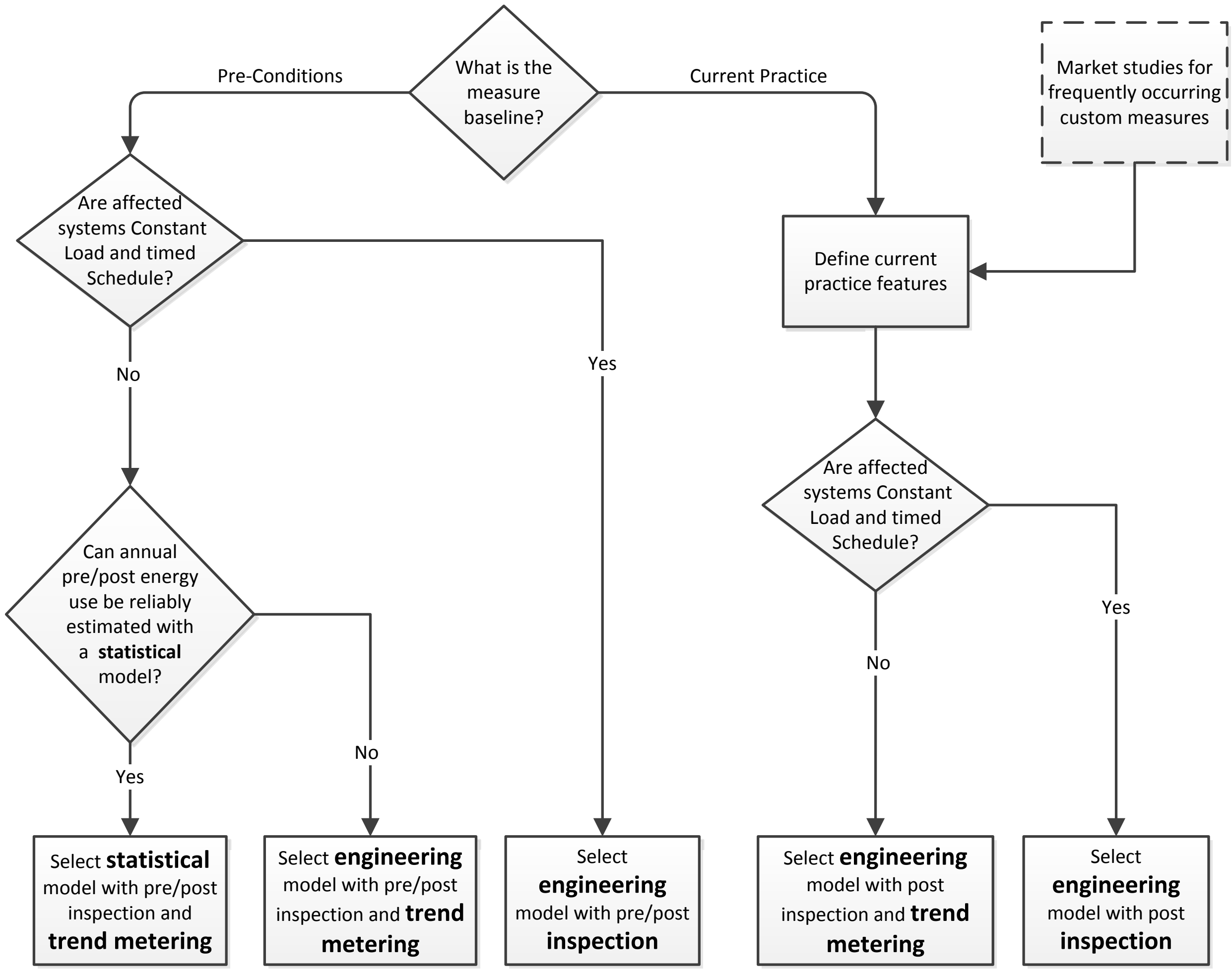

Figure 3: Selecting a Method for Savings Estimation.

4.3.4. Within Measure Sampling

Some measures may comprise many pieces of equipment located throughout an end user site, e.g., motors, fans, terminal units, blowers or light fixtures. It may be possible to estimate the total energy use of all such units by selecting a random sample of units and collecting the necessary data for that sample. Various techniques such as stratification by unit capacity may be useful in decreasing the size of the sample required to achieve the sampling precision goal. In general, sampling should not be used unless it is practical to achieve relative error in the estimate of mean unit energy use equal to or less than ±20% at a confidence level of 80%, without introducing substantial bias. Further guidance on relevant sampling techniques can be found in BPA M&V Guidelines (2011).

4.3.5. Data Collection Methods

This section provides guidance on data collection methods. When data should be collected will be determined by the data requirements of the savings estimation method selected according to the guidance provided in section 4.3.3. For example, if the measure has a pre-conditions baseline, the affected systems have variable loads, and savings are to be estimated with a statistical model, then trend metering of power might be needed in both baseline and efficient-case periods.

4.3.5.1. Inspection and Interview

Data should be gathered from one or more visits to the end user site. Data are obtained by inspecting the affected systems and by interviewing end user staff or vendors who are familiar with operations, maintenance and performance of the affected systems. Inspection will help the practitioner understand the system's physical layout and collect nameplate information needed in obtaining manufacturer specifications and performance data. If pre-conditions data are required, the site visit should also be used to document operating conditions of the affected systems. The practitioner needs to come away from the inspection with accurate data on the control sequences and settings. It may be necessary to make arrangements prior to the inspection to have time from staff or vendors who can operate the controls interface so that this data can be observed. This may also be an opportunity to enable relevant trend logging. In either pre or post period, taking pictures is advised as an efficient means of primary data collection and cross check with other forms of data collection.

4.3.5.2. Trend Metering

The method selected and the configuration of the affected systems will determine the number and location of trend metering points. These points might involve metering of power, temperature, pressures, flow rates, status or many other parameters over a period of time and at intervals required by the model. The number of points and their location may also be determined by within-measure sampling as discussed in section 4.3.4. The duration of the trending will be determined by the need to observe a large portion of the variability in energy use and significant determinants. The goal is to achieve reliable measurements for each metering point for the required periods and achieve that at the lowest cost possible.

Safety is a paramount concern. There are many potential hazards including potential falls, electric shock and damaging interactions with moving parts. Knowledge of relevant safety procedures is required for the practitioner as stated in section 4.2.

In addition, precautions should be taken to prevent any unintended interruption to the end user's equipment or systems. If interruptions are required, they must be approved in advance by the end user and should be carried out by the end user's staff.

4.3.5.2.1. Sensor Selection, Installation and Calibration

A variety of sensors may be needed to satisfy the estimation plan. Some of these may already be in place as part of the end user's control systems. Regardless of whether they are installed under the supervision of the practitioner or are already in place, the practitioner must be concerned about sensor selection, installation and calibration.

- **Selection**. Sensor must be an appropriate type and sized to achieve a measurement with acceptable accuracy. For example, an oversized current transformer (CT) may record values that are only a small fraction of full scale and thus have large errors.
- **Installation**. Sensors must be installed at the correct location. For instance, input power to a VFD is different than input power to the motor it controls, the latter being required if VFD losses need to be included in the power metering. Installation work involving high voltages should be performed by a qualified electrical worker. The practitioner should consider using a member of the end user's staff or the end user's trusted electrical contractor to do this work.
- **Calibration**. Where possible, the practitioner should use sensors that have been recently calibrated. If this is not possible, take redundant short-term measurements with high quality calibrated instruments to confirm that critical sensors are transmitting reasonable values. In some instances, it will not be practical to take either redundant measurements or to have sensors calibrated. It will then be left to the practitioner to determine whether the measurements are usable and to make necessary adjustments to the estimation model.

4.3.5.2.2. Data Acquisition

Data acquisition may be accomplished either by installing data loggers specifically to support savings estimation or by obtaining data from end user monitoring and control system trend logs. In either case, the practitioner needs to confirm at the beginning of any data collection period that the data are being acquired and are usable. Where possible, redundant measurements with high quality calibrated instruments should be compared to output from the data loggers or end user trend logs at the beginning of the data acquisition period. If data collection is needed over extended periods, routine checks should be made to confirm that usable data is being acquired and if not, remedial action should be taken. Especially for pre-conditions data, there may only be a limited window of opportunity for data collection and acquisition failures may be impossible to remediate.

4.3.5.2.3. Preparing Analysis-Ready Trends

Even when applying best practices in sensor selection, installation, calibration and data acquisition, it is still likely that the data acquired will contain some defects. The practitioner should establish range gates (expected high and low values) and examine all data that falls outside these bounds. One common occurrence is that small calibration errors can appear as negative values in the trend logs. Checks should also be made for unexpected sequences of identical values that do not correspond to known modes of operation for the equipment being measured. Other checks should also be performed to identify unexpected relationships between measurements and their primary determinants such as power to a chiller and outside air temperature.

Each of these checks described above may result in the practitioner not being able to use certain intervals of measurements. These should be eliminated from the data used in savings estimation. It is good practice to automate these data editing actions, e.g., using formulae in a spreadsheet, to document how the data has been modified. If this is not possible and data must be edited manually, it should always be done on a copy of the primary data and documentation prepared that describes how the data were modified.

4.3.5.3. Secondary Sources

Many valuable pieces of information can be acquired from secondary sources. These include but are not limited to the following.

- **Design documents**. These will show location, connections and specifications of affected equipment. Be wary of deviations between such documents and actual conditions. This is true even if they are of recent vintage and labeled as as-built documentation. Spot verification is always a good idea if practical.
- **Manufacturer specifications**. These include results from standardized tests and performance curves, nameplates, and other data. These will be particularly critical when engineering models are used to estimate savings. Acquiring this information from the contractor or trade ally most directly involved in the measure's delivery is typically preferred.
- **Equipment databases**. These include equipment performance data or specifications. An example is MotorMaster+ , which contains very useful data concerning the efficiency curves for motors of many types and sizes.
- **Current practice baseline**. See section 4.3.1.4.1 for listing of the data that may be relevant to establishing the properties of current practice baselines.
- **Weather data**. In many cases, it is necessary to use site-specific trends for weather parameters. The National Weather Service maintains and extensive array of measurement sites that can often meet the needs for weather data in the estimation model.

4.3.6. Savings Estimation

The practitioner should estimate site-specific savings for Custom Protocol measures using the data collected in accordance with section 4.3.4 and 4.3.5 and a reliable analysis method selected in accordance with section 4.3.3.

4.4. Quality Standards

The savings estimation method should rely on the best practical and reliable model.

- Practical means that the required data collection and analysis can be carried out with available resources. Generally, the budget available for this work should not exceed 10% of the measure cost (not the incentive cost).
- Reliable means that the method includes tests of the model that demonstrate it is free of substantial measurement bias. For statistical models this requires at least achieving a good fit with the trend metering. For engineering models, the trend metering should either be

used directly, such as in a spreadsheet bin model or used indirectly to calibrate a simulation model. In addition to fit and calibration the practitioner should show convincing evidence that the model (statistical or engineering) accurately extrapolates short-term trend metering (if that is all that is practical) to estimate annual energy use. Finally, to be considered reliable, a method that involves within-measure sampling should satisfy the relative error target in section 4.3.4.

Detailed quality control procedures are left to each agency responsible for operating programs that deliver Custom Protocol measures in the region. However, each agency is encouraged to consider incorporating a peer review process in their quality control procedures. Having practitioners check each other's work not only contributes to quality but it also propagates skills and experiences among the region's practitioners. This is especially true if practitioners employed by different firms perform the peer review.

4.5. Documentation Standard

4.5.1. Site-Specific Savings Estimation Plan

The Savings Guidelines do not require formal documentation of a savings estimation plan. However, a site-specific plan should be formulated for each Custom Protocol measure that will result in a reliable estimate of savings for the measure. If the plan specifies baseline data collection, that data must represent typical conditions found at the site during the baseline period. Similarly, the plan must represent typical conditions after measure delivery (post-period). The plan must specify the data analysis and modeling to be used to estimate savings and must call for the collection of all data needed to satisfy the input requirements of the savings estimation model.

Frequently, various aspects of the planned data collection and analysis will need to be adjusted as they are implemented. There is no general requirement to formally document either the initial plan or updates to the plan as they occur. Such formal documentation requirements for the plan are left to the discretion of the program delivering the measure. However, the as-implemented data collection and analysis must be documented in the site-specific savings report described in the next section.

4.5.2. Site-Specific Savings Report

The savings estimation report should address the topics specified below. Other reporting requirements may be specified by the program operators delivering the measure. The report should also conform to the transparency requirements in section 4.5.3.

- **Measure Description**. Description of the baseline and post-period conditions of the affected system(s). Includes a summary of the measure and the mechanism by which it changes energy use.

- **Delivery Verification**. Description of the inspection procedure, testing, and documentation review completed by the practitioner to determine whether the measure operates as intended and is capable of achieving energy savings.
- **Data Collection**. Description of the data collected during the baseline and post periods. All measurement points should be enumerated along with the calendar period of data collection, the data logging interval, sensor type and placement, and data logging method. All other data directly relevant to the savings calculation such as equipment performance specifications should be listed along with the source of the information.
- **Within-Measure Sampling Procedure**. If applicable, this section describes the method for sampling units. Sampling is applicable when the measure comprises a large quantity of units at a specific site. The sampling objective should be described, such as estimating the mean unit capacity, along with the information available for the population of units that allows for the relative error of the sample to be tested
- **Savings Estimation**. Description of the computational procedure used to estimate the annual change in use for all affected fuels. All input assumptions and the source of each must be documented. For measures with pre-conditions baselines and RUL of ten years or less, the report should document or reference the assumptions and analyses performed to estimate savings between the time when the RUL expires and the measure lifetime expires.

4.5.3. Transparency

The report should be sufficiently detailed to allow for quality control review by an appropriately skilled analyst. The data collected, data editing and data analysis should be documented so that the reviewer can reproduce the savings results, assuming that they have access to the software used by the practitioner. If the practitioner performs the analysis in custom spreadsheet models they should adhere to the following practices:

- Organize sheets into sections for each savings calculation:
  - Summary of Results
  - General Fixed Inputs – baseline and post
  - Curve Fits – baseline and expected post
  - Equations – list and explanation
  - Calculations – by category (occupancy, equipment status, day type, etc.)
- Constants should be labeled and placed in their own cell. Do not bury literal constants inside formulas
- Explain any uncommon constants
- List equations including explanations of variables
- Use names for variables in formulas instead of cell references as much as practical
- Consider breaking long calculations into multiple steps where helpful for clarity.

- Where a breakup of a long calculation will increase clutter, thereby reducing clarity, provide an explanation of the calculation in a cell comment or on a separate worksheet.

If the practitioner performs the analysis using general-purpose software, such as DOE 2.1R, eQUEST and AIRMaster+, all input data and files should be included in the report documentation. The report should clearly state what runs were performed with the software and how the outputs were used in estimating savings.

If general-purpose software is used, it should be accessible and reasonably priced for all practitioners in the region. Either the software must be inherently transparent or it must be fully documented. Fully documented means the exact algorithms for all calculations are completely described in a document accessible to all practitioners or that the analysis method is documented along with the results of a validation process, similar to ASHRAE Standard 140, which demonstrates the comparability of the method to other accepted calculation methods.

4.6. Method Development and Maintenace

Custom Protocol measures require site-specific plans, estimation methods, data collection and reports for estimating and documenting savings. However, there can be many similarities between delivered measures. Program operators can improve the quality and reduce the cost of estimating savings by sharing plans and reports among practitioners. This is particularly true for sites involving new types of measures or applications of measures for new types of end users and end use systems. Practitioners can reduce costs by utilizing relevant portions of prior plans and reports in the development of methods and estimates for each delivered measure. Appropriate precautions should be taken, as determined by each program operator, to protect the confidentiality of end user data.

5. PROGRAM IMPACT EVALUATION

Program impact evaluations estimate savings from a period of program operation. This section provides guidance on how to conduct program impact evaluations that result in reliable estimates of savings from programs that deliver any combination of Unit Energy Savings (UES), Standard Protocol and Custom Protocol measures. The audience for this section is any evaluator, either staff of an agency that operates programs or staff of an independent evaluation firm working under contract to such agencies.

Programs vary widely in delivery method, target markets and delivered measures. Programs are operated in this region by a wide variety of agencies. Private retail utilities and the Energy Trust of Oregon operate under the oversight of state regulators and public utilities have a similar relationship to general or special purpose local governmental boards. Impact evaluations should be designed to achieve reliable estimates of savings while accommodating the special requirements of the program's delivery methods, target markets, efficiency measures, operating agency, and regulatory environment.

5.1. Eligible Measures

The energy savings from any program may be estimated in accordance with the guidance of this section. How each of the measures, which comprise a program is treated during an impact evaluation is determined by its savings estimation method as shown in Figure 4, and further discussed in section 5.4. The figure illustrates the sampling process and shows that programs may comprise measures having one or more savings estimation methods.

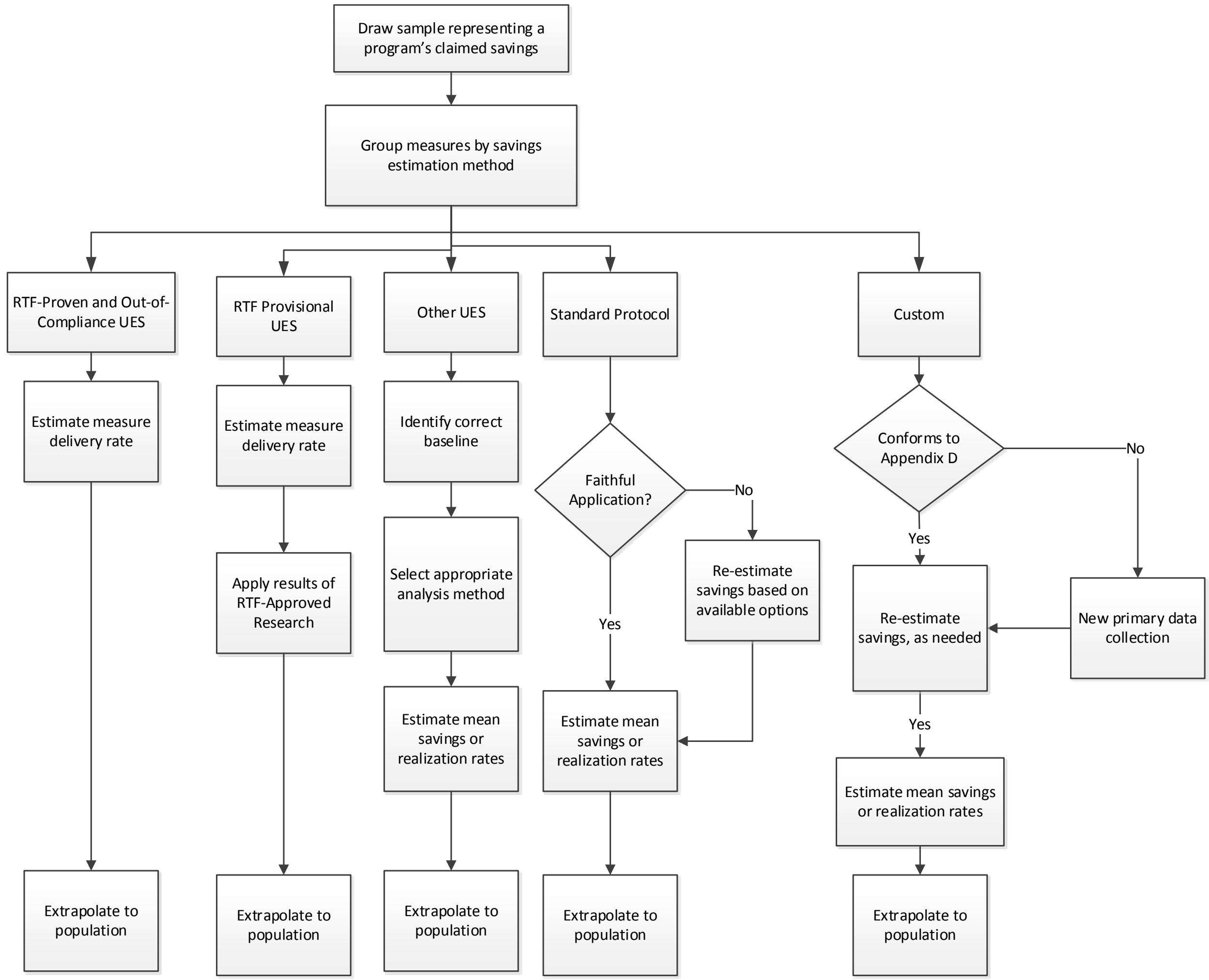

Figure 4: Overview of Program Impact Evaluation Process

5.2. Required Knowledge and Skills

Generally, a team of professionals performs program impact evaluations. The term "evaluator" refers to the team responsible for a specific program impact evaluation.

The evaluator should have a full understanding of the following:

- These guidelines, particularly as they pertain to program impact evaluation.
- Savings values, specifications, baseline conditions and effective dates for RTF approved UES and Standard Protocol measures that are relevant to any particular program impact evaluation.

The team should be able to successfully perform the following tasks:

- Select representative and efficiently designed samples, i.e., maximizing precision for a given sample size.

- Collect and prepare analysis-ready site-specific data, e.g., surveys, inspection, measurement and billing data.
- Estimate savings using a variety of engineering and statistical techniques for sampled measures.
- Extrapolate sample findings to the study populations and quantify the uncertainty in this extrapolation.
- Prepare transparent and clearly written report describing the study methodology and findings.

5.3. Portfolio Assessment

For any agency delivering energy efficiency measures, a portfolio is the collection of all measures claimed to have been delivered by the agency's programs during a given period. A portfolio may comprise a number of programs and their constituent measures. An agency's evaluation plan for a given period will comprise one or more studies, which will estimate energy savings (referred to as savings in the balance of this section) for portions of a portfolio's programs or measures.

The purpose of this portfolio assessment is to make decisions about the impact evaluation of programs, not about how to fund or authorize programs. The portfolio assessment provides an opportunity to adjust and modify program impact evaluations so they can adapt to expected changes in the character of the programs.

The assessment has three primary purposes:

- Identify the programs that need impact evaluations
- Identify the programs expected to involve pre-conditions baseline measures (see section 5.4.2) before the start of the reporting period so there is time to plan for and deploy pre-conditions data collection efforts[5].
- Set the confidence and relative error target for each impact evaluation study

5.3.1. Factors to Consider During the Assessment

The following factors should be considered during the assessment:

- Size of savings
- Level of uncertainty about the savings
- Degree of innovation

[5] Study designs that rely solely on billing data, to characterize pre-conditions, may only need to ensure that the appropriate period of data collection will be available for the sampled end users.

5.3.1.1. Size of Savings

Programs with large expected savings are always candidates for evaluation. Thus, the first criterion is the size of the expected savings for each program as a percent of total portfolio savings. If a program warrants an impact evaluation, then the evaluator should consider the relative size of the savings for the measures that comprise the program. It may be appropriate to allocate evaluation resources to the measures with greatest savings.

5.3.1.2. Uncertainty of Savings

One of the main purposes of any evaluation is to reduce uncertainty about key parameters associated with savings. The uncertainty around any one key parameter (delivery verification rate, savings realization rate, and mean savings) should help to guide the allocation of finite evaluation dollars. For example, if lighting is an important measure, there are two key parameters of interest—delta Watts and operating hours. If there is a fair amount of confidence in the delta-Watts parameter, but not in the operating-hour parameter, then evaluation resources should be shifted from estimating delta-Watts to estimating operating hours. By doing so, the uncertainty around savings of lighting measures will be minimized.

5.3.1.3. Degree of Innovation

Recent programs that are innovative, in terms of their delivery methods or the measures delivered, are more likely to have implementation problems compared to mature programs and thus have saving which are more uncertain. Innovative programs, if they are expected to have large market potential, should receive, all things being equal, a larger portion of the assigned evaluation budget.

In addition, any program that has substantially changed since its last evaluation (e.g., measure specification, incentive levels, delivery methods and participant characteristics) may be a candidate for evaluation.

5.3.2. Programs that Need Impact Evaluations

Considering all the factors discussed above, programs that meet the following criteria should be evaluated:

- A large (one expected to account for more than 10% of portfolio savings for a reporting period), mature[6] program whose savings are uncertain because the delivery approach, measure mix or end user mix has substantially changed since its last evaluation.
- A large, mature program that has been not been evaluated during the previous three years.

[6] A mature program is one whose delivery methods have been routinized, are considered effective, and have not changed over a number years.

- An innovative program with highly uncertain savings that is expected to eventually achieve large savings.

The sum of savings for programs that are expected to account for less than 10% of the portfolio should not exceed 20% of portfolio savings. When this condition cannot be met, programs with similar delivery methods that are targeted at similar markets should be combined and evaluated like large programs.

5.3.3. Program Savings Relative Error

The design of each program impact evaluation should target a relative error in the estimate of savings equal to or better than ±20% at a confidence level of 80% and be free of substantial bias. As explained in section 5.4, this level of confidence and precision is not a requirement since there are many legitimate reasons why it might not be achieved. Ultimately, the evaluator must have the flexibility to respond to data issues as they arise in order to maximize the reliability of the savings. While this section does not provide a procedure for estimating relative error at the portfolio level[7], the portfolio-level estimate of savings will in most cases meet or exceed the target set for the individual programs, if reasonable efforts are made to achieve the target for each program evaluation.

5.3.4. Savings Claim Verification

Program operators track the delivery of measures under each program and claim savings for each period of program operation based on these accomplishments. Each program's design must include methods for documenting measure delivery. Such proof varies based on program delivery design and measure. An upstream program might document shipments of efficient products to distributors or retail outlets, by type of product. At the other extreme, a custom grant program might require documentation from detailed post-period site inspection of the delivered measures. The portfolio assessment should include claim verification for all programs. This would involve review of documentation for reliable samples for each program to determine whether the delivered quantity, measure applications and eligibility data are adequate and that the savings are consistent with the electronic database used in reporting the program claimed savings. If the program needs an impact evaluation, the impact evaluation sample may be used to complete the claim verification.

5.4. Planning a Program Impact Evaluation

Evaluators should consider four primary issues in formulating a plan for each program impact evaluation:

- Type of Measure (as defined by the applicable savings estimation method)

[7] The quantification of portfolio-level error is complex. Given the target for individual studies, it may not be worth the resources to quantify.

- Baseline (Current Practice or Pre-Conditions)
- Research Design
- Sampling Plan

Planning for program impact evaluation should occur prior to program initiation. This is particularly important for programs involving measures with pre-conditions baselines (section 5.4.2.2). For these programs, the plan should be in place so that measures can be sampled and baseline data collected according to the evaluation design.

5.4.1. Type of Measure

Measures delivered by programs should be categorized and treated as described below.

- **Unit Energy Savings (UES)**. The treatment of these measures depends on their RTF status and stage of development.
  - Proven UES Measures. Savings for these are estimated as described in Section 5.4.6.1.1.
  - Provisional UES Measures. For these measures, delivery verification is required as for Proven measures (Section 5.4.6.1.1). The savings estimates for these measures should reflect the results of the Provisional measure's RTF approved research plan.
  - Other UES Measures. This type includes the UES categories of Small Saver and Planning. Also included are UES measures that have been created by program operators but are not recognized by the RTF. Savings are estimated by conducting one or more studies that may require site-specific data collection. A general discussion of the methods for estimating savings for Other UES Measures is presented in section 5.4.6.1.2.
- **Standard Protocol**. Savings are determined by conducting a "faithful application" review for a sample of delivered measures as described in Section 5.4.6.2.
- **Custom Protocol**. Savings are estimated by conducting one or more studies that may require new site-specific data collection, as described in Section 5.4.6.3. For portions of the sample whose savings estimates are prepared in conformance with section 4, analysis or data collection may not be required.

5.4.2. Defining Measure Baseline

There are two possible baselines, current practice and pre-conditions, defined in the Roadmap. Determining the appropriate baseline is a critical step in estimating savings for all measures. For RTF Proven, Provisional, Small Saver and Planning measures, the RTF has already determined the appropriate baseline in estimating the UES value. Similarly, the RTF has determined the appropriate baseline for RTF approved Standard Protocol measures. Refer to measure listings on the RTF website for more details. Evaluators responsible for a study must determine the correct baseline for Other UES (Small Saver, Planning, and non-RTF UES) measures not recognized by the RTF. Starting with a review of program operator documentation the evaluator also determines whether Custom Protocol measures have the appropriate baseline.

5.4.2.1. Current Practice

In determining what constitutes "current practice," the evaluation needs to focus on what equipment choices and installation configurations would have been adopted in the absence of the program, the so-called counterfactual (Mohr, 1995; Salmon, 1998). This can be challenging if the measure baseline is not governed by prevailing codes and standards. There are a number of possible indicators that current practice is the appropriate baseline, as defined in the Roadmap. Studies to establish the typical performance characteristics of current practice equipment or practices may be more appropriate for collaborative region-wide efforts than for a single program evaluation.

5.4.2.2. Pre-Conditions

The program operator can establish pre-conditions by collecting data (billing data, site inspections, measurement of equipment performance and operation) prior to measure delivery. As described in Section 5.4.5.4, the program operator should conduct this data collection in conformance with the sample design, procedures, training, and oversight provided by the evaluator. When the required data cannot be collected for the period prior to measure delivery, it may be possible to reliably estimate pre-conditions based on interviews with those at the participant site and a review of available documentation. As needed, adjustments to the observed or estimated pre-conditions should be made to normalize savings to:

- Post-period operating practices (including occupancy and production levels), which are not part of the changes caused by the measure
- Typical weather conditions

5.4.3. Selecting an Evaluation Approach

A program may deliver one or more types of measures as illustrated in Figure 4. The evaluation approach will be dictated by a program's complement of measures and their respective savings estimation methods. In all cases, savings are estimated for a representative sample. The results from the sample are extrapolated to estimate program savings. This extrapolation may be based on either a savings realization rate or mean savings, whichever results in the smallest estimation error. The savings realization rate is the ratio between the program operator claim of savings and the program impact evaluator's estimate of savings.

The following sections describe the decisions that should be made in selecting an evaluation approach for each type of measure.

5.4.3.1. UES Measures

5.4.3.1.1. RTF Proven UES

For RTF Proven measures, the research design is relatively straightforward since the RTF has already established the UES values. Program savings are largely dependent on measure delivery verification as described in section 5.4.5.1.2. The remaining decisions are:

- Targeted level of confidence and precision for the measure delivery verification rate
- Required sample size
- How the results will be extrapolated to the population.

5.4.3.1.2. Other UES Measures

For Other UES (Small Saver, Planning, and non-RTF UES) measures, the designs are more complex, requiring the evaluator to determine:

- The targeted level of confidence and precision for various parameters such as the measure delivery verification rate, realization rate, or mean savings
- Required sample size
- How to identify and manage the sources of bias
- For measures other than RTF Small Saver and Planning, the correct baseline (current practice or pre-conditions) for estimating savings
- How, for a pre-conditions baseline, key data can be collected by the program operator and what training, procedures, and oversight must be provided to the program operator to ensure the success of the data collection
- The correct method for estimating savings (see section 5.4.6.1.2), and the associated mean savings or realizations rates
- How the results will be extrapolated to the population.

5.4.3.1.3. Provisional UES Measures

The research design for these measures is described in the approved RTF research plan (see 2.4.2).

5.4.3.2. Standard Protocol Measures

For Standard Protocol measures, the design decisions are

- The targeted level of confidence and precision
- Required sample size
- How to implement various aspects of the faithful application review (see section 5.4.6.2)

- How the results will be extrapolated to the population.

The RTF approved Standard Protocol defines all the data collection and analyses needed to reliably estimate savings for each sampled measure. The evaluator is responsible for specifying the sample of measures for which the protocol needs to be applied as part of any program's evaluation. To reduce costs, this may be the same sample used by the program operator in establishing the savings claim for a reporting period. For Standard Protocol measures involving pre-conditions baseline, the evaluator must determine whether the same sample is to be used prior to the start of the reporting period.

5.4.3.3. Custom Protocol Measures

For Custom Protocol measures, the design challenges are complex, requiring the evaluator to determine:

- Whether the savings calculations conform to section 4.
- The targeted level of confidence and precision for various parameters such as the measure delivery verification rate, realization rate, and mean savings.
- How to identify and manage any sources of bias.
- Whether the correct baseline (current practice or pre-conditions) has been used by the program operator in estimating savings.
- How, for a pre-conditions baseline, key data can be collected by the program operator and what training, procedures, and oversight must be provided to the program operator to ensure the success of the data collection.
- The correct method for estimating savings (see section 5.4.6.3), and the associated mean savings or realizations rates.
- How the results will be extrapolated to the population.

5.4.4. Methods for Addressing Sources of Error

An impact evaluation design should mitigate various sources of error in estimating savings. Figure 5 presents a typology of the various sources of error

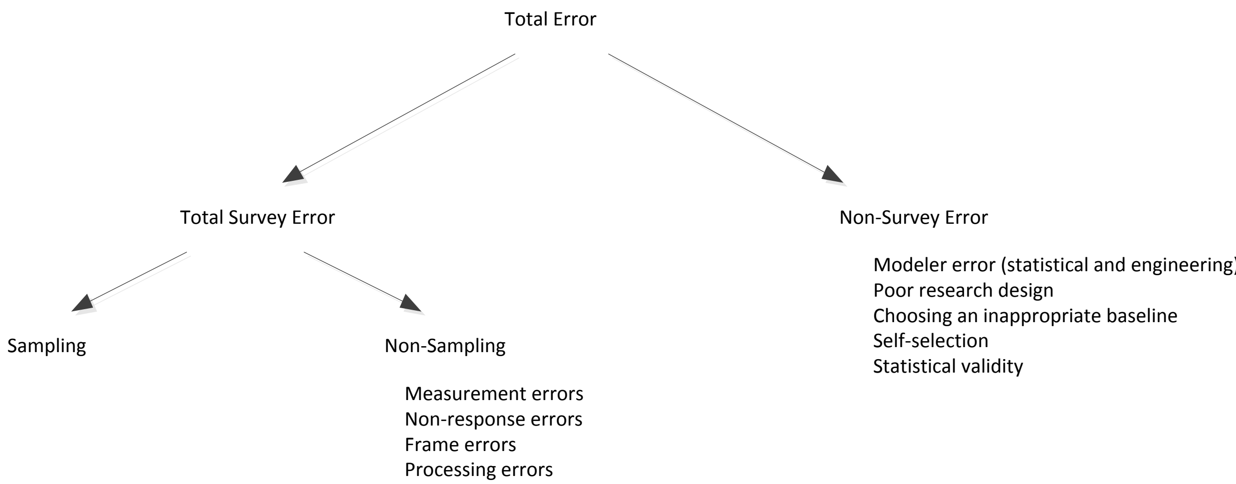

Figure 5: Sources of Error in Estimating Savings

Some of the more important sources of error shown in Figure 5 include:

- **Sampling Error**. It is rarely possible to conduct a census of any program participant population. Generally, program populations are large and the cost of a census would be prohibitive, and when the populations are large, a census is not required to achieve the necessary reliability. Instead, random samples drawn from these populations are used as a way to estimate various characteristics of these populations. The specific approaches to sampling are left up to the evaluator, but must be in accordance with accepted statistical standards. For example, one can choose from a variety of sample procedures recognized in the statistical literature, such as sequential sampling, cluster sampling, stratified or simple random samples, and stratified ratio estimators. Any of these, and others, could be appropriate depending on the circumstances. There are many available publications on sampling techniques that can be used as reference, including Cochran (1977), Thompson (2002), TecMarket Works (2004 and 2005, and Sarndal et al. (1992).
- **Measurement Error.** Evaluators must attempt to minimize measurement error associated with all data collected to support either engineering or statistical models. For example, they should make sure that all data collectors (e.g., those collecting project documentation, conducting telephone or in-person interviews, and collecting data on site) are properly trained. In addition, they should establish quality control systems to monitor data as it is collected to identify and address potential measurement problems (e.g., out of range data or missing data).
- **Baseline Error**. Evaluators should aim to insure that the appropriate baseline (current practice or pre-conditions) is identified for Custom Protocol and Other UES (Small Saver, Planning, and non-RTF UES) measures (except for RTF Small Saver and Planning UES measures) by adequately training evaluation staff and providing adequate quality control review.
- **Modeler Error**. For both engineering and statistical models, evaluators should make sure that all modelers are properly trained and given adequate quality control review. For

example, in the case of engineering models, evaluators should train modelers to correctly calibrate model results to the observed energy use. For statistical models, evaluators should make sure modelers conduct the appropriate regression diagnostics (e.g., influential observations, autocorrelation, multicollinearity, omitted variables) and take steps to mitigate any identified problems.

Some evaluators make the mistake of focusing almost exclusively on reducing sampling error by insisting on large samples. Relatively little attention is devoted to addressing the many other sources of error. As a result, some studies achieve a high level of confidence and precision around a biased estimate. Therefore, focusing on sample error, while giving relatively little attention to these other sources of error, would compromise the RTF's objective of obtaining reliable estimates of savings. The evaluator must have the flexibility to respond to non-sampling errors as they arise in order to maximize the reliability of the savings (Sonnenblick and Eto, 1995; California *Evaluation Framework,* 2004; California Protocols, 2005).

Section 5.3.3 sets targets for program-level confidence and precision (relative error). These are targets as there are many legitimate reasons why they might not be achieved. For example, in the planning stage, the expected variability in the savings could be so great that it would be impossible to meet the precision target. Or, once the evaluation is launched, survey or on-site non-response could be much higher than expected, requiring more resources devoted to addressing this source of bias and fewer resources devoted to achieving the targeted confidence and precision

5.4.5. Data Collection Methods

This section describes various methods for collecting impact evaluation data from program participants and establishing current practice baseline conditions. The roles of program evaluators and program operators are also clarified and the value of regional collaborative studies is discussed.

5.4.5.1. Site-Specific Data Collection Method

Impact-related data may be collected for specific sites using survey research, on-site verification, and on-site measurements.

5.4.5.1.1. Survey Research

Data may be collected from participants and non-participants using mail questionnaires, telephone interviews, on-line tools, and in-person interviews to estimate impact evaluation parameters, e.g., operating hours, temperature set points, and occupancy. Surveys often limit their measurements to those that can be standardized and repeated over large numbers of persons. Surveys gain their inferential power from the ability to measure groups of persons that are representative of large populations.

Part of the evaluator's task is to determine how to minimize the total survey errors shown in Figure 6.

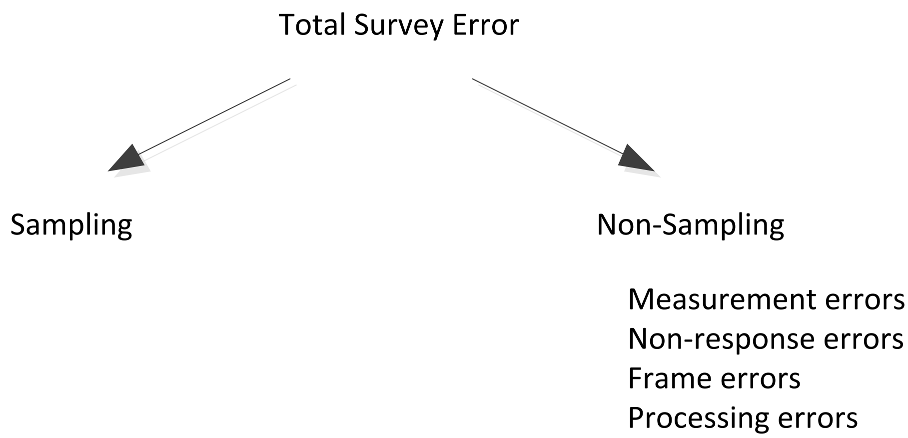

Figure 6: Total Survey Error

Some of the most important decisions include:

- How will the potential sample members be identified?
- What approach will be taken to contact those sampled, and how much effort will be devoted to trying to collect data from those who are hard to reach or reluctant to respond?
- How much effort will be devoted to evaluating and testing questions that are asked?
- What mode (e.g., mail questionnaire, telephone interview, on-line tools, and in-person interview) will be used to pose questions and collect answers from respondents?
- What is the upper limit on the length of the data collection instruments?
- If interviews are involved, how much effort will be devoted to training and supervising interviewers?
- How much effort will be devoted to checking the survey responses for accuracy and internal consistency?
- What approaches will be used to adjust the survey estimates to correct for errors that can be identified (e.g., item non-response and unit non-response)?

There are many available publications that address all these issues and more, including Biemer et al. (2004), Lyberg et al. (1997), Groves (2004), and Biemer and Lyberg (2003).

5.4.5.1.2. Claim and Delivery Verification

With the exception of Standard Protocol and Custom Protocol measures that conform to the reporting requirements of section 4, all sampled measures in any program evaluation should receive both claim and delivery verification. Standard Protocol measures and conforming Custom Protocol measures may only receive claim verification (section 5.3.4).

5.4.5.1.3. Primary Data Collection for Other UES and Custom Protocol Measures

Studies of Other UES (Small Saver, Planning, and non-RTF UES) and Custom Protocol measures will require detailed planning for appropriate on-site data collection. The data collection needs to focus on the most uncertain and significant determinants of savings such as operation hours or part load efficiencies. See section 4 for further guidance on the identification of significant determinants. What data collection is needed for the sampled sites will depend on whether the measure is current practice or pre-conditions baseline and what engineering or statistical model is best suited to estimate savings. As discussed in section 5.4.6.1.2, these models may treat each sampled site separately or they may involve regression techniques applied to groups of sampled sites. Some of the choices that need to be made for Other UES measures are similar to those faced by practitioners in estimating savings from Custom Protocol measures. This is particularly true if the savings estimation process calls for site-specific savings estimates. In that case, the evaluator should reference section 4 for relevant advice.

5.4.5.2. Regional Current Practice

Various data can be collected that can provide some insights into current practice. For Custom Protocol measures, the acceptable sources of current practice information are described section 4.3.1.4.1 For Other UES measures (excluding RTF Small Saver and Planning measures), acceptable sources of current practice information are described in Section 5.4.6.1.2. Collaborative, region-wide studies may be useful in establishing current practice for important and common Other UES measures. Current practice is defined by the RTF for all other types of measures.

5.4.5.3. Other Studies

Occasionally, studies are conducted outside the area served by an agency that might be applicable to programs and measures in the agency's portfolio. If these studies can be used, then the cost of conducting a similar study can be avoided. The results of these other studies must be applicable to programs and measures within the agency's portfolio. Following are some basic criteria that should be considered before using results from other studies. Compared to the agency's service area, the other study should have been conducted:

- On the same sector (e.g., residential, small commercial, large commercial, agricultural)
- With end users with similar demographic and firmographic characteristics
- On similar types of buildings (structure, vintage, etc.)
- In regions with similar heating and cooling degree days (for weather sensitive measures)
- On measures with similar specifications including program delivery methods

In addition, the methods used to estimate the parameter of interest (e.g., savings, operating hours, and full load equivalent hours) should be clearly described so that potential users can judge whether the methods are consistent with these guidelines. In particular, consistency of baseline assumptions (current practice or pre-conditions) is of particular concern when savings values or savings realization rates used are from another study.

5.4.5.4. Role of Evaluators and Program Operators

The roles of program evaluators and program operators, while distinct, can overlap when it comes to collecting key impact evaluation data. The responsibilities of each are listed below.

The program operator should:

- In accordance with these guidelines, and training and oversight provided by the evaluator, collect data and estimate savings for Standard Protocol and Custom Protocol measures.
- Develop and maintain a transparent and well-documented program-tracking database containing key evaluation data such as end user contact information, measure description, delivery verification documentation, delivery dates, and estimated savings. For UES measures, descriptions contained in these databases must be adequate to allow for reliable assignment of delivered measures to the appropriate UES values.

The evaluator should:

- Develop and implement evaluation plans according to the Savings Guidelines
- Design and implement samples.
- Conduct studies as needed to establish baseline specifications for current practice baseline measures.
- Provide detailed specifications and training to the program operator and regular oversight to ensure successful collection of pre-conditions data required for the impact evaluation.
- Estimate total savings.
- Prepare all interim and final evaluation reports.

Evaluators must also adhere to certain standards of professional behavior including the American Evaluation Association's *Guiding Principles for Evaluators* and the *Code of Standards and Ethics* set by the Council of American Survey Research Organizations.

5.4.5.5. Regional Collaborative Studies

Given that evaluation resources are relatively scarce, it is in the interest of all regional stakeholders to look for opportunities to maximize the cost-effectiveness of impact evaluations. One way would be to pool their resources in conducting studies that are of regional interest. The greatest opportunity for collaborative studies would arise when multiple program operators are delivering a UES measure that is currently not Proven. Such studies can be designed to provide the information needed to develop a Proven measure or to resolve compliance issues for RTF measures that have been assigned the status out-of-compliance. This will substantially reduce the future cost of impact evaluation studies for all program operators who deliver the measure.

5.4.6. Savings Estimation Methods

Because programs typically comprise different types of measures (section 5.4.1), different research designs and analytical techniques will be needed to estimate savings.

5.4.6.1. UES Measures

5.4.6.1.1. RTF Proven Measures

Delivery verification is carried out for a reliable random sample. Modeling tools may be used to to mitigate selection bias. For the sample, information is obtained from either documentation or direct inspection, which is needed to match the verified units to the measure specifications (see the Roadmap). This allows a UES value to be associated with each delivered unit that is consistent with the latest version of RTF approved values prior to the program delivery period. Savings for the units delivered during a program period can then be computed as the sum of the delivered count multiplied by the respective UES value for each measure.

5.4.6.1.2. Other UES Measures

For other UES (Small Saver, Planning, and non-RTF UES) measures, the appropriate approach depends on whether the baseline is pre-conditions or current practice.

For measures with pre-conditions baseline, a participant pre/post design could be used that employs any of the following estimation techniques:

- Regression models involving groups of participants with monthly pre- and post-installation billing and weather data (e.g. cross-section time-series or interrupted time-series designs);
- Site-specific statistical model to estimate pre and post energy use with the difference representing savings;
- Calibrated site-specific engineering models (e.g., DOE-2) to estimate pre and post energy use with the difference representing savings.

For measures with current practice baseline, the study design could employ either site-specific or statistical approaches:

- **Site-Specific Approach**. Take post-only participant measurements and compare the average efficient-case energy use with one that represents current practice. Features of the current practice baseline should be determined by applicable codes and standards, or one of the following if they are more efficient, practical to obtain and applicable to the delivered measure’s location.
    - Assumed baseline in the most recent Council Power Plan
    - Vendors’ description of what they would normally do for this type of end user
    - Information on recent shipments or sales of relevant equipment or services gathered from manufacturers, trade associations, distributors, retailers or other studies and

databases that establish current practice. Other databases such as Motormaster can also provide relevant information.

If no other source is practical, the current practice may be determined by the evaluator's opinion on what would normally be done, based on prior experience with similar measures.

Once the relevant features and specification of the current practice baseline for a measure have been established, engineering models as described in section 4.3.3.1 can used to estimate baseline energy use.

- **Statistical Approach**. Estimate regression models involving both participants and non-participants. One of the most powerful designs in terms of defining the counterfactual, although it does have its challenges, is the true experimental design. The research question is, "What do those randomly assigned to the program group do with respect to the purchase and installation of energy efficient measures or the adoption of energy efficient behaviors compared to the control group?" When such designs are not feasible, other quasi-experimental designs (designs in which random assignment is not possible) may be considered. These include non-equivalent control group designs, which rely on statistical analysis of data from existing groups of participants and non-participants. The following sources provide more details regarding using regression analysis in these types of designs: Johnson (1984); Pedhazur (1997) Ridge et al. (1994); Cook and Campbell (1979); Shadish, Cook and Campbell (2002).

5.4.6.1.3. Provisional UES Measures

The savings estimates for these measures should reflect the results of the Provisional measure's RTF approved research plan. The Provisional measure research results should be applied to random samples that are the subject of delivery verification. **If the RTF approved research plan has not been implemented then the measure should be treated as described for Other UES (Small Saver, Planning, and non-RTF UES) measures, potentially requiring primary data collection for a sample of delivered measures.**

5.4.6.2. Standard Protocol Measures

There are two possible approaches for evaluating Standard Protocol measures. In the first, all data collection and analysis is performed by the evaluator. Data is collected and analyzed by the evaluator, in accordance with the approved protocol for a sample of delivered measures. If the measure has a pre-conditions baseline the evaluator will have to coordinate closely with the program operator so that measures can be selected prior to delivery, to all for collection of baseline data.

Alternatively, the evaluator and program operator may collaborate in conducting the study. In which case, the program operator is responsible for implementing the Standard Protocol data collection or data collection and site-specific analysis, in accordance with the Standard Protocol. The program operator carries out this work for a sample that is drawn in accordance with a sample design provided by the evaluator. In this approach, the evaluator verifies that the measures were delivered by inspecting the data and documentation provided by the program operator and confirms that the Standard Protocol is faithfully applied. The following criteria should be used to identify those cases that meet or fail to meet the faithful application standard:

A. The measure must be eligible.

B. There must be evidence that the measure delivery was verified.

C. The data collected, must conform to the requirements of the protocol, including within-measure sampling and Provisional requirements.

D. The savings were estimated in conformance with the RTF approved protocol in force at the time the measure was delivered.

The options available to estimate measure savings depend on the review findings regarding each of these criteria:

- For cases that fail either of the first two criteria (A or B), the savings are set to zero.
- For cases that passed the first two criteria (A and B) but failed the third criterion (C), if the failure rate (of C) is less than 30% of the entire sample, then the mean savings for the cases that did not fail the third criterion (C) can be applied to each that did fail.
- For cases that fail the last criterion (D), the evaluator must re-estimate the saving in accordance with the protocol.
- For cases that meet all four criteria, the savings are accepted and considered verified.

If more than 30% of the entire sample fails the third criterion (C), the savings realization rate from a new sample (only feasible for current practice baseline measures that do not require pre-period data collection) or a sample from the previous evaluation can be applied to the entire sample. The source of these realization rates must be evaluations that conform to these guidelines. If the previous reporting period also has too many failures, mean savings can be estimated after setting all that fail the first three criteria (A, B and C) to zero.

For pre-condition baseline measures, if the evaluator finds that the sample for which data has been collected by the program operator does not conform to the sample design the evaluator provided and that it is not representative of the population then savings may only be claimed for the sample itself and cannot be extrapolated to the program population. In other words, the savings for all measures not sampled is set to zero. For measures with current practice baseline, a new sample may be selected by the evaluator and used to estimate savings in conformance with the Standard Protocol.

5.4.6.3. Custom Protocol Measures

For portions of the sample whose savings estimates are prepared in conformance with section 4, additional primary data collection may not be required, at the discretion of the evaluator. The evaluator uses the data available in the savings estimation report for each of the sampled cases to confirm or re-estimate the savings. When the evaluator determines that a report does not conform to data collection guidance of section 4, new primary data collection is required and the best practical estimation of savings is performed for the sample consistent with section 4.

5.4.6.4. Programs Not Evaluated for a Reporting Period

Finally, programs not evaluated in a period should use the realization rate from the most recent similar evaluation conducted by the agency or, if none is available, the realization rate from the most recently evaluated similar program in the region. These realization rates must be from evaluations that conform to these guidelines.

5.4.7. Population Estimation Methods

When samples are used to estimate savings, various parameters (e.g., measure delivery rate, mean savings, and savings realization rate) are estimated. These estimated parameters are then extrapolated to the population from which they were drawn in order to produce an estimate of the total savings. In addition, if the samples are stratified random samples rather than simple random samples or some variety of cluster sampling, then sample weights have to be taken into account in the extrapolation. In addition, the size of the savings associated with delivery rates and realization rates also have to be considered in the estimation. This section does not attempt to present all the possible calculations but only to highlight the critical issues in estimating population parameters.

The parameters that will typically be estimated, based on random samples and extrapolated to the population, vary depending on the measure classification.

- For RTF Proven and Provisional UES Measures, savings-weighted measure delivery verification rates are estimated.
- For Standard Protocol, Custom Protocol and Other UES (Small Saver, Planning, and non-RTF UES) Measures, either savings-weighted savings realization rates or mean savings are estimated. For these measures, delivery verification rates may be estimated and applied separately or be subsumed in the overall estimate of the savings realization rate or mean savings for each sample measure.

Once estimated, these parameters can be extrapolated to estimate the ultimate parameter of interest, total energy savings. Each of these estimates has some uncertainty, reflected in the standard error that should be used to construct confidence intervals. Useful guidance on how to calculate these parameters and extrapolate these values to the populations from which they were drawn is presented in Cochran (1977) and Levy and Lemeshow (2008).

5.5. Reporting

In order to be able to judge whether the savings estimates are reliable and can be used for planning purposes and assessing progress throughout the region, the evaluation reports must be clearly written, consistently present key variables of interest, and be readily accessible.

5.5.1. Methodological Transparency

The evaluation methods must be described in sufficient detail so that others can assess the reliability of the reported results. The key methodological elements include:

- **Bias**. Discuss what efforts were made to control bias both in terms of the data collection, baseline definition, and the use of both engineering and statistical models.
- **Sample design, implementation, and disposition**. The description of the sample design should provide the targeted relative precision, a priori assumptions regarding variance of the parameter of interest, and formulas for calculating sample sizes. If the sample design was stratified, describe methods used to determine strata boundaries and, if the sample is disproportionate, how the weights were be calculated. The achieved confidence and precision for each parameter estimated and how the sample-based results were extrapolated to the population must also be reported. The extent of any non-response bias should also be reported as well as steps taken to mitigate any bias.
- **All data collection instruments**. All survey questionnaires and on-site data collection instruments and procedural manuals should be provided. For each instrument, a discussion of pre-test results should also be provided.
- **Baselines definitions**. For pre-conditions baseline, evidence that the equipment or practice replaced by the measure had remaining useful life. For current practice baseline, description of how the current practice features where identified.
- **Key steps in the analysis**. The steps in the analysis should be sufficiently clear (including data cleaning methods and model functional forms) that another analyst could replicate the results.

5.5.2. Comparability

Because there are a number of parameters for which it is useful to compare across the region, certain information should be consistently reported. These include:

- Total number of participants
- Total number of delivered measures by type of facility, e.g., single family, multifamily, office, retail
- Total number and types of measures delivered
- Measure specifications including delivery method
- For each estimated parameter (e.g., total annual savings, mean operating hours, delivery verification rate, savings realization rate), estimated values, the population size, the sample size, standard error, and 80 and 90 percent confidence intervals should be reported
- The key terms, see section 1.2, should be used consistently.

5.5.3. Access

Once an evaluation report is completed, it should be converted to a PDF file and links to file placed on a publicly accessible web site. Along with the PDF file, the following information should be provided to enable potential users to judge the relevance of the study to their needs:

- Title of Report
- Author
- Sponsoring agency
- Program operator
- Publication date
- Program year
- End user sector (residential, commercial, industrial, agricultural)
- Abstract

5.6. Peer Review of Evaluation Research Designs

The RTF will play a clearinghouse role for regional collaboration on impact evaluation. Its particular focus will be research that is needed to support the development of Provisional and Proven UES values and standard protocols, but all relevant and useful research will be considered. Parties may bring proposed research to the RTF for review. The RTF will not directly fund such research, but if it determines the research is important it will facilitate peer review of the research design and regional coordination leading to implementation. The RTF will work closely with NEEA's Northwest Research Group in accomplishing these tasks.

5.7. Additional Guidance and Relevant Protocols

Substantial work has been done by many organizations on the development of guidelines and protocols that aid researchers in designing program impact evaluations. Guidelines and protocols that should be considered in the design of impact evaluations include the following.

- *Evaluation Methods for Achieving Diverse Energy-Efficiency Policy Objectives* --Webinar (both audio and supporting materials). (http://library.cee1.org/content/evaluation-methods-achieving-diverse-energy-efficiency-policy-objectives-part-1).
- *Energy efficiency Guidebook for Public Power Communities* (Chapter 14 on evaluation), prepared by Energy Center of Wisconsin. (http://www.ecw.org/publicpowerguidebook)
- *Model Energy Efficiency Program Impact Evaluation Guide*, part of the National Action Plan for Energy Efficiency. (http://www.epa.gov/cleanenergy/documents/suca/evaluation_guide.pdf).
- *Scaling-Up Energy Efficiency Programs: The Measurement Challenge*, prepared by the Alliance to Save Energy to showcase the critical importance of effective evaluation,

measurement and verification (EM&V) of energy savings, especially as the U.S. continues to witness unprecedented growth in investments for energy efficiency. (https://www.ase.org/resources/scaling-energy-efficiency-programs-measurement-challenge).

- California evaluation protocols (http://www.calmac.org/events/EvaluatorsProtocols_Final_AdoptedviaRuling_06-19-2006.pdf).
- California Evaluation Framework (http://www.calmac.org/publications/California_Evaluation_Framework_June_2004.pdf).
- *American Evaluation Association's Guiding Principles for Evaluators* (http://www.eval.org/p/cm/ld/fid=51).
- *Code of standards and ethics set by the Council of American Survey Research Organizations* (http://www.casro.org/?page=thecasrocode).
- *Uniform Methods Project*. U.S. Department of Energy and the National Renewable Energy Laboratory. (http://www.nrel.gov/extranet/ump)

6. References

NREL. 2013. *The Uniform Methods Project: Methods for Determining Energy Efficiency Savings for Specific Measures*. Golden, CO: National Renewable Energy Laboratory, U.S. Department of Energy. Download from: http://www.nrel.gov/extranet/ump.

BPA M&V Protocols. 2011. *Existing Building Commissioning an M&V Protocol Application Guide. Engineering Calculations with Verification Protocol*. *Verification by Energy Modeling Protocol. Verification by Equipment or End-Use Metering Protocol. Verification by Energy Use Indexing Protocol.* Portland, OR.: Bonneville Power Administration. Download from: www.conduitnw.org.

BPA M&V Guidelines. 2011. *Regression for M&V Reference Guide. Sampling for M&V Reference Guide. Measurement and Verification Protocol Selection Guide and Example M&V Plan. Absent Baseline Application Guide*. Portland, OR.: Bonneville Power Administration. Download from: www.conduitnw.org.

BPA M&V Glossary. 2011. *Glossary for M&V: Reference Guide*. Portland, OR.: Bonneville Power Administration. Download from: www.conduitnw.org.

ASHRAE. 2002. ASHRAE Guideline 14-2002 – Measurement of Energy and Demand Savings. Atlanta, Ga.: American Society of Heating, Refrigerating and Air-Conditioning Engineers. Purchase from: http://www.techstreet.com/cgi-bin/detail?product_id=1645226.

Belsey, D. A., E. Kuh, and R. E. Welsch. *Regression Diagnostics: Identifying Influential Data and Sources of Collinearity*. New York: John Wiley. 1980.

Blalock, H. (1970), *Estimating Measurement Error Using Multiple Indicators and Several Points in Time*. American Sociological Review, 35, pp. 101-111.

Campbell, Donald T. and Julian C. Stanley. *Experimental and Quasi-Experimental Designs for Research*. Chicago: Rand McNally College Publishing, 1963.

Cochran, William G. *Sampling Techniques*. New York: John Wiley & Sons, 1977.

Cook, Thomas D. and Donald T Campbell. *Quasi-Experimentation: Design & Analysis Issues for Field Settings*. Boston, MA.: Houghton Mifflin Company, 1979.

Johnston, J. *Econometric Methods*. New York: McGraw-Hill Book Company, 1984

Kerlinger, F. N., and E. J. Pedhazur. *Multiple Regression in Behavioral Research*. New York: Holt, Rinehart, and Winston. 1973.

Kennedy, Peter. *A Guide to Econometrics*. Cambridge, MA: The MIT Press, 1992

Kish, Leslie. *Survey Sampling*. New York: John Wiley & Sons, 1965.

Rossi, Peter and Howard E. Freeman. 1989. *Evaluation: A Systematic Approach*. Newbury Park, California: SAGE Publications.

Sayrs, Lois. *Pooled Time Series Analysis*. Newbury Park, CA: SAGE Publications, 1989.

Sarndal, Carl-Eric, Bengt Swensson and Jan Wretman. *Model Assisted Survey Sampling*. New York: Springer-Verlag, 1992.

Shadish, William R., Thomas D. Cook, and Donald T. Cambell, 2002. *Experimental and Quasi-Experimental Designs for Generalized Causal Inference*. New York: Houghton Mifflin Company

Sonnenblick, Richard and Joe Eto (1995). *Calculating the Uncertainty in Estimates of DSM Program Cost-Effectiveness*. International Energy Program Evaluation Conference: Chicago, IL. pp. 759-768.

TecMarket Works, (2004). *The California Evaluation Framework*. Prepared for the Southern California Edison Company. 2004.

TecMarket Works Team. 2006. *California Energy Efficiency Evaluation Protocols: Technical, Methodological, and Reporting Requirements for Evaluation Professionals*. April 2006.

San Francisco, Calif.: California Public Utilities Commission.

Thompson, Steven K. 2002. *Sampling. (2nd Ed.)*. New York, N.Y.: John Wiley & Sons, Inc.

Weiss, R. S. and M.Rein. *The Evaluation of Broad-Aim Programs: Difficulties in Experimental design and an Alternative*. In C. H. Weiss (ed.) Evaluating Action Programs: Readings in Social Action and Education. Boston: Allyn and Bacon, 1972.

GUIDELINES FOR THE ESTIMATION OF INCREMENTAL MEASURE COSTS AND BENEFITS

REGIONAL TECHNICAL FORUM

Release Date: December 8, 2015

TABLE OF CONTENTS

1. INTRODUCTION

The Regional Technical Forum (RTF) develops and maintains a series of documents that provide guidance on how to assess energy efficiency measures. This document (the Cost Guidelines) provides guidance on how to estimate the incremental costs and benefits associated with the delivery of an efficiency measure.

1.1. Purpose

The purpose of these guidelines is to provide a systematic approach to developing estimates and documenting approaches and sources when estimating costs and benefits. This document refers only to the *incremental* costs and benefits resulting directly from the efficiency aspects of an energy efficiency measure. These guidelines only address the monetizable costs and benefits that result directly from delivery of a measure.

This document does not provide guidance on the following electrical and gas system costs and benefits: energy, capacity, transmission and distribution capacity deferral, line losses, CO_2, risk mitigation credit, and Power Act credit. It also does not provide guidance on energy efficiency program administration costs, measure sponsor cost shares, financing costs, discount rates or program life. These elements of costs and benefits are specified in the ProCost model and inputs adopted and revised periodically by the RTF.

1.2. Key Concepts

1.2.1. Roadmap Concepts

A number of key concepts such as “Measure” and “Savings” are defined in the *Roadmap for the Assessment of Energy Efficiency Measures* (the Roadmap). These Cost Guidelines assume the reader has a thorough understanding of the Roadmap. Additional concepts that specifically apply to these Guidelines are described below.

1.2.2. Measure Costs and Benefits

Measure costs and benefits are the incremental and monetizable results from the delivery of a measure. These costs and benefits are categorized as capital costs, maintenance, operations-fuel, operations – non-energy consumables and other non-energy impacts. Costs represent an increase in the required financial commitment relative to the baseline and are expressed as positive incremental effects. Benefits represent a reduction in the required financial commitment and are expressed as negative incremental effects.

1.2.3. Elements

Elements are discrete measure costs and benefits that may have unique methodologies for estimating values. Depending on the measure, an element may either be a cost or a benefit. For detailed discussion of each measure cost and benefit element, see section 4.

1.2.4. Incremental Impacts

Incremental impacts are the effects of delivering a measure, relative to the baseline condition (i.e., not delivering the measure). Two baseline conditions are defined in the Roadmap: Current Practice and Pre-Condition. For measures with a Current Practice baseline, costs and benefits would be incurred as a result of establishing the baseline condition, e.g., installation of the market average light bulb. Incremental impact is the difference in cost and benefits between the baseline change and delivery of the measure, e.g., the difference between the market average light bulb and a LED bulb. For a Pre-Condition baseline, the baseline is no change and thus baseline costs and benefits are zero. The incremental impact of the measure for pre-conditions baseline is the full cost and benefit associated with delivery of the measure. Measures may have more than one specific application, e.g., single family weatherization, includes both window upgrades and replacements. Therefore, measures may have applications with both current practice and pre-condition baselines.

1.2.5. Substantial Cost or Benefit

All elements or groups of elements that account for a substantial portion of the either the measure's costs or benefits (for any measure application) should be included in the estimate of costs and benefits. A portion of costs or benefits is substantial if the RTF determines that it is likely to account for more than 10% of the measure costs or benefits.

1.3. Treatment of Remaining Useful Life (RUL)

In some cases, the expected lifetime (see Lifetime Guidelines) of a measure with a pre-conditions baseline will exceed the remaining useful life (RUL) of the pre-condition system, equipment or practice. If RUL is ten years or less, costs and benefits should be estimated for the following time periods: a) RUL: time between measure delivery and when the baseline conditions would have changed if the measure had not been delivered, and b) Balance of Measure Lifetime (BML): time between RUL expiration and the end of the measure lifetime. The RTF expects careful consideration of costs and benefits during this period, but does not impose any specific quality standards on these estimates beyond the use of best available data and professional judgment.

The assumptions made and analysis performed to estimate costs and benefits for the first and second periods should be documented, respectively, in the Summary and SummaryRUL sheets of the Measure Assessment Workbook for UES measures. For Standard Protocol measures, these assumption and analysis should be documented in the appropriate sections of the protocol. For Custom Protocol measures, the assumptions made and analysis performed should be documented in the site-specific savings report.

If the estimated RUL is longer than ten years, then the RTF will assume that RUL equals measure lifetime, and only one estimate of costs and benefits is required.

1.4. Development of RTF Approved Costs and Benefits

The best practical and reliable analysis methods and data sources should be used in estimating costs and benefits. Practical means that the required data collection and estimation can be carried out with proven techniques and resources deemed reasonable by the RTF. Measure costs and benefits analyses must provide unambiguous estimates. The adequacy of the estimation is left to the discretion of the RTF. Documentation of data sources, how they were used in estimating costs and benefits, and estimates of uncertainty[1] are necessary for the RTF to determine whether estimates of costs and benefits are sufficiently reliable.

The RTF's role in the estimation of measure costs and benefits depends on which method is used to estimate measure savings:

- **UES Measures –** Require RTF approval of costs and benefits, which are inputs to the ProCost model.
- **Standard Protocol Measures –** Require RTF approval of a protocol for estimating costs and benefits. This may be either specific costs and benefits estimates (for measures where the savings are expected to vary significantly, but not the costs and benefits) or protocols for determining costs and benefits, such as a list of elements to provide estimates for, data collection sources and methods, and analysis approaches.
- **Custom Protocol and Program Impact Evaluation Measures** – Does not require costs and benefits to be reviewed by the RTF. However, costs and benefits should be estimated and documented as described in these Guidelines, as appropriate. The RTF may review the research plan for a Program Impact Evaluation, including costs and benefits estimation methods, if requested.

1.5. Supporting Documents

The following supporting documents are referenced throughout these guidelines.

- **Measure Assessment Template –** The Excel® Measure Assessment Template contains the ProCost model, ProCost inputs and outputs, energy and costs and benefits analyses, and summary tables. It is used to fully document a measure. The template contains the Summary sheet which is used to document the approaches, supporting data and other aspects of the analyses that yield estimates of measure costs and benefits. The template also contains a checklist that can be followed in estimating costs and benefits for a measure.
- **Standard Information Workbook –** This Excel® workbook provides RTF-approved generic values for various parameters needed in completing measure assessments. For example, it

provides hourly costs for various labor categories that can be used in estimating capital and O&M costs for a measure

- **Marginal Cost and Load Shape (MC and Loadshape)**. This workbook contains marginal energy costs for electricity and gas.

2. Estimation of Costs and Benefits

The following presents guidelines for how to estimate measure costs and benefits. The estimation should address all substantial elements (or element groups) of costs or benefits that result from the delivery of the measure.

2.1. Measure Specification

The Roadmap describes the information needed to specify a measure and its specific applications. This includes identification of the baseline as either current practice or pre-condition; characteristics resulting in a difference in energy consumption between the baseline and efficient conditions; the measure application's delivery mechanism(s); and any other requirements, such as geographic region, market sector, building type, or business type. The estimate of costs and benefits must be consistent with the baseline and other characteristics of each measure application.

2.2. Element Identification

The Summary worksheet found in the Measure Assessment Template is used to describe the approaches used to estimate costs and benefits. The summary describes how each substantial element of cost or benefit is treated. The elements referenced may include those defined in section 4. However, the elements presented within these guidelines are not exclusive and additional elements may be included. The RTF may accept these additional elements if incremental costs and benefits estimates can be provided consistent with the guidelines described within this document.

2.2.1. Include Substantial Elements

All elements or groups of elements that account for a substantial portion of the either the measure's costs or benefits (for any measure application) should be included in the estimate of costs and benefits. A portion of costs or benefits is substantial if the RTF determines that it is likely to account for more than 10% of the measure costs or benefits. In many cases, data sources are limited and reflect costs and/or benefits of several elements combined, for example, contractor interviews or project invoices may be used to determine full capital costs. Such groups of elements should be included if they represent a substantial portion of a measure's costs and benefits.

2.2.2. Exclude Inapplicable or Insubstantial Elements

Although all substantial measure cost and benefit elements should be considered, most often, the majority of the possible elements defined in section 4 will be excluded from the analysis of a particular measure. Reasons for exclusion include the following:

- The element is not applicable to the measure. For example, design and engineering costs or benefits would not be applicable to white goods measures (e.g., residential appliances).
- There is no incremental impact of a particular element. For example, expected ongoing maintenance labor costs for the baseline and measure are identical.
- The expected costs or benefits are not substantial, i.e., less than 10 % of total costs or benefits. For example, a natural gas measure having a spark ignition rather than a pilot light would not require consideration of the costs and benefits of the electricity consumed by the spark igniter.

2.2.3. Limit Costs and Benefits to Efficiency Features

The costs and benefits of additional features that may be delivered with a measure, such as aesthetic upgrades or convenience factors, should be excluded from the incremental costs and benefits. For example, the incremental cost between baseline efficiency appliances with enamel finishes and high-efficiency appliances with stainless-steel finishes should not include the incremental cost of the finish. In another example, the RTF may prefer to estimate single-family window retrofit costs based on program-collected cost data from multifamily window retrofit measures, rather than cost data from single-family window retrofit measures, because single-family installations are likely to include costs of aesthetic upgrades.

2.3. Data Sources

The following presents guidelines for how to identify and evaluate data sources to be used in estimating measure costs and benefits.

2.3.1. Standard Information Workbook

For elements of costs and benefits addressed in the *Standard Information Workbook,* it is used as the generic source of data and analysis methods. However, where data or methods that are more appropriate to the measure are available, they should be used instead of those specified in the Standard Information Workbook. For example, if the labor rates specified in the Standard Information Workbook do not reflect the type of expertise, union affiliation, or other characteristics of a specific measure, more appropriate labor rates should be developed.

2.3.2. Other Data Sources

Where more measure-specific data are available, and for elements not addressed in the Standard Information Workbook, appropriate data sources must be identified and used.

The relative and absolute appropriateness of various data sources can be evaluated by considering several preferred data characteristics. Evaluating data sources in terms of these characteristics can help to prioritize data sources for a specific analysis. The following are preferred data characteristics, which apply to both primary research and secondary data sources:

- **Consistent reference for incremental values** – For measures with a current practice baseline, baseline and measure data that come from comparable sources (ideally the same source) allow for a consistent comparison of the incremental values. Depending on the baseline cost data source, adjustments may need to be made in the analysis to avoid including costs unrelated and unnecessary to the measure.
- **Sufficient granularity** – Datasets should be sufficiently detailed to understand any nuances that relate to the measure and to provide sufficient basis for analysis assumptions and adjustments.
  - Greater amounts of detail/granularity in data allow for greater variety and depth of analysis (for example, targeting a specific measure).
  - This may also serve to provide an increased level of confidence in the analysis.
- **Comprehensiveness** – Where multiple costs and benefits elements must be considered, some data sources may provide more comprehensive costs and benefits data than others.
  - For example, project invoices would provide data on the full capital cost of a measure, whereas distributor interviews would only provide data on the capital material cost.
  - If sales tax is included in values, it should be controlled for, as described in Section 4.1.
- **Reliability** – Data sources should accurately reflect the measure. Data sources that are not derived from the actual prices paid by customers may be biased. Secondary data sources may contain errors, and should be vetted where possible.
- **Representativeness –** Data sources should reflect the variation and relative magnitudes of likely customer purchase, for example by model, by region, by type of store, and by delivery channel. Data should also be relatively recent.
  - Adjustments to the dataset may be needed to reflect differences in delivery mechanism, region, or other specifications. For example, an upstream program may track wholesale costs, whereas a downstream program would track retail costs. Data should be sufficiently detailed to apply the most appropriate adjustments or make inferences.
- **Reasonable cost to obtain –** The resources required to obtain data should be reasonable relative to the significance of the elements being evaluated.
  - For example, the expense of conducting in-store retail research to determine the cost of a relatively inexpensive ongoing maintenance-materials item would not be warranted.

2.3.3. Use of Avoided Costs

While avoided costs most accurately represent societal measure costs and benefits, they are typically only available for electricity and natural gas. Therefore, if reliable marginal costs are not available, retail costs of other fuels should be used in estimating the costs and benefits of other fuel impacts.

2.4. Analysis

Analysis conducted for costs and benefits must be appropriate, replicable, and reflective of the expected measure costs and benefits through the sunset date.

2.4.1. Factors Affecting Costs and Benefits

Data sources should be as representative, thorough, and trustworthy as is feasible within the time and budget constraints of the analysis. In most cases, secondary data sources (i.e., data previously collected for other purposes) will be used. Data sources used to estimate costs and benefits should reflect—or be adjusted to reflect—the measure. Considerations may include, but are not limited to, the following factors that affect costs and benefits:

- **Measure specifications**. For example, cost data collected from a weatherization program not requiring air sealing would need an adjustment prior to use for a weatherization measure requiring air sealing.
- **Program delivery mechanism**. For example, cost data from a direct-install program may not directly apply to a retail program.
- **Access channels**. For example, if an efficient technology is mostly purchased at small drug-stores, an online survey of prices may not reflect the true costs to participants.
- **Location**. For a Northwest regional measure, data used from other regions outside of the Northwest should be adjusted as necessary to reflect costs and benefits expected in the Northwest.
- **Measure time period**. Data used from other times should be adjusted to reflect the expected costs and benefits through the sunset date. Typically, current costs and benefits are used as a proxy for this.

Section 6 describes some approaches that may be used to adjust data so that they better reflect the measure. When reliable data sources are unavailable for a measure element that is expected to be significant, professional judgment may be used.

2.4.2. Avoid Conflicts with Savings Estimation

Some factors may be significant for savings (see Savings Guidelines) and for costs and benefits. Estimating the effect on savings has precedence.

2.4.3. Appropriate Analysis Methods

Analysis methods (e.g., average value, linear regression, and lower quartile) should reflect the realities of the data and their representativeness of the specified measure. For example:

- A relatively small or scattered set of data would not warrant a regression analysis.

- Shipments or sales data may need to be factored into averages (i.e., weighted average) to reflect true market average costs where unit cost range is significant.
- The average value from all contractor bids may overestimate a measure's costs since actual costs may tend toward the lower bids. Using the first quartile value may provide a more appropriate estimate of actual measure costs.
- The median value may better represent the actual measure cost in a small dataset with significant difference between the average and median values.

2.4.4. Consistent and Clearly Stated Base Year

All costs and benefits estimates must be provided in a consistent and clearly stated base year.

3. Documentation Standards

The following presents the requirements for documenting measure costs and benefits analyses and estimates. The Costs and Benefits section of the Summary worksheet should be completed. On other worksheets, any information required to complete the analysis must be documented. Each step of the analysis must be clearly labeled and include adequate information for the analysis to be replicated by a third party using the same sources and analysis methods. The measure costs and benefits analysis and datasets should be included in the measure assessment workbook or a separate file of a suitable format.

3.1. Analysis

The costs and benefits analysis of each element or group of elements should be transparent and made readily available.

The analyses should clearly state the following:

- The elements being estimated.
- The sources of data used.
- The datasets.
- The base year for all costs and benefits estimates.
- Any modifications to the original datasets (for example, the removal of duplicates, incomplete records, or outliers).
- The analysis method (see section 6) for analysis guidance.
- The results (element estimates).

The analysis should be contained within a workbook based on the Measure Assessment Template, if feasible.

3.2. Datasets

If a dataset is used to develop estimates, the cleaned dataset should be provided in a workbook based on the Measure Assessment Template, along with the data source(s) and a clear description of the data-cleaning process. The data source(s) and cleaning process details should be clear enough for a reviewer to reproduce the cleaned dataset from the original dataset.

If the datasets are not amenable to inclusion in the workbook (e.g., if the datasets are too large or if the analysis was done in a programming platform other than Excel®), then the datasets should be provided in an accessible file format and made available on the RTF website. A reference to where data can be accessed (e.g., a website not managed by the RTF) is not sufficient.

Proprietary data sets can be used, even if they cannot be made available on the RTF website. In this situation, the RTF must decide if the information provided by these data outweighs the inability to publically post a completely transparent analysis. RTF staff should review datasets in advance of measure proposal and the Measure Assessment Workbook should contain as much of the data as permissible.

- For example, data might be provided with associated firm or individual names removed.
- In another example, the result (e.g., mean value and standard deviation) may be provided if the full data set cannot be provided.

3.3. Costs and Benefits Summary

See the Measure Assessment Template for an example of the Summary worksheet. The costs and benefits section of the summary worksheet must be completed. This section documents the following information:

- One or more sets of measure identifiers describing applications of the measure, e.g., single family homes west of the cascades. More than one set of identifiers will be required if costs and benefits vary substantially between different applications of a measure.
- The category of costs or benefits as defined in section 4.
- Description of the analysis approach used in estimating that category of costs or benefits.
- Description of the data source(s) that support the analysis approach.
- For each data source, the source type, picked from the list of types described in section 5.
- Year of the dollars estimated by each analysis approach.
- The analyst's estimate of the uncertainty of the estimate, expressed as a percent range (+/-%) around the cost or benefit estimates for each analysis approach. Can also be a description such as high, medium or low uncertainty. If costs and benefits are taken directly from a representative sample of measure deliveries the range should be the confidence interval for the sample and the confidence level should be noted.

3.4. Measure Costs and Benefits Checklist

See the Measure Assessment Template for an example of the Checklist worksheet. The guidelines checklist must be completed. The checklist is used to confirm that the requirements stated in this costs and benefits guidelines document are fulfilled.

4. Definition of Cost and Benefit Elements

This section defines the measure costs and benefits elements. It includes the most commonly considered elements; however, it is possible that others exist.

4.1. Capital Costs

Capital costs are the costs incurred in the acquisition and installation of an energy-efficient measure. The following are capital costs elements:

- Material
- Ancillary material
- Disposal
- Labor
- Design and engineering
- Permitting/licensing
- Markups
- Delivery

Sales tax is not considered in RTF measure costs and benefits analyses because the total resource cost test, which the RTF uses to determine cost-effectiveness, is not affected by this type of transfer within a society.

4.1.1. Material

Material refers to the primary equipment installed for the baseline and measure. Some examples are the difference in cost between light-emitting diodes (LEDs) and incandescent light bulbs and between a high-efficiency clothes washer and a standard efficiency washer.

4.1.2. Ancillary Material

Ancillary material refers to the components and consumables required to complete the proper installation of equipment and systems. Ancillary material does not include the material cost of the actual measure or the labor costs associated with the installation. Ancillary materials are associated with the installation and not the ongoing maintenance, which is covered under separate measure elements.

Examples of ancillary materials include wiring, exhaust/flue piping, pipe solder, fasteners (e.g., nails, screws), adhesives or sealants (e.g., glue, caulk, and spray foam), equipment mounting materials (e.g., footings, anchors, and concrete), or cleaning supplies that are used exclusively on the measure.

4.1.3. Disposal

Disposal refers to the removal, hauling, and discarding or recycling of existing equipment or by-products of new equipment installations. Disposal includes decommissioning associated with the existing equipment disposal process.

Disposal excludes any disposal associated with the ongoing maintenance of the measure, which is instead covered under the measure element Consumables Disposal (Section 4.4.3. Disposal also excludes demolition or any other labor associated with activity required to complete the measure installation (e.g., modifying a structure to accommodate the new installation).

When considering disposal options, the least expensive option should be assumed unless the measure includes the specification of a particular disposal/recycle process, in which case the cost of that process should be used. For example, recycling costs should be counted if recycling is required as a measure specification.

Disposal costs should not differentiate between different disposal times for baseline and measure activities. For example, for Pre-Condition baseline measures, if disposal would occur at a later date in the baseline case (at the end of useful life) that disposal cost should be subtracted from the disposal cost for the measure case. No value should be given to delaying the disposal cost in the baseline case. Therefore, disposal impacts should only be addressed in cases where implementation of a measure results in a change in disposal requirements—either in quantity or method of disposal. For example, an appliance recycling measure may incur recycling costs that exceed the standard disposal costs for baseline appliances.

4.1.4. Labor

Labor refers to the direct effort associated with the *installation* of equipment and systems or modification of operational practices. Labor is typically estimated in units of time (for example, person-hours), to which standard labor rates are applied.

Labor includes installations that are conducted by hired contractors or the measure recipient's own staff/owner:

- **Hired contractors –** Labor hours include those hours for the contractors associated with the installation (i.e., the same hours that would appear on a customer invoice).
- **End User staff –** The hours worked and hourly rates of the staff involved with the installation should be counted. Labor costs are typically arrived at by multiplying labor units of time by a fully loaded (i.e., including taxes, paid benefits, and organizational overhead) wage rate that is appropriate for the work performed.

Labor associated with ongoing maintenance is addressed in Section 4.2.1.

4.1.5. Design and Engineering

Design and engineering refers to the design and specification of project details. Design and engineering does not include the labor described in Section 4.1.4. Design and engineering typically includes the costs associated with selecting and sizing appropriate equipment, developing the work plans to guide installation staff, developing and modifying designs to optimize performance, and analysis to verify performance. Design and engineering costs are typically incurred from specifying engineers; mechanical, electrical, and plumbing firms; or the equipment supplier. These costs are usually incurred for complex and custom measures where measure design and specification depends on the characteristics of specific sites.

4.1.6. Permitting and Licensing

Permitting and licensing refers to the fees paid to a code enforcement agency in order to install equipment. This element is highly variable across jurisdictions. If the impact of this element is substantial, a broad data collection effort may be necessary to capture the variability in permitting and licensing fees across the region.

4.1.7. Markups

Markups refer to the additions to the cost of an item applied at various transaction points in the product delivery stream from the manufacturer to the end user. Where material and ancillary material cost data are not collected from the perspective of the end user (e.g., wholesale distributor pricing and retail pricing for contractor-installed products), markups may need to be applied to costs to reflect the true cost to the end user.

Markups may also include overhead associated with using in-house staff for labor (see Section 4.1.4). Sources such as invoices or work orders that detail in-house labor rates do not generally include organizational and administrative overhead, which may be significant and incremental for some measures. The analysis should provide details showing that this markup is not double-counted in the labor element.

4.1.8. Delivery

Delivery refers to costs incurred delivering material and ancillary material to the site of installation (e.g., shipping costs, or delivery fees from a retailer).

4.2. Maintenance

Maintenance is broadly defined as the "performance of routine, preventive, predictive, scheduled and unscheduled actions aimed at preventing equipment failure or decline with the goal of increasing efficiency, reliability, and safety." [1] The maintenance elements are as follows:

- Ongoing maintenance – labor
- Ongoing maintenance – materials
- Ongoing maintenance – disposal

Maintenance requirements between baseline and measure equipment are typically the same, resulting in no incremental costs or benefits. These maintenance elements should only be considered for cases where the costs and benefits differ between the baseline and measure.

4.2.1. Ongoing Maintenance – Labor

Ongoing maintenance labor is incurred by maintenance staff or contracted maintenance providers at sites where the measure has been installed. Ongoing maintenance labor may include the labor associated with general equipment upkeep, cleaning, minor parts replacement (e.g., filter replacements), and preventive care (e.g., bearing lubrication) that are required to maintain measure equipment at its designed operating capacity.

4.2.2. Ongoing Maintenance – Material

Ongoing maintenance materials describes costs for consumable products that are necessary for normal measure operations, such as materials required for general equipment upkeep, cleaning, and preventive care (e.g., bearing lubricant) that are required to maintain measure equipment at its designed operating capacity.

4.2.3. Ongoing Maintenance – Disposal

Ongoing maintenance disposal describes costs incurred for the disposal of consumables that are necessary for normal measure operations. Ongoing maintenance disposal may also include the disposal of materials associated with general equipment upkeep, cleaning, and preventive care (e.g., used bearing lubricant) that are required to maintain measure equipment at its designed operating capacity.

4.3. Operations – Fuel

For the purposes of these guidelines, "operations" are defined broadly and include energy consumed as the result of normal use of the equipment addressed by the measure.

[1] U.S. Department of Energy, http://www1.eere.energy.gov/femp/pdfs/omguide_complete.pdf

Avoided costs are used for electricity and natural gas because they are readily available in the ProCost model[2]; for all other fuels retail costs are used as a proxy for avoided costs.

4.3.1. Electricity (Avoided cost from ProCost)

Electricity avoided energy costs are developed using ProCost.

4.3.2. Natural Gas (Avoided cost from ProCost)

Natural gas avoided energy costs are developed using ProCost.

4.3.3. Propane, Heating Oil, and Wood

Impacts on propane, heating oil, and wood consumption are captured by this element.

4.3.4. Other Fuel

The measure costs and benefits of any fuels not explicitly listed as measure elements are captured here.

4.4. Operations – Non-Energy Consumables

Operations – Non-Energy Consumables refers to all costs incurred from the consumption and disposal of materials as part of the operation of the measure equipment. The Operations – Non-Energy Consumables elements are:

- Water
- Consumable Materials
- Consumable Disposal

4.4.1. Water

Water refers to the retail costs of water consumed by the end user. This measure element should not be used to capture the avoided costs associated with the supply (i.e., non-retail costs) of water.

The cost per gallon associated with water should account for both freshwater supply and wastewater disposal. For residential and commercial applications, wastewater is typically included in the per-gallon water cost. For agriculture and industrial applications, separate (fresh) water and wastewater impacts may need to be calculated as these may not necessarily be coupled.

---

[2] Natural gas and electrical energy consumption impacts are not included in the scope of these guidelines. See Section 1 for more details.

In contrast to other elements in the Operations – Non-Energy Consumables category, the embedded energy of water consumption may be included in the energy analysis.[3] If the embedded energy implications are included in the energy analysis, then the cost of that energy should be netted out of the water and wastewater costs. This will eliminate double counting of the value of that embedded energy (once as the value of energy savings in the energy analysis and a second time a part of the total water cost).

4.4.2. Consumable Materials

Consumable materials refer to materials that are used in the ongoing operation of a measure. One example of this would be laundry detergent: the difference in detergent consumption (type or quantity) between types of washing machines (e.g., top vs. front loaded).

4.4.3. Consumables Disposal

Consumables disposal, which refers to the cost of disposing of spent materials, is similar to the disposal element described in Section 4.1.3. There may be ambiguity between disposal costs in this operations category and the maintenance category of measure elements. In these cases, to avoid double counting, one category must be selected to apply these costs. This decision should be documented clearly in the Costs and Benefits section of the Summary worksheet.

4.5. Other Non-Energy Impacts

Other non-energy impacts are defined as any effects, positive or negative, that result from a measure that are not captured through the energy analysis or the categories of capital costs, maintenance, and operations. Examples of other non-energy impact measure cost elements may include the following:

- Building Owner
  - Building value
  - Rent premiums
- Building Operator
  - Equipment downtimes
  - Renewable energy credits
  - Resale of on-site generation
- Business
  - Marketing and public relations

[3] This unique treatment of embedded energy from water is justified because it is the one consumable that is almost entirely processed within the region; the vast majority of embedded energy in water consumed in the region comes from regional sources.

  - Productivity
  - Absenteeism
  - Attracting and retaining top tenants and employees
- Building Occupants
  - Occupant comfort
  - Occupant illness
  - Indoor environmental quality
- Utility
  - Reduced customer calls, shutoffs, and reconnections for delinquency
  - Reduced cost collection activities
  - Reduced arrearages and carrying costs for arrearages
- Societal
  - Income generated from measure installation
  - Avoided costs for unemployment benefits
  - Reduced heat island effect

Conclusive estimates of the monetary impact of these elements do not generally exist. Other non-energy impacts may be included in a measure cost analysis if it can be sufficiently demonstrated to the RTF that the impacts are significant and monetizable.

5. Common Data Sources

This section discusses common data sources used for estimating costs and benefits. Additional data sources not described in this appendix may also be appropriate.

5.1. Program Tracking Data

Energy efficiency program tracking databases maintained by utilities or other incentive providers often contain measure cost data collected from program participants. They may also include factors contributing to cost, such as type or size of installation. Depending on the measure, data may be limited to the material cost, or may indicate the full measure capital cost (i.e., including installation and disposal costs).

Strengths of this data source include the following:

- Provides local/regional retail prices paid by customers for a specific measure
- Sales volume information reflects the range and distribution of projects by cost and benefit influencing factors (e.g., product type, installation type).
- Program data could be used to inform the measure specification (e.g., efficiency level, typical installation type).

Weaknesses of this data source include the following:

- Current practice baseline costs may not included.
- Program requirements or the existence of the program may influence the costs.

Some additional considerations include the following:

- Program data will be reflective of a particular type of delivery mechanism (e.g., downstream rebate) and region.
- Data content, quality, and availability will vary by program implementer.

5.2. In-Store Retail

In-store retail refers to material and ancillary material cost data collected in retail stores.

Datasets typically include information on costs, equipment types and size, efficiency levels, and non-energy-related feature variations.

Strengths of this data source include the following:

- Provides local/regional retail prices paid by customers
- Provides consistent data source for both measure and baseline equipment

Weaknesses of this data source include the following:

- Data may not indicate sales volume.

- Installation costs are not included.

Some additional considerations include the following:

- Using a representative sample of data is particularly important when using in-store retail data. Analyses should consider data from both urban and rural settings in order to fully characterize the market. Analyses should also include data from a wide range of store types including grocery, big-box hardware, specialty hardware, and drugstores, as appropriate.
- Additional information can be gathered through interviews with store staff. For example, staff can estimate relative sales volumes of various products.
- Typically, this is the most expensive way to collect primary data due to travel and labor requirements.

5.3. Contractor and Project Invoices

Contractor invoices are a documented source of actual costs to end users. Invoices often include useful information on the equipment installed such as model number and manufacturer that facilitate market characterization in addition to incremental cost development. Typically, invoices are available for aggregated capital costs. Invoices may also be available for maintenance service and upkeep of installed measures.

Strengths of this data source include the following:

- Provides local/regional retail prices paid by customers
- Invoices typically allow for a bundled analysis of the entire capital cost category of elements.

Weaknesses of this data source include the following:

- Invoices for current practice baseline installations may not be available.

5.4. Contractor Price Sheets

Contractor price sheets summarize expected costs, and often include both baseline and measure products. They are used by contractors to develop quotes for projects.

Strengths of this data source include the following:

- Baseline and measure equipment can be compared from the same source.

Weaknesses of this data source include the following:

- While the incremental costs between standard and efficient equipment are reliable, the prices listed in the price sheet may not represent the actual costs paid by customers. This is problematic for pre-condition baseline measures, where the full measure cost *is* the incremental cost.

5.5. Online Retail

Online retail refers to costs collected from online retail venues.

Strengths of this data source include the following:

- Reflects prices paid by customers
- Baseline and measure equipment can be compared from the same source.

Weaknesses of this data source include the following:

- Typically does not capture local or regional prices.
- Installation costs are not included.
- May not reflect actual costs for measures where the majority of purchases are made through channels other than the Internet.
- Limited to simple and common measure types.

Additional considerations of this data source include the following:

- Online retail surveys are typically the least expensive method for collecting data when secondary sources are not available.

5.6. Contractor Interviews

Cost information from contractors can be gathered through telephone interviews, or other surveying methods. Contractors generally have specialized knowledge and experience with specific measures.

Strengths of this data source include the following the following:

- Contractors typically have first-hand knowledge of the requirements for operating and maintaining the equipment that they install.
- Baseline and measure equipment can be compared from the same source.
- Interviews can be tailored to provide more nuanced information, details, or to support complex measures.

Weaknesses of this data source include the following:

- Potential biases in estimates because data are not provided as competitive bids.

Some additional considerations include the following:

- Contractor interviews can provide more nuanced information about factors that influence many of the costs and benefits of a measure. Some examples of information where contractors can be especially useful include costs for labor, markups, materials, and operations and maintenance.

- Contractors are also a useful source for calibrating data. Interviews can be structured in a way so that draft analyses (or portions of analyses) can be presented to contractors to gauge how closely the analyses results match their own work practices.

5.7. Distributor Interviews

Cost information from distributors can be gathered through telephone interviews, or other surveying methods. Distributor interviews can be used to obtain wholesale material and supporting material costs for large regions.

Strengths of this data source include the following:

- Baseline and measure equipment can be compared from the same source.
- Relative to contractors, distributors may have a broader knowledge of sales trends and equipment prices.
- Interviews can be tailored to provide more nuanced information/details, or to support complex measures.

Weaknesses of this data source include the following:

- Distributors have a limited view of end-user costs.

Some additional considerations include the following:

- Markups from the distributor to the contractor and ultimately to the customer are required supplements to information gathered.
- Distributors may not provide information readily as this can be sensitive and competitive information.

5.8. Market Actor Interviews

Interviews with professionals in industries relevant to the measure can be a source for costs and trends, as well as providing references to secondary cost research (e.g., industry surveys).

Strengths of this data source include the following:

- Information on current program and market conditions
- Interviews can be tailored to provide more nuanced information/details, or to support complex measures.

Weaknesses of this data source include the following:

- May not provide actual cost values

5.9. Maintenance Staff Interviews

Maintenance staff can be interviewed either on-site or over the phone to gather information on the maintenance and operation of installed measures. Maintenance staff may either be in-house staff or hired service contractors who maintain equipment and perform regularly scheduled upkeep. Information from maintenance staff interviews may include maintenance labor and material needs, operation material needs, and disposal requirements.

Strengths of this data source include the following:

- Baseline and measure equipment can often be compared from the same source (individual).

Weaknesses of this data source include the following:

- Only applicable to measures in buildings that have a maintenance staff.

Some additional considerations include the following:

- Maintenance staff is also a useful source for calibrating data. Interviews can be structured in a way so that draft analyses (or portions of analyses) can be presented to maintenance staff to gauge how closely the analyses results match their own work practices.

5.10. Professional Judgment

For basic measures, individual analysts may provide their own estimates for certain elements such as installation hours, delivery costs, and maintenance requirements.

Strengths of this data source include the following:

- Tailored analysis to specific needs
- Leverage analysis staffs' expertise

Weaknesses of this data source include the following:

- Prone to subjectivity and interpretation
- May not be supported by documented sources

Some additional considerations include:

- Professional judgment is an appropriate approach to representing elements with relatively minor costs or benefits. For more significant elements, professional judgment may be used when reliable data sources are unavailable, but will be subject to RTF review.

6. Analysis Methods

This section discusses analysis methods for determining measure costs and benefits. These methods should be considered while ensuring that results (1) represent the measure and (2) do not include costs or benefits unrelated and unnecessary to the measure.

6.1. Data Cleaning

Original datasets should be reviewed and cleaned before being analyzed. Cleaning steps may include the following:

- **Formatting** – Records should be formatted for consistency. Common formatting activities include converting numbers stored as text into numeric entries and standardizing the spelling and naming of categories.
- **Treatment of incomplete and duplicate records** – Incomplete and duplicate records should be identified and addressed.
- **Identification of outliers** – Outliers are observations that are abnormally far from other values in the dataset. Outliers may occur naturally in the dataset or due to error in data entry and/or conversion. Naturally occurring outliers can arise from atypical circumstances at the source of the record (e.g., extremely high measure installation costs at a hazardous site). Statistical tests can be used to identify true outliers in datasets.[4]
- **Treatment of outliers** – The decision of whether to include or remove outliers needs to be defensible and depends on the nature of the outlier and how it relates to measure implementation. The outliers should be removed if comparable values are not anticipated in measure implementation (e.g., program participants will likely avoid high-cost installations and opt for lower-cost alternatives); if the measure specifications are tightened to exclude likely causes of the outlying data; or if the outlier(s) cause the analysis method to result in unrepresentative values.

6.2. Estimation Approaches

There are various analytical methods for estimating measure costs and benefits element impacts from relevant datasets. Estimation methods include the following:

- Average (arithmetic mean)
- Weighted average
- Median
- Regression modeling

[4] Two examples are the Grubbs' Test (for testing for a single outlier) and the Tietjen-Moore Test (for testing for multiple outliers).

- Lower quartile
- Built-up cost estimates

6.2.1. Average (Arithmetic Mean)

The average (also referred to as the arithmetic mean) provides a single-point estimate to represent a set of values. It is most appropriate when the dataset represents a random sample, and an estimate of the expected value in the underlying population is desired. Provided the data are from a random sample, an unbiased estimate is produced.

Skewed distributions and outliers can influence this method. Despite this, the average can provide representative results so long as the sample is clearly defined and restrictive (e.g., specific model numbers, specific cost resource).

Examples where averages are appropriate include the following:

- Compact fluorescent lamp retail costs by wattage and package size
- Capital costs for duct sealing based on project costs in an incentive program tracking database

6.2.2. Weighted Average

Using a weighted average is most appropriate when the dataset is not from a random sample of the underlying population, and the expected value of the population is desired. Weights can be defined as market variables (e.g., market share of particular manufacturers with known markup differences, equipment size) or cost influential feature sets (e.g., distribution of stainless-steel models versus less expensive white models or distribution of efficiencies).

Like the average, the weighted average can be influenced by outliers and skewed distributions. Outlying data points and skewed distributions should be checked for when using the weighted average.

Examples where weighted averages are appropriate include the following:

- Residential measure material costs weighted by retail (e.g., big box store) or contractor. Residential installations may originate from either retail locations or contactors for a given measure in a given market. Weightings by source help ensure that the specific mix of supply streams is represented in the final results. This is especially important where one source may include premiums or other markups (i.e., contractor markups).
- Determining average regional costs of water or labor, when costs in several regions are known, as are the consumption levels or population in these regions.

6.2.3. Median

The median provides a representation of a population's central tendency, and is an appropriate analysis method when outlying values exist in the sample. Medians are not influenced by outliers because the median does not communicate asymmetry in the extremes of the sample distribution. Medians should primarily be used when there is a demonstrated concern that outliers in the dataset skew the average value significantly.

Examples where medians are appropriate include the following:

- Design and engineering costs for complex measures. Complex measures may cover a wide range of applications and include cost considerations that are site-specific. Therefore, a median can describe the central tendency of the cost of design and engineering associated with a given measure.
- Project cost estimates obtained from a survey of industry experts

6.2.4. Regression Modeling

For measure costs and benefits analysis, regression analysis is preferred to the arithmetic mean if there is interest not just in the overall mean, but in how the mean is conditioned by the observable characteristics.

Regression modeling can be used to disaggregate cost effects of energy from non-energy features and to determine the effect of observable characteristics on the mean value of the measure element cost.

Where data gaps exist, regression is appropriate for interpolation, but not for extrapolation.

Examples where regressions are appropriate include the following:

- Estimating the incremental cost of increased appliance efficiency, when appliances also contain a variety of non-energy features (e.g., dishwasher settings and finish material)
- Estimating the costs of a product in a range of sizes/capacities, where some size ranges are poorly represented in the dataset (e.g., estimating the cost of a high-efficiency chiller for each of several capacity ranges, where there are many data points for most size ranges but only a few data points for one of the intermediate size ranges)

6.2.5. Lower Quartile

The lower quartile method estimates measure costs by defining the range and selecting the cost at the lower quartile (i.e., the 25^{th} percentile), to reflect the bidding process where less expensive options are given priority. Bids from this method are typically normally distributed; any skewness should be fully described when using this method.

Costs that are competition-driven may cover a wide range, particularly for contractor bids where material and labor may be combined. In addition to capturing competitive pricing, this range can encompass differences in material selections (e.g., premium versus standard), installation methods, labor rates and hours (e.g., skilled/experienced contractors may estimate fewer hours), and warranties.

The lower quartile cost estimate method can be applied to solicited bid (e.g., from contractors) on a specified project.

6.2.6. Built-Up Cost Estimates

In some cases, it may not be possible to isolate the cost of a measure by directly comparing the costs of efficient and baseline technologies (i.e., if the same non-energy features are not offered for standard and high-efficiency equipment.) In these cases, a built-up cost approach may be used to estimate the incremental cost between baseline and efficient case specifications. This method would separate out any variations in price that result from differences in non-energy factors, making the true incremental cost accessible for measure analysis.

The built-up cost method is more complex than statistical methods. Estimates must be supported by other sources of information for inputs; they require expert understanding of technology and installation requirements, and they may require defensible assumptions on separation of energy- and non-energy- related features and installation scenarios.

Examples where built-up cost estimates are appropriate include the following:

- Lighting fixtures where the material cost includes decorative housings and trim (e.g., outdoor lamp fixtures). If aesthetic features are not constant between baseline and measure equipment, then the full equipment cost will reflect more than just the cost difference in lighting technology. Therefore, building up the cost by analyzing the individual components (in this example, lamps and ballasts) will allow the energy-related features to be isolated.
- Complex variable frequency drive (VFD) retrofit installations. Some installation costs may be site specific and include demolition, calibration, and testing specific to a singular installation. Costs that are necessary to support the measure installation may vary significantly and make it difficult to characterize measure incremental costs. Building up costs for the major installation components (e.g., material, labor, and design and engineering) will allow these features to be individually characterized so that the incremental costs specific to a desired installation scenario can be described.

.

GUIDELINES FOR THE ESTIMATION OF MEASURE LIFETIME

REGIONAL TECHNICAL FORUM

Release Date: December 8, 2015

TABLE OF CONTENTS

1. Introduction

The Regional Technical Forum (RTF) develops and maintains a series of documents that provide guidance on how to assess energy efficiency measures. This document (the Lifetime Guidelines) provides guidance on how to estimate the lifetime of measure savings.

1.1. Purpose

The purpose of these guidelines is to describe a systematic approach to developing estimates, and documenting approaches and data sources when estimating the lifetime of measure savings (referred to as measure lifetime).

1.2. Key Concepts

1.2.1. Roadmap Concepts

A number of key concepts such as "Measure" and "Savings" are defined in the *Roadmap for the Assessment of Energy Efficiency Measures* (the Roadmap). These Lifetime Guidelines assume the reader has a thorough understanding of the Roadmap. Additional concepts that specifically apply to these Guidelines are described below.

1.2.2. Measure Lifetime

Measure lifetime is defined as the median number of years during which at least half the deliveries of a measure are in place and operable, i.e., produce savings. Measure lifetime should not be confused with a measure's sunset date, which ends the period during which a measure's savings estimation method is RTF-approved. In addition, measure lifetime is a different concept than Remaining Useful Life.

1.2.3. Remaining Useful Life (RUL)

Remaining useful life (RUL) is applicable to measures with a pre-conditions baseline. RUL is the number of years that the system, equipment or practice that comprise the measure's pre-condition baseline would have persisted if the measure had not been delivered.

1.2.4. Balance of Measure Lifetime (BML)

Balance of measure lifetime (BML) equals the measure lifetime minus RUL.

1.2.5. Lifetime Factor

Many factors may affect measure lifetime, including but not limited to: delivery method, equipment sizing, maintenance practices, operating conditions and operating hours.

1.2.6. Substantial Lifetime Factor

All factors that account for a substantial portion of the measure's lifetime (for any measure application) should be included in the estimate of lifetime. A factor is substantial if the RTF determines that it is likely to increase or decrease the measure lifetime by more than 20%.

1.2.7. Lifetime Reference Table

This table contains lifetime values for various measures or groups of measures, and data sources and analysis approaches that support those values. The table contains the same information as is needed when completing the measure lifetime portion of the Summary sheet in the Measure Assessment Template (see section 3.3). The information in this table should be considered in developing and documenting lifetime estimates for a measure. This table can be found in the Standard Information Workbook (see section 1.4).

1.3. Development of RTF-Approved Lifetime

The best practical and reliable analysis methods and data sources should be used in estimating measure lifetime. Practical means that the required data collection and estimation can be carried out with proven techniques and resources deemed reasonable by the RTF. The analysis of measure lifetime must be unambiguous. Documentation of data sources, how they were used in the estimation, and an estimate of uncertainty[1] are necessary for the RTF to determine whether an estimate of measure lifetime is sufficiently reliable.

The RTF's role in the estimation of measure lifetime depends on which method is used to estimate measure savings:

- **UES Measures –** Require RTF approval of lifetime, which are inputs to the ProCost model.
- **Standard Protocol Measures –** Require RTF approval of the protocol for estimating lifetime. This may be either specific lifetime estimates (for measures where the savings are expected to vary significantly, but not the lifetime) or protocols for determining lifetime, such as a list of components to provide estimates for, data collection sources and methods, and analysis approaches.
- **Custom Protocol and Program Impact Evaluation Measures** – Does not require lifetime to be reviewed by the RTF. However, lifetime should be estimated and documented as described in these Guidelines. The RTF may review the research plan for a Program Impact Evaluation, including lifetime estimation methods, if requested.

[1] Although maximum levels of uncertainty are not set, it is necessary to provide the RTF with the information needed to gauge uncertainty, such as a complete description of data sources and analysis approaches.

1.4. Supporting Documents

The following supporting documents are referenced throughout these guidelines.

- **Measure Assessment Template –** The Excel® Measure Assessment Template contains the ProCost model, ProCost inputs and outputs, energy and costs and benefits analyses, and summary tables. It is used to fully document a measure. The template contains the Summary sheet, a portion of which is used to document the approaches, supporting data and other aspects of the analyses that yield estimates of measure lifetime. The template also contains a checklist that can be followed in estimating lifetime for a measure.
- **Standard Information Workbook –** This Excel® workbook provides RTF-approved generic values for various parameters needed in completing measure assessments. It contains the Lifetime Reference Table providing information on relevant data sources, analysis approaches and estimates of lifetime that can be used as a starting point in the analysis of a measure's lifetime.

2. Estimation of Lifetime

The following presents guidelines for how to estimate measure lifetime. The estimation should address all substantial factors that determine the lifetime of a delivered measure.

2.1. Measure Specification

The estimation of measure lifetime must be consistent with the measure specification (the Roadmap describes all aspects of the measure that are included in the specification). The specification defines measure identifiers, which determines the number of separate measure applications. Each measure application requires an estimate of lifetime. If needed, measure identifiers may be included which separate applications of a measure that have substantially different lifetimes. In addition, the measure specification defines the savings baseline, implementation standards, product standards and sunset date. All of these must be considered in estimating lifetime.

2.2. Data Sources

The estimation of measure lifetime may be based on one or more data sources. The Lifetime Reference Table, in the Standard Information Workbook, may be used as a starting point in estimating measure lifetime, if it contains data sources for the same or similar measures. If no specifically applicable data sources are found in this table, a search should be conducted for other sources of relevant data.

At a minimum, the search for relevant sources of data should include a review of documents and data available from ASHRAE and from the standard setting processes operated by the US Department of Energy. In addition to the Lifetime Reference Table, Section 5 provides references for other data sources that should be considered.

The following should be considered in determining whether a data source is used in estimating a measure lifetime:

- Are the methods for lifetime estimation and the supporting data collection sound and well documented?
- If the data is for another similar, but not identical measure, is there a practical method for applying it to the measure specification?

Data sources involving primary data collection should be assessed in accordance with the guidance provided in section 4.

If no relevant data sources are available, new data collection may be conducted, such as:

- interviews with equipment vendors or other trade allies
- review of manufacturer warranty or other product lifetime information

If no relevant sources are available and the RTF determines that new data collection is not feasible, professional judgment may be used as a data source, provided that the rationale is described and documented.

2.3. Analysis

Analysis conducted to estimate measure lifetime must be appropriate, replicable, and reflective of the expected measure lifetimes for measures delivered during the period defined by the measure sunset date.

2.3.1. Factors Affecting Lifetime

Many factors may have a substantial impact on measure lifetime. All substantial factors should be considered in the estimation. A factor is substantial if it would increase or decrease the measure lifetime by at least 20%.

Factors that may have a substantial impact measure lifetime include, but are not limited to, the following.

- **Program delivery method**. Measures directly installed may last longer than measures delivered via mail for self-install, because self-installers may be less skilled and may not install according to manufacturer expectations, such as appropriate placement.
- **Installation practices**. Does the installation adhere to equipment manufacturer requirements for the class of equipment and comply with the product warranties? Adjustments may be needed to lifetimes originally estimated if they assumed practices not consistent with likely installation practices.
- **Sizing and rating**. Is the equipment sized and rated for the likely operating schedules and duty cycles, and are these consistent with the manufacturer's recommendations and warranty? Over and under sizing the equipment can change the lifetime of the measure.
- **Maintenance**. Is maintenance performed in a fashion that is consistent with the manufacturer requirements or best practices for the equipment and its associated controls or measure components? Is maintenance likely to be performed over the life of the measure? Deferred maintenance can decrease the lifetime.
- **Delivery verification**. Is measure delivery verified, including equipment and controls testing per manufacturer requirements?
- **Region or climate zone**. Region or climate may affect measure lifetime in many ways. For example, differences in climate zones may lead to changes in loading on the affected equipment.
- **Operating hours**. Operating hours, determined by installation location, business type, or climate, might affect lifetime of certain measures, such as changes to lighting, HVAC equipment. This may not be a factor for other measures such as insulation.

- **Operating conditions and practices**. Adjustment to lifetime might be needed if operating conditions are "dirtier" than manufacturer recommendations or on/off switching occurs frequently.
- **Occupancy Changes**. Changes in occupancy, such as those caused by business turnover, may change lifetime. For example, measure lifetime estimated for all commercial applications may not be appropriate if the measure applies only to one sector, such as restaurants, where ownership and occupancy changes frequently.
- **Remodeling practices**. The lifetime should account for removal of the measure due to remodeling prior to its expected physical failure.

2.3.2. Avoid Conflicts with Savings Estimation

Some factors may be significant for savings (see Savings Guidelines) and for lifetime. Estimating the effect on savings has precedence. For example, installation rate may be a factor in estimating savings. If savings per unit is adjusted downward to reflect observed installation rates, it is not appropriate to also adjust lifetime to reflect the zero lifetime of the measures that were not installed.

2.3.3. Lifetime Estimation

Measure lifetime should be estimated as follows:

- **Select best data sources**. Assess the available data sources and select those that most closely conform to the measure specifications. Apply the techniques discussed in section 4 in assessing sources that are based on primary data collection efforts. For example, for a multi-family residential CFL measure delivered by direct mail or direct installation, the best available lifetime data source may be a retention study done in California for a single-family CFL measure delivered by direct installation. In addition, relevant data might be available for operating hours from Northwest studies of direct installation and direct mail delivery of CFLs to multi-family residences.
- **Identify substantial factors**. Using the insights gained from the review of available data sources and professional judgment; identify the factors from section 2.3.1 that substantially affect measure lifetime. For the CFL example, differences in occupancy change rates and operating hours might be the substantial factors.
- **Develop analysis approach**. Develop an analysis approach that best utilizes the selected data sources and accounts for the factors that substantially affect measure lifetime. For the CFL example, the analysis approach might involve adjustment for differences between multi-family and single family CFL operating hours, observed for each delivery method. However, it may not be practical, given lack of data, to include adjustment for occupancy change rates in the analysis.
- **Estimate lifetime, accounting for measure specification**. Use the analysis approach to estimate lifetime, accounting for differences between the measure as defined in the

selected data sources and the measure's specification. Develop lifetime estimates for each application of the measure as defined by the measure specification (see Roadmap). In the CFL example, the lifetime for the direct install multi-family CFL might only be adjusted for the difference in operating hours between single and multi-family residences, while the lifetime for direct mail would also be adjusted for the differences in operating hours observed for the delivery methods.

2.3.4. Remaining Useful Life Estimation

Remaining useful life (RUL) should be estimated for all measures that have a pre-conditions baseline. Consideration should be given to the system, equipment and practices that comprise the baseline condition affected by the measure. The question to be addressed is how long the system, equipment or practice would have persisted in its baseline state if the measure were not delivered. The RTF expects careful consideration of the RUL, but does not impose any specific quality standards on these estimates beyond the use of best available data and professional judgment.

The assumptions made and analysis performed to estimate RUL should be documented in the SummaryRUL sheet of the Measure Assessment Workbook for UES measures. For Standard Protocol measures, these assumption and analysis should be documented in the appropriate sections of the protocol. For Custom Protocol measures, the assumptions made and analysis performed should be documented in the site-specific savings report.

If the estimated RUL is longer than ten years, then the RTF will assume that RUL equals measure lifetime.

3. Documentation Standards

This section describes the documentation standards for measure lifetime analysis and data. To the extent practical, all documentation should be included in a workbook based on the Measure Assessment Template. This will include completing the measure lifetime portion of the Summary worksheet along with other sheets that contain the data and analyses performed to estimate lifetime for each measure application. The workbook should document how data sources were used to estimate the lifetime for the measure or for separate measure applications.

3.1. Analysis

Documentation of the analysis performed should clearly state the following:

- The sources of data used
- The datasets used
- Any modifications to the original datasets (for example, the removal of duplicates, incomplete records, or outliers)
- The analysis approach
- The results

The analysis should be contained within a workbook based on the Measure Assessment template, if practical. However, if analyses are performed using software other than Excel™, such as Proc Lifereg, lifetest, or similar SAS (or other) procedures, the source code used to perform the analysis should be included along with reference to the software and version.

3.2. Datasets

If a dataset is used in estimating measure lifetime, the cleaned dataset should be provided in the workbook based on the Measure Assessment Template along with the data source(s) and a clear description of any data-cleaning processes applied. The procedure used in dataset cleaning should be clear enough for a reviewer to reproduce the cleaned dataset from the original dataset.

If it is not practical to include the dataset(s) in the workbook (too large, or not compatible with Excel®) the datasets should be made available on the RTF website. A reference to where data can be accessed (e.g. a website not managed by RTF) is not sufficient.

3.3. Measure Lifetime Summary

See the Measure Assessment Template for an example of the Summary worksheet. The Measure Lifetime section of the summary worksheet must be completed. This section documents the following:

- One or more sets of measure identifiers describing applications of the measure, e.g., single family homes west of the cascades. More than one set of identifiers will be required if measure lifetime varies substantially between different applications of a measure.
- Description of the analysis approach used in estimating measure lifetime.
- Description of the data source(s) that support the analysis approach.
- The analyst's estimate of the uncertainty of the estimate, expressed as a percent range (+/-%) around the measure lifetime estimate. The uncertainty could also be characterized as high, medium or low. If lifetime is taken directly from a representative sample of measure deliveries the range should be the confidence interval for the sample and the confidence level should be noted.

3.4. Measures Lifetime Checklist

The lifetime portion of the checklist sheet in the Measure Assessment Template must be completed to confirm that the requirements stated in this lifetime guidelines are fulfilled.

4. Assessing Measure Lifetime Studies

This section describes characteristics of lifetime studies that should be considered when assessing their quality and applicability to the estimation of measure lifetime.

4.1. Program and Measures Studied

It is important to clearly understand the program and measures that were the subject of the study. This would include all elements of the measure specification as described in the Roadmap, such as delivery method, installation procedures, and building, system or customer characteristics that determined eligibility. Clarifying these elements provides information that may determine how the study's results are generalized to other programs and measure specifications.

4.2. Participant Characteristics

Lifetime studies will typically involve samples of program participants. The source for the participant data is usually the program's tracking database. Information from the program tracking database should include the number and specification of the delivered measures, the installation locations (city, state and climate zone), and equipment removed. Delivery verification data may also be available and is useful in assessing a lifetime study.

4.3. Sample Design and Procedures

The statistical reliability of the study will depend on its sample design and sampling procedures. Important aspects of the sample design and procedures include:

- Scope of the sample population, e.g., types of customers, period of measure delivery, measure applications.
- Definition of the sampling element, e.g., premise, measure, or affected equipment.
- Sample selection procedure, e.g., simple random, stratified random, multi-stage.
- Stratification criteria and the definition of strata boundaries.

The most important consideration is whether the study used a random sampling technique, which is required if the results are to be statistically reliable.

4.4. Survey Instruments

Survey instruments should be designed to collect several key pieces of data about the delivered measures. Key data includes (in decreasing order of importance):

- Whether the measure is still in place and operating

- Date of failure, replacement or removal, or questions that bracket the date of removal if specific dates aren't known.
- Reason for failure, replacement or removal
- Confirmation of measure delivery
- Operating conditions and usage
- Demographics or firm-o-graphics of the participants
- Other significant measure lifetime factors[2]

The study documentation should provide a copy of the survey instruments.

4.5. Survey Data Collection

The study should use data collection methods appropriate to the measure and participants – phone, on-site, web, other[3]. For example, data on single, identifiable measures may be collected accurately by phone (water heaters in a home). However, a facility manager may have difficulty in a phone survey describing which lighting fixtures were affected by the measure delivery. It may be useful to review other studies to determine whether the selected method is appropriate for the measure or customer group. The study should include pre-testing of survey instruments followed by appropriate revisions to the instruments and procedures. Study documentation should include response rates and the reasons for refusals for each sample stratum. The use of refusal surveys can help identify bias.

4.6. Data Preparation

The study should report efforts taken to investigate possible bias in the results. Any bias corrections should be justified, including treatment of outliers, missing data points, and any other data eliminated in the process of preparing the final analysis dataset.

4.7. Model Specification and Estimation

Most Lifetime studies use Proc Lifereg, lifetest, or similar SAS (or other) procedures for estimating the lifetime (50% of the measures in place and operable). The study should model multiple distributions, not just a default distribution. The model should test the need for introduction of exogenous variables (omitted factors) that may affect lifetimes in the model specification, for example indicators of operating hours, business types, and climate zone   The study should consider and address heterogeneity, errors in variables, influential data points and

---

[2] If behavioral measures are included, additional relevant questions might address frequency of behavior retention by one / all members of the household or business.

[3] In each case, best practices for the collection of survey data should be used. For phone, it may include variations on Dillman methods, pre-notifications, 5 calls before discarding sample, multiple rounds, and call-backs for non-response bias control. Quality data collection and survey training is important.

other modeling considerations. The model selection and all statistical results for the rejected and selected models should be included, discussed, and compared to other relevant studies and to lifetimes currently in use. Results should be presented with associated confidence intervals and standard errors.

5. Selected Sources for Data and Methods

5.1. Measure Lifetimes

- CADMAC database – searches on measure life studies in their searchable database will identify more than 50 lifetime studies. www.calmac.org
- ASHRAE Equipment Service Life Values - an online database providing averages / medians for lifetimes of measures removed, self-reported by contractors. The data can be filtered by geographic location. The figures are based on reports, not statistical Lifetime studies. ASHRAE Database of Equipment Service Life, referenced at "http://xp20.ASHRAE.org/publicdatabase/service_life.asp"
- California DEER Data Base for Energy Efficient Resources, at www.energy.ca.gov/DEER. Database of lifetimes for hundreds of measures, with source citations. (checked March 2012)
- California Public Utilities Commission, "D0111066 Energy Efficiency Policy Manual", http://docs.cpuc.ca.gov/published/Final_decision/11474-13.htm; uncertain date (checked March 2012)
- California Public Utilities Commission, "Appendix F, Effective Useful Life Values for Major Energy Efficiency Measures", www.calmac.org/events/APX_F.pdf · PDF file (checked March 2012)
- Efficiency Maine, 2010, "Technical Reference User Manual (TRM) No. 2010-1", August, available on web at "www.efficiencymaine.com/.../Maine-Commercial-TRM-8-31-2010-Final.pdf" · PDF file
- Efficiency Vermont, 2010, "Technical Reference User Manual (TRM)", February, available on web at www.veic.org/Libraries/Resource_Library_Documents/EVTTRM_VEIC, PDF file
- GDS Associates, 2007. "Measure Life Report - Residential and Commercial / Industrial Lighting and HVAC Measures", Prepared for the New England State Program Working Group (SPWG), Manchester NH.
- Goldberg, Miriam, J. Ryan Barry, Brian Dunn, Mary Ackley, Jeremiah Robinson, and Darcy Deangelo-Woolsey, 2009, "Business Programs: Measure Life Study", prepared for State of Wisconsin Public Service Commission of Wisconsin Focus on Energy Evaluation, Madison, WI, August.
- MassSave, 2010, "Massachusetts Technical Reference Manual", October, www. ma-eeac.org/docs/MA%20TRM_2011%20PLAN%20VERSION.PDF · PDF file
- New Jersey Board of Public Utilities, 2009, "New Jersey Clean Energy Program Protocols to Measure Resource Savings", December, See Appendix A (dated 2001), available on web at www.njcleanenergy.com/files/file/Library/Protocols%20Final%2012-7", PDF file

- New York Evaluation Advisory Contractor Team, 2009. "New York Standard Approach for Estimating Energy Savings from Energy Efficiency Programs, Single Family Residential Measures", Prepared for New York Department of Public Service, Albany, NY, December. Note: similar documents for other sectors.
- Skumatz, Lisa A. Ph.D;, and John Gardner, 2005. "Revised / Updated EULs Based on Retention and Persistence Studies Results", Prepared for Southern California Edison.
- Skumatz, Lisa A., Ph.D., M. Sami Khawaja, Ph.D., and Jane Colby, 2010, Lessons Learned and Next Steps in Energy Efficiency Measurement and Attribution: Energy Savings, Net to Gross, Non-Energy Benefits, and Persistence of Energy Efficiency Behavior, prepared for Center for Energy and Environment, Berkeley, California.
- Skumatz, Lisa A., Ph.D. and Allen Lee, 2004. "Attachment G - Assessment of Technical Degradation Factor (TDF) Study", prepared as Attachment to "Review of Retention and Persistence Studies for California Public Utilities Commission (CPUC)", San Francisco, CA.
- Skumatz, Lisa A. Ph.D., Rose Woods, and Scott Dimetrosky, 2004. "Review of Retention and Persistence Studies for the California Public Utilities Commission (CPUC), San Francisco, CA.
- Vermont Energy Investment Corporation (VEIC), 2011. "Mid-Atlantic Technical Reference Manual, Version 2.0", Managed by NEEP, July available on web at "neep.org/uploads/EMV%20Forum/EMV%20Products/A5_Mid_Atlantic_TRM_V2"... · PDF file
- Vermont Energy Efficiency Corporation (VEIC), 2010. "State of Ohio Energy Efficiency Technical Reference Manual", August, available on web at "amppartners.org/pdf/TRM_Appendix_E_2011.pdf" · PDF file

5.2. Factors Affecting Lifetime

5.2.1. Hours of Operation

- Gaffney, Kathleen, Miriam Goldberg, Paulo Tanimoto, and Alissa Johnson, 2010, "I Know what you Lit Last Summer: Results from California's Residential Lighting Metering Study", Proceedings from the ACEEE Summer Study on Energy Efficiency in Buildings, Asilomar, CA, August.
- KEMA, CFL Metering Study. Prepared for California's Investor-Owned Utilities (PG&E, SCE, SDG&E and SoCalGas), February 2005
- Williams, Alison, Barbara Atkinson, Karina Garbesi PhD, Erik Page PE, and Francis Rubinstein FIES 2012,"Lighting Controls in Commercial Buildings", Leukos, Volume 2 number 3, January 2012.

5.2.2. Turnover by Business Type

- Hickman, Curtis and Lisa A. Skumatz, 1994, "Effective ECM and Equipment Lifetimes in Commercial Buildings: Calculation and Analysis", Proceedings of the American Council for an Energy Efficient Economy (ACEEE) Summer Study on Buildings, Asilomar, CA, August.

5.2.3. Region and Climate

- Conduct comparisons of measure lifetimes by region on ASHRAE database. Source: ASHRAE Equipment Service Life Values - an online database providing averages / medians for lifetimes of measures removed, self-reported by contractors. The data can be filtered by geographic location. The figures are based on reports, not statistical Lifetime studies. ASHRAE Database of Equipment Service Life, referenced on web at "http://xp20.ASHRAE.org/publicdatabase/service_life.asp"
- Compare measure lifetimes from various state / regional lifetime tables.
- DEER provides values for multiple climate zones. California DEER Data Base for Energy Efficient Resources, at www.energy.ca.gov/DEER. Database of lifetimes for hundreds of measures, with source citations. (checked March 2012)

5.2.4. Delivery Method

- Nexus Market Research, 2008, "Residential Lighting Measure Life Study", submitted to New England Residential Lighting Program Sponsors, Cambridge, MA, June, and references therein.

5.2.5. Vintage and Early Replacement

- KEMA, 2008, "Summary of EUL-RUL Analysis for the April 2008 Update to DEER", from web, www.energy.ca.gov/DEER.
- New York State, 2011(?), "Appendix M – Guidelines for Early Replacement Conditions", available on web at "www3.dps.ny.gov/.../$FILE/Appendix%20M%20final%205-05-2011.pdf" · PDF file.
- Skumatz, Lisa A., 2011, "Remaining Useful Lifetimes and Persistence – Literature and Methods", Proceedings of the IEPEC Conference.
- Skumatz, Lisa A., Ph.D., M. Sami Khawaja, and Jane Colby, 2010. "Lessons Learned and Next Steps in Energy Efficiency Measurement and Attribution: Energy Savings, Net to Gross, Non-Energy Benefits, and Persistence on Energy Efficiency Behavior", report prepared for California Institute for Energy and Environment (CIEE), Berkeley, CA, January.
- Welch, Cory, and Brad Rogers, 2010, "Estimating the Remaining Useful Life of Residential Appliances", Proceedings from the ACEEE Summer Study on Energy Efficiency in Buildings, Asilomar, CA, August.

SUPPORTING DOCUMENTS

REGIONAL TECHNICAL FORUM

Release Date: December 8, 2015

TABLE OF CONTENTS

1. Measure Assessment Template

This Excel workbook is used for all classes of measures. It contains a Summary sheet which documents the measure specifications, estimation methods (savings, cost and lifetime) and the data sources that support these estimates. Example Summary sheets are included in the template for UES, standard protocol and custom measures. Also included is the Checklist sheet, which documents the major milestones that are relevant to the development and maintenance of measures in conformance with the Guidelines. In addition, this template includes ProCost and sheets needed to supply and document input and output values for this cost-effectiveness model.

Download Measure Assessment Template.

2. Standard Information Workbook

This Excel workbook provides RTF-approved default values for various parameters needed in completing measure assessments. For example, it provides hourly costs for various labor categories that can be used in estimating capital and O&M costs for a measure. It also contains the Lifetime Reference Table, which may be used as a starting point in estimating measure lifetime. It also contains commonly used input parameters for determining UES savings, including: weightings used for heating/cooling zones, house sizes, standard SEEM inputs, lighting hours of operation, and water-heater temperature set points.

Download Standard Information Workbook.

3. Marginal Cost and Load Shape (MC and Loadshape)

This workbook contains marginal energy costs and measure load shapes. These are used by ProCost in estimating measure cost-effectiveness.

Download MC and Loadshape.

4. Standard Protocol Template

This Word document provides a template, by example, for describing a standard protocol measure. A standard protocol must also have a completed Summary and Checklist sheets (in the Measure Assessment Template) and a calculator (as described in the Savings Guidelines). In addition, a research plan is required for provisional standard protocol measures.

Download Standard Protocol Template.

5. Research Plan Templates

5.1. Research Strategy

This Word document provides a template for documenting a research strategy. Such strategies are required for the approval of a Planning UES or Standard Protocol measure.

Download Research Strategy Template.

5.2. Research Plan

This Word document provides a template for documenting a research plan. Such plans are required for the approval of a Provisional UES or Standard Protocol measure.

Download Research Plan Template.