But I Thought a Statewide TRM Would Solve Everything?

Ryan J. Del Balso, Navigant Consulting, Inc., Verona, WI
Kevin Grabner, Navigant Consulting, Inc., Verona, WI

ABSTRACT

Developing a statewide TRM involves more than just identifying energy savings estimates for a variety of measures. There are many pitfalls along the way that do not involve the details of the measures themselves. Once the decision has been made to develop a statewide TRM, there are four main pitfalls which can derail the effective development and implementation of it: (1) lack of common understanding of the TRM's purpose, (2) not comprehending and planning for its effects on implementation and evaluation, (3) lack of planning for long-term upkeep of the TRM, and (4) underestimating its overall cost. The best way to avoid these pitfalls is to learn from others who are further along the path and hopefully not repeat the problems they encountered. This paper presents experiences from multiple states at various stages of TRM development, each with unique issues with which to deal. The information presented is pertinent to jurisdictions that have already made a decision to pursue a statewide TRM and those who are considering one.

Introduction

Several states have recently entered the path of developing a statewide TRM. Some are a few years into it, some are just beginning. Some may still be asking the question, "Should we develop a statewide[1] Technical Reference Manual (TRM)?" Regardless of their current position on the path, experience has shown there are often unexpected pitfalls along the way to developing a statewide TRM. If they are considering taking the plunge, or have already jumped into the deep end, the stakeholders[2] need to fully understand the ramifications of such a decision and carefully plan for its development and application. The best way to do this is to learn from others who farther along the journey. This paper identifies some lessons learned in order to help others as they contemplate or plan the development of their own statewide TRM.

What is a TRM?

A TRM, as related to energy efficiency programs or their equivalent, is a manual that specifies a standardized methodology for implementers to estimate and claim savings (energy, demand, fuel, water, greenhouse gases, etc.) for many common, mass marketed, energy efficiency measures. They are also sometimes used by evaluators as the yardstick against which the implementers will be judged. For

[1] Although this paper generally refers to "state(s)" as the applicable jurisdiction for a TRM, it is understood that some jurisdictions cross state/territorial/provincial boundaries, cover only a portion of a state/territory/province, or possibly apply to a single utility. For ease of terminology, the use of "state(s)" or "statewide" within this paper will generally include any territory a TRM applies to, whether it is a state or other jurisdictional boundary.

[2] Because each jurisdiction may use different terms to describe just about anything related to a TRM or its use, to avoid listing all of the possible terms throughout the paper a few definitions are in order. All public regulators and their staff will be referred to as "regulator(s)." All electric and gas utilities, electric distribution companies, and their implementation contractors will be referred to as "implementers" unless specifically discussing a particular jurisdiction. Evaluation contractors will be referred to as "evaluator(s)," and all other interested parties, including ratepayers, installation contractors, manufacturers, etc., will be referred to as "the public." Collectively, they all (regulators, implementers, evaluators, and the public) constitute the "stakeholders."

jurisdictions with multiple implementers offering the same measure, this ensures all parties are claiming savings for the measures in a similar manner, and sometimes using the same deemed savings estimate.

In jurisdictions without a TRM, it is typical for each implementer offering an energy efficiency program to claim measure savings using their own methodology and estimates. This commonly results in each program claiming a different savings for a given measure, even though there may be no indication of actual differences between the program offerings and measure savings. A TRM reduces this inconsistency by providing a representative average "deemed" savings value or standard "deemed" savings algorithm for each measure in the TRM to be used by all implementers.

TRMs usually include "fully deemed"[3] measures and "partially deemed"[4] measures, but rarely include protocols for custom measures. Different terminology may be used in various jurisdictions, but in general, "fully deemed" refers to measures for which a single average "deemed" savings value is provided in the TRM to be used no matter what the actual customer conditions are. No customer specific inputs are required to claim savings. For example, some TRMs provide a single deemed savings estimate for all recycled refrigerators regardless of size, location, age, configuration, etc. Fully deemed measure protocols work best for large scale mass market measures where there is strong empirical data to derive an estimate for average savings for the population, or measures with conditions of installation that rarely fluctuate significantly from a known value. The advantage of fully deemed measures is that they enable very cost effective implementation due to their simplicity. Their disadvantage is that if the population of participants is relatively small, the participating population measure consumption differs in some way from the assumptions, or there is limited reliable data to support the savings estimates, then the deemed savings estimates may be inaccurate and unreliable. They also generally are not reliable predictors of savings for an individual customer.

"Partially deemed" measures are somewhere between fully deemed and custom measures. For measures which vary significantly in their installation characteristics, the TRM may use a standard "deemed" savings algorithm rather than a fully deemed savings value. The protocol may include some "deemed" variables which must be used by implementers for all customers, and some variables which have default values for each jurisdiction or measure characteristic, but which may use customer specific inputs in place of the defaults if known. Some variables may not include a default value at all, but instead the implementer is required to obtain customer specific data. These measures are not appropriate to be fully deemed because the true savings fluctuate widely from customer to customer and a representative average is difficult to determine. Examples include commercial and residential HVAC measures, non-residential lighting projects, variable speed drives, etc.

Custom measures are on the other end of the spectrum from fully deemed measures. They are generally one of a kind measures for a given customer, and/or so complicated or rare that average savings estimates cannot be reliably derived. Examples of custom measures include modifications to a unique industrial process, a large chiller plant upgrade with multiple chillers and complicated control sequences which does not fit common TRM measure parameters, or installation of a newer technology that has not yet been evaluated as part of an energy efficiency program. Custom measures generally require project specific savings estimates to be derived once the project details are known. Custom measures are more complicated to implement than fully deemed or partially deemed measures, and therefore, many implementers prefer to include as many measures as possible in a TRM as fully or partially deemed.

Due to the unique needs of each jurisdiction implementing energy efficiency programs across North America, there is a diversity of approaches to naming, developing, using and maintaining each jurisdiction's equivalent of a TRM. What is considered a "measure savings protocol" in one TRM may be termed a

[3] Sometimes referred to as "deemed," "prescriptive," "stipulated," etc.
[4] Sometimes referred to as "semi-deemed," "quasi-deemed," "quasi-prescriptive," etc.

"measure substantiation document" in a different one, "unit energy savings" in another or a "measure estimation sheet" in yet another.

TRM's have been developed in many forms including stand-alone text documents, stand-alone spreadsheets, downloadable programs, web-based applications, and any combination of these. The most common format is a text document with or without supporting spreadsheets, however, several jurisdictions maintain databases of energy efficiency measure savings which contain similar content and serve a similar purpose as a TRM, but are not called a TRM. As an example, California has an extensive database of deemed measure savings titled the California Database for Energy Efficiency Resources (California Public Utility Commission, 2008), more commonly known as the DEER. This database has been developed overtime through significant research, metering studies, and evaluations. The Michigan Energy Measures Database (MEMD) (Morgan Marketing Partners, 2013) is another tool that is similar in purpose and function as a TRM, but which resides in a database rather than a text document. The Pacific Northwest Electric Power and Conservation Planning Council's Regional Technical Forum (RTF) uses multiple documents together which collectively serve a purpose similar to a TRM. The RTF has established four different savings estimation methods which can be used for energy efficiency measures, two of which when combined would be similar to a TRM, the "Unit Energy Savings" (UES) and "Standard Protocol" methods (Regional Technical Forum, 2012). The US Department of Energy funded Uniform Methods Project (UMP) (National Renewable Energy Laboratory (NREL) and The Cadmus Group, Inc, 2012) is an attempt to develop "a set of model protocols for determining energy and demand savings that result from specific energy-efficiency measures or programs." The vision is that the UMP protocols will serve as generally accepted industry standard framework that can be incorporated into a TRM and modified as needed.

Regardless of the terminology used, at a minimum they all include protocols to estimate savings for measures which are incented in energy efficiency programs. This typically includes residential and non-residential electric energy efficiency measures which are incented in a prescriptive manner. Some TRMs also include gas and other fuel energy efficiency measures, and even custom measure savings protocols. The type of measures (electric, gas, other) included is generally based on the needs and scope of the applicable efficiency portfolio standards (EEPS[5]) or equivalent legislation.

Whatever the format, the content within each measure protocol includes at a minimum, the methodology for estimating energy savings, whether it is from electricity, fuel, or both. This may be in the form of a fully or partially deemed savings estimate, and in a few cases custom measure protocols. The protocol may include a methodology for estimating electric peak-demand savings, water savings, measure lifetimes, required/allowable incentive levels, incremental measure cost assumptions, total resource cost (TRC) estimates, and/or any other information the stakeholders establish as appropriate for their jurisdiction. Some TRMs include measurement and verification (M&V) requirements in addition to the savings estimates. The scope of the TRM measure protocols will be based on the needs of the stakeholders developing the TRM.

Due to the varied nature of their development and purposes, some TRMs are quite complete and thorough in their documentation. These standalone TRMs include common cross-cutting assumptions, the purpose of the TRM and its proper application within the TRM document itself, in addition to the measure protocols. Some jurisdictions maintain several documents which collectively serve as a TRM, with each document focused on a specific measure or providing specific guidance on the use or development of the measure protocols.

[5] "EEPS" will be used throughout this paper as a general reference to all legislation, regulation requirements, or utility decisions which result in the requirement for a given entity to develop portfolios of energy efficiency programs within the applicable jurisdiction in order to meet set energy and/or demand reduction compliance targets.

TRM Design – Choices and Tradeoffs

Once the decision has been made to develop a statewide TRM, there are four main pitfalls which can derail the effective development and implementation of it: (1) lack of understanding of the TRM's purpose, (2) not comprehending and planning for its effects on implementation and evaluation, (3) lack of planning for long-term upkeep of the TRM, and (4) underestimating its overall cost.

Purpose

States develop TRMs for various reasons; this is clear from the many different formats and guidelines for their use across North America. Without a clearly defined purpose, it is difficult to make many decisions about what to include in a TRM, how it will be used by implementers and evaluators, and how it will get updated. Defining the primary purpose of the TRM should be done in the early stages of development.

Stakeholders should not make the incorrect assumption, however, that the purpose of a TRM is to develop measure savings protocols. These are only an outcome of the process, not the purpose itself. The underlying purpose should be carefully thought through as early in the development process as possible, preferably prior to making a decision to develop a TRM at all.

The primary purposes of a TRM are typically twofold: to increase reliability of savings claims and to reduce risk. There are often several secondary purposes, but they are usually a subset of these two primary ones.

Two of the primary parties with the most at stake in the decision to pursue a TRM are the regulators and implementers, and their purposes, although similar, tend to stem from different needs. The following sections will look at some of what influences both regulators' and implementers' purposes for a TRM and how these relate to the TRM itself. After that we look at how the regulatory environment can affect how the TRM achieves its defined purpose.

Regulators' Purpose

Regulators are generally driven by a desire to increase reliability of savings claims to reduce ratepayer risk of paying for savings that were not achieved. Under this perspective, the purpose of the TRM is to increase standardization of measure savings claims within the jurisdiction, which is hoped will result in improved reliability of those claims. Theoretically, if the TRM is accurate, this will reduce ratepayer risk by ensuring ratepayers only pay for real savings achieved through measures incented by the program implementer.

Unfortunately, the accuracy of a TRM is not something that should be assumed at the outset. Due to the complexity of programs and measures, lack of reliable data, and the ease of human error, accuracy for all measures in a TRM is often difficult to achieve. The larger and more complex a TRM is, the more difficult it is to maintain accuracy for all measures. If the regulator strives for reliability in the TRM, they need to make accuracy a requirement, but to achieve this some tradeoffs may be needed. Accuracy generally comes with higher cost and more time. As a result, if the regulator requires a certain level of accuracy, one tradeoff may be that fewer measures get included in the TRM. This is because it costs more to develop and maintain measure protocols with a higher level of accuracy. Another tradeoff may be that it takes more time to develop or modify.

Accuracy is something that can be improved over time through an established process to develop and maintain the TRM. While all jurisdictions aim for accuracy, some jurisdictions have intentionally made accuracy a high priority. Understanding the risk associated with an unreliable TRM, they have established a very rigorous process to get measures approved for inclusion in the TRM. Requirements may include achieving minimum statistical requirements on savings estimates, requiring a metering approach be used or

calibrated computer simulation models be used to develop savings estimates, and/or a formal review process. A robust stakeholder review process, which includes sufficient time and budget, with multiple opportunities for protocol review, correction and additional review, is necessary to increase the chances of developing a reliable and more accurate TRM.

Secondarily for regulators, standardization also reduces the regulators' risk of perceived unfairness. This can occur when one implementer claims X savings for a given measure while another implementer claims Y savings for the same measure. This is often more a matter of avoiding misperception by ratepayers and regulators who are not experts in the uncertainty associated with savings estimates. It may be that these two savings claims, although different, are statistically the same because they are both within the bounds of the uncertainty range for the measure savings estimates. It can be perceived as unfair, however, to grant one implementer higher savings than another implementer for the same measure when there are financial penalties or bonuses on the line, even though they are statistically the same. Use of a TRM can reduce this risk to the regulator of misperception by the stakeholders.

A third benefit that is less often considered is that more reliable savings estimates will reduce the uncertainty in forecasting generation needs. Again, this is related to the primary purpose of reduction of risk.

Implementers' Purpose

Implementers are generally driven by a desire to reduce their own risk of not meeting their EEPS compliance targets by having a set standard (the TRM protocols) against which they will be judged for their energy savings compliance targets, whether that standard represents real savings or not. In most jurisdictions the implementers either face financial penalties for not meeting their targets, or miss out on financial incentives if they do not exceed their targets. This is real financial risk for the implementers.

Evaluation adjustments to program or portfolio level savings claimed by the implementers are typically applied through program or portfolio level realization rates (RR). The RR is the ratio of the evaluator verified gross[6] savings divided by the implementer claimed gross savings. Implementer compliance with their EEPS targets is generally judged against the claimed savings multiplied by the RR, sometimes also adjusting for net savings. Without a TRM used as a compliance standard, evaluators typically have freedom to use a different methodology than the implementer to estimate verified savings if they believe it to be more reliable, or they may use the same algorithm with different variable inputs than what the implementer used. This may lead to substantial RR adjustments to the claimed savings. When a TRM is in place as the standard, the implementer at least knows what standard they are being judged against and as long as they appropriately apply the TRM they can have confidence that their claimed savings will not be adjusted for differences in assumptions or methodologies.

Having a TRM as the standard simplifies implementation because the implementer does not have to gather as much customer specific information, which can be difficult to obtain, so long as they use what's in the TRM. Their risk is reduced because when they properly follow the TRM to claim savings, they do not face the possibility of those savings being significantly adjusted downwards through evaluation. They still face adjustments if they incorrectly applied the TRM, or made errors in their measure quantities, but theoretically this is easier to control for than the higher risk of not having a standard at all. Without a TRM against which they are judged, then they face this significant risk on an annual basis (assuming annual evaluation adjustments).

[6] "Gross" savings generally refer to measure savings prior to considering the effects of participant free-ridership (related to program influence on the participant decision making process), spillover (did the program influence additional energy efficient measure installations outside the program), and other program influences. "Net" savings include adjustments for these influences and are generally accounted for through a Net-to-Gross (NTG) ratio.

In jurisdictions where the implementers face financial penalties for missing a compliance target (or bonuses for exceeding them), using a TRM is perceived as a substantial risk reduction tool for the implementers as they are better able to plan to meet their compliance targets.

Standardization does not always reduce the implementers' exposure to risk, however. Let's look at the case in Pennsylvania. In 2008, the state of Pennsylvania passed a law (Act 129) setting a four year electric energy and demand reduction target for the largest electric distribution companies (EDCs) in the state and required the use of a statewide TRM (State of Pennsylvania, 2008). In this case, the EDCs and the regulator agreed on annual updates to the TRM because they found that the originally adopted TRM was missing many measures which were included in the EDCs' program plans, and many of the measure protocols were less accurate and reliable than desired. While the TRM was improving in accuracy and reliability through the annual updates, what also occurred with each update was an overall reduction in savings per measure as many measure protocols in the original TRM significantly overestimated savings. Because the measures in the TRM accounted for a large portion of the overall EDC portfolios, this in turn increased the EDC's acquisition costs per unit of energy savings year over year. What did not occur, however, was a similar reduction in compliance targets nor an increase in allowable spending. Each year the implementers saw reduced savings per measure and increased acquisition costs, but still had to meet their compliance targets with the same amount of funds. As the originally filed EDC energy efficiency and conservation plans were based on the savings in the first TRM, this annual reduction in savings per measure caused an increased risk to the EDCs of not meeting their compliance targets and facing significant financial penalties. While the increased risk in Pennsylvania was real, it is likely that this is not the norm as most jurisdictions do not have the same regulatory structure.

RRs vary from year to year regardless of whether a statewide TRM is in place or not, the question is by how much and is the long-term mean RR closer to 1.0 with a TRM or without. Theoretically, having a TRM that is used by all implementers within a jurisdiction should result in RRs closer to 1.0 on average than a jurisdiction without a TRM, especially if the evaluation is also required to use the TRM for verified savings estimates. If this is true it would make a strong case that using a TRM as a standard for all implementers will usually result in lower risk to the implementers of significant RR adjustments.

Although it is difficult to do an apples-to-apples comparison of various jurisdictions due to the regulatory and implementation differences, evaluation reports from several jurisdictions were reviewed and their RRs compared. Residential lighting and consumer product program evaluation RRs from multiple program years were gathered from 13 different implementers; 35 program evaluations with a TRM were compared to 15 without. Comparing the RRs of these program evaluations, for those jurisdictions with a TRM the mean RR was 0.98 with a standard deviation of 0.11. For those without a TRM the mean RR was 0.99 with a standard deviation of 0.23. Although the mean RR for both are close to 1.0, clearly the much higher standard deviation shows there is a larger range of RRs for those jurisdictions without a TRM, thus opening up the implementers to higher risk.

Similarly, RRs from 60[7] commercial and industrial program evaluations were compared; 30 with a TRM and 32 without. For the programs with a statewide TRM, the mean RR was 0.94 with a standard deviation of 0.15. For the programs without a statewide TRM, the mean RR was 0.96 with a standard deviation of 0.17. The results for the C&I sector show that there is not much difference in risk between having a TRM and not having one. The mean RR was closer to 1.0 for those without a TRM and the standard deviation was only slightly higher. There are some limitations to this comparison in that each jurisdiction uses the TRM differently for evaluation, and while all of the programs reviewed include TRM

[7] Total program counts do not match sum of TRM and non-TRM program counts because two programs included in the analysis were evaluated for one program year using both a TRM approach and an approach with no TRM to observe the RR effects between the two methods. In one C&I prescriptive program the TRM approach yielded a RR of 0.95 and the no-TRM approach resulted in a RR of 1.04. For the small business program the TRM approach found a RR of 1.03 and the no-TRM approach resulted in a RR of 0.86. Although interesting, this is too small a sample to draw any conclusions from.

measures, some of the programs in the comparison also include custom measures which may be skewing the results. This was common to both the TRM and non-TRM programs though and so it should not significantly affect the findings. It may be that the implementers need to review whether their assumption of risk reduction for C&I programs with a TRM is real.

Achieving the TRM's Purpose in a Regulatory Environment

When establishing the purpose of a TRM for a given jurisdiction, stakeholders should also consider the impact the regulatory environment will have on achieving that purpose. For example, is the regulatory environment penalty based, incentive based, or balanced? This can impact the level of effort the implementers put into fighting changes to the TRM, even though the changes may be increasing its accuracy. If the implementers face significant financial penalties if compliance targets are missed, how willing are they going to be to introduce information that may improve the accuracy of the TRM, but will also lower a measure's savings? If they have information leading to higher savings, however, they are more likely to be very willing to share the information. Unfortunately, penalties may set up an environment where all parties tend to be less willing to share information that may hurt them financially. Over time this can affect the reliability of the TRM and its purpose can be undermined. To counteract this some regulators or stakeholders may suggest including a conservative bias in savings estimates within the TRM. This too becomes a problem though as it may result in under-investment in beneficial measures which may no longer pass cost/benefit analysis tests even though they would pass if more accurate savings were used.

On the other hand, if the regulatory environment is less punitive and/or provides greater ease of adapting programs to TRM changes, the implementers may be more willing to pass on information that will improve the accuracy and reliability of the TRM, even if it results in lowering the savings estimates in future TRMs, thus reducing ratepayer risk. When the implementers have less risk of penalties, using a TRM may result in less risk to the ratepayers as well.

The impact of regulatory requirements and expectations should also be considered. We can see these impacts by looking at the differences between Pennsylvania and Illinois. In response to the Pennsylvania Act 129 (State of Pennsylvania, 2008), the regulator established a statewide TRM as the standard against which savings would be judged. As part of Act 129, the legislature required that the affected EDCs submit energy efficiency and conservation plans to the regulator for approval. As part of the process, the plans are released for a public comment and reply period. This allows the public an opportunity to review and comment on the implementers' plans prior to approval. If an implementer wishes to modify their programs they must submit a modified plan to the regulator which then must be released for a public review and comment period prior to approval. This process may take several months and includes many hurdles to get approval for even minor changes to a plan. As a result, if there are substantial reductions in measure savings in a TRM update, it is a difficult, lengthy and costly process for the implementers to make necessary changes to their plan in order to still meet their compliance targets. Although the regulator recognized this barrier and showed flexibility in creating an expedited review process for minor program changes, it is still a significant hurdle to overcome. The use of a TRM in this jurisdiction, coupled with difficulty in modifying implementation plans, has increased the risk of the implementers failing to meet their compliance targets (with failure subject to financial penalties). Thus, these well intentioned regulatory requirements have created barriers to achieving the purpose of the TRM to reduce risk.

In Illinois on the other hand, where there are also financial penalties for the implementers associated with missing compliance targets, the regulatory requirement is that the implementers are expected to quickly modify their plans (even within a program year if necessary) without going through regulatory approval when it is found that a measure saves less or more energy than originally anticipated. This regulatory expectation works well with a TRM as the standard because implementers have the ability to quickly adjust to changes and manage their risk if a measure's savings are decreased through a TRM update. Whereas in

Pennsylvania the implementers have more limited ability to rapidly adjust to TRM changes, even in the midst of rising risk.

Another consideration which can impact the objective of a TRM to reduce risk is whether all implementers are starting from a similar position when first establishing the TRM. Again using Illinois as an example, this question became a real challenge to overcome while developing the TRM. Before a statewide TRM was developed, each of the utilities in Illinois had their own energy efficiency implementation plans that were approved by the regulator through official dockets. Unfortunately several of the dockets had different deemed savings estimates for similar measures, thus the utilities would be impacted unevenly when the TRM was first instituted. During the TRM development process, the utilities discovered that some of them were going to see significant reductions in planned savings for measures for which they already had Commission approval, and this would restrict their ability to meet their compliance targets which had not changed. In some cases, all the utilities faced lower measure savings thus increasing their risk. For some utilities these impacts were large and added risk that was not expected when choosing to develop a statewide TRM. These issues had to be worked out in the TRM development process, however, the intended purpose for the TRM of reducing risk did not turn out quite as anticipated.

To avoid the pitfall of having a TRM which does not fulfill its intended objectives, the purpose of the TRM should be established and the impact all of these issues have on fulfilling that purpose should be thoroughly deliberated and addressed in the early stages of developing a statewide TRM, preferably before commissioning one. There is much more required than simply hiring good engineers with the technology expertise to create measure savings protocols. There are major policy issues that need to be thought through as well. To avoid confusion or misapplication of such a public document, we recommend that at a minimum a TRM include a section which provides common cross-cutting assumptions, definitions, and a clear description of the purpose and application of the TRM. Unfortunately defining the purpose and application of a TRM tends to be one of the most controversial parts of its development, and the importance and ramifications of these are often underestimated when making the decision to develop a statewide TRM. It is the authors' experiences with these efforts that initially led to the development of this paper.

Effects on Implementation and Evaluation

Another potential pitfall which can derail an otherwise well-written TRM is not comprehending and planning for its effects on implementation and evaluation. The relationships between implementation, evaluation, and the TRM should be considered and planned for at the start of TRM development, not after. To meet regulators' desires to improve reliability, programs must be evaluated to confirm the TRM specified savings are real. Implementers may prefer evaluators to only verify compliance with the TRM to meet their goal of reducing compliance risk. Will regulators have to compromise and accept a little less reliability, or will implementers need to compromise and be exposed to more risk? Is there a way to balance the needs of both? The framework and guidelines for how a TRM will be used by implementers, evaluators, and regulators should be written down early in the process as it affects what should go into the individual measure protocols. Without clear guidance the TRM may become a very expensive tool which no one knows how to use, or they will make their own interpretation as to how it should be applied which may undermine its intended purpose.

Two of the first questions to decide about a TRM are, "Will the TRM be a regulatory order to which implementers are "required" to follow, or will it be just a "guidance" document which the implementer may or may not use to estimate savings?" In other words, when can and should the implementer deviate from the TRM? Similarly, "Will the evaluator be "required" to follow the TRM when verifying implementer savings, or will it simply serve as another "reference" for the evaluator in their efforts to verify the actual/real savings for a measure?"

Mandatory or not?

Why would a jurisdiction develop a TRM, but not make it mandatory for the implementer to use? TRMs are developed by many people with various levels of training, often over a short time frame, on relatively small budgets (relative to their importance), and typically with incomplete or unreliable data. TRM measure savings estimates are just that, estimates, and they are often based on data that is either old, incomplete, from a different jurisdiction (often from the other side of the country), derived from a similar but different measure or program, and often based purely on engineering or professional judgment. Frequently all of these may be the case. It is also common that the writers of the protocol make mistakes. These issues can all lead to large inaccuracies for individual measures and if a TRM is a regulatory order that is required to be followed, the stakeholders may be uncomfortable allowing claimed savings that are suspect of being completely wrong, especially if they know what more accurate savings estimates should be. This is particularly true of implementers who are generally uncomfortable following a TRM measure protocol if they know there is a mistake in it. If the public found out savings were based on errors and overestimated by an order of magnitude, would they be happy with the implementer or the regulator?

It is reasonable to make the TRM a requirement for implementers to follow, but if there aren't exceptions to when a TRM must be applied, then these errors or inaccuracies will lead to less reliable claimed savings and possibly upset stakeholders. It is also reasonable to develop the TRM as a document that implementers agree to follow, but are not required to. This gives them freedom to deviate from it when they have better estimates. Even if a TRM is established as a requirement, there are times when implementers should be free to use better savings estimates, as long as they provide justification for why they are more accurate or reliable than the TRM estimates. When and how it is acceptable for the implementer to deviate from the TRM should be decided early in the process. If not in the TRM itself, the decisions should be documented in a policy framework that carries as the same level of authority as the TRM.

Auditor or Evaluator?

Whether the TRM is established as a regulatory order or not, stakeholders still need to decide what the role of the evaluator will be with the TRM in place? Will the evaluator be tasked with only verifying implementer compliance with the TRM; will the evaluator verify what the real savings are for each program; or both?

If the regulators' primary purpose for the TRM is to reduce ratepayer risk by improving the reliability of savings estimates, then claimed savings must be evaluated to confirm the savings specified by the TRM are indeed accurate. What we have seen in practice is that increased standardization through use of a TRM produces consistency, but does not always result in increased reliability of savings claims and reduced ratepayer risk. This is especially true when evaluation is reduced to simply verifying that the implementer followed the TRM methodology correctly, essentially performing the role of an auditor rather than the role of a true evaluator. This may actually result in increased risk to the ratepayers because the savings are being judged against the standard, but the standard does not necessarily represent the real savings.

TRMs are developed with history as a guide, but measure savings change over time depending on actual participation and market changes. The initial version of TRM measure many protocols are often developed quickly, using many assumptions, and may not accurately represent real savings. In some cases they may have been accurate when originally written, but the marketplace changes over time thus altering the real measure savings rendering the once accurate protocol obsolete. Hence the need for regular updates as well as regular evaluation (two separate issues). A principal purpose of energy efficiency program impact "evaluation" has historically been to determine if the implementer "claimed" savings have an inherent bias on the whole, either overestimating or underestimating savings. If a bias is found the evaluation makes a realization rate adjustment to estimate the verified savings. Notice the intentional use of "claimed" savings

rather than "measure" savings, because this often leads to misunderstanding of the role of evaluators in updating the TRM. We will come back to this issue in the Long Term Upkeep section below.

If evaluation is not allowed to do its traditional job of determining whether there is a bias (positive or negative) in the claimed savings at the program level or portfolio level, then it will not be known whether the savings are reliable or not. All we will know is whether the implementer followed the TRM correctly or not. If the evaluators' role is reduced to only verifying that the implementers followed the TRM correctly, then the regulators will lose out on their desire to have reliable savings estimates and may have a false sense of reliability in the savings claims. The only bias the evaluator will be able to identify is whether the implementer was consistent in applying the TRM or not.

The stakeholders need to determine if this is an acceptable risk as they think about how the TRM will be used and how it applies to the evaluators. In some cases it may be acceptable, particularly if there is a robust process to develop accurate TRM protocols. It may be that it is more cost effective to fund evaluators to do measure specific studies to update the TRM protocols than to fund a comprehensive program impact evaluation every year. In cases without such a robust TRM development process, it may be more cost effective to fund full impact evaluations on an annual basis.

A TRM can change the framework for implementation and evaluation. To improve reliability, there is a need by the regulator to show through evaluation whether the TRM claimed savings are real or not. To reduce their risk, however, implementers may prefer that evaluations only verify compliance with the TRM, not the magnitude of the real savings. How should the needs of both the regulator and implementer be balanced, and how does that affect the evaluator? These competing issues must be thoroughly contemplated and decisions as to how to handle them should be made based on a clearly stated purpose. Having an understanding of the TRM's purpose and the process to develop TRM protocols will help in making these decisions. It is best to think through these issues prior to developing a statewide TRM due to their impact on cost of both evaluation and TRM development.

Long Term Upkeep

TRM upkeep is often only vaguely acknowledged as a need to be addressed in the future until after the first edition of a TRM is completed. It shouldn't be such an afterthought as it can have negative ramifications as to the long term viability of the TRM. Stakeholders should plan for how will it be updated and how that will affect program planning and implementers' ability to meet compliance goals. As discussed above, there is potentially significant risk to implementers if the TRM changes measure savings in the middle of a program cycle, but there is also risk to the regulators if the TRM is not updated to reflect 'real' savings.

There are several questions dealing with TRM upkeep that need to be answered by stakeholders. How often should the TRM be updated; once a year, any time new information is available, only at the beginning of a planning/compliance cycle? What should the timing of those updates be? If the TRM is updated in the middle of a planning/compliance cycle, how does that affect program plans and compliance targets? What if there is an error in the TRM? How are new measures incorporated into the TRM? When can they be incorporated? What if an evaluation shows the savings are significantly less than what the TRM states? Are updates applied retrospectively? How are Evaluation results to be used for updating the TRM? When is it appropriate to update a measure protocol? How will it be determined that a measure should receive an update? Who will be responsible for maintaining and updating the TRM and how will this be paid for? It is difficult to answer these questions here as the regulatory environment can have a significant impact on the answers to these questions, but there are a few additional concepts that are worth pointing out.

To Update or Not?

It should be recognized that regardless of their differences, all TRMs are intended to represent the average annual energy savings over a period of time (generally over the effective useful lifetime of the measure) for a given energy efficiency measure, across a "population of installations". This is a key point that cannot be reiterated enough. A TRM measure protocol should not be interpreted to represent the true savings of any one particular installation. It is easy to fall into the trap of believing that a TRM measure protocol is unreliable because it doesn't accurately represent the savings for one particular installation of a measure or even for many installations. Because TRM protocols represent savings for a population of measures, they are generally more reliable for measures that have a greater number of installations relative to the variability of savings than for measures with only relatively few installations because there are more data to pull from and the estimates get closer to the real average as a result. But this only holds true when looking at the average savings for all the installations, not the individual installations.

That does not mean TRM protocols are not useful for less frequently installed measures. It is less important to have high accuracy for protocols that represent only a few installations with small savings since the measure may only represent a small portion (<<1%) of overall portfolio savings and being off by 100% will not materially affect overall compliance. Even if inaccurate, from an implementation standpoint it is more cost effective to have a standard savings estimate on hand for these small impact measures than it is to make them custom measures.

Due to this trait of TRMs, it is common to assume that when a new reference is found that provides a different savings estimate for a measure, then the existing protocol must be wrong or inaccurate. What needs to be taken into account (and often isn't) is whether or not the new estimate is within the error bounds or confidence bounds of the original estimate, and/or whether the new estimate is truly more reliable (e.g. tighter precision). It may be that the new estimate is more accurate because it was based on a more thorough study, or because errors were found in the old protocol and it should be updated, but this is not the scenario we see as a pitfall , becausethe protocol should be updated in these cases. It is more common that the new study is simply "newer" or it is closer geographically or climatically than the original reference, but it has the same flaws as the original reference, is based on a smaller sample or has more significant weaknesses. In these cases it is necessary to look at the difference between the references and determine if there really will be an improvement in reliability and accuracy by updating the TRM to use the new reference.

Every study has an error band and range of reliability and, although these are often not specifically stated, they must be considered prior to making an update. If the new study and old study both have an error band of ±25% and they overlap by 15%, then it is possible that one is not better than the other and it may not be worth updating the protocol. On the other hand, if the original study has an error band of ±25% and the new study is within that error band, but it only has an error band of ±5%, then it may be worth updating the protocol using the new reference. These are questions that should be reviewed every time a protocol is considered for updating.

It is possible that a protocol should not be updated based on any one study, but if several studies indicate a bias in one direction or the new study is much more thorough and reliable than previous references, than the protocol likely should be updated.

Who Will Identify Needed Updates?

At this point we would like to return to discussing the difference between "claimed" savings and "measure" savings. In this context "claimed" savings is used to represent the total program or portfolio level savings claimed by the implementer. "Measure" savings are simply the savings estimates for any one measure as specified in the TRM. This is discussed in this section, because a common error in assumptions by regulators and implementers is that needed TRM measure updates will be identified through the normal evaluation activities. Along with this, it is assumed that the evaluators will also be collecting all the necessary information to update the identified TRM measure protocols.

Evaluators rarely have the opportunity or budget to delve deeply into each and every one of the measures in an implementer's portfolio, let alone all the measures in a TRM which may cover multiple portfolios. Thankfully, to determine bias of overall program or portfolio level "claimed" savings they don't have to. It is possible through a sampling approach to determine if an implementer had a bias in their "claimed" savings at the program or portfolio level, but if this were required for each "measure" savings, the budget for evaluation would need to be increased dramatically. This is because each measure would need a statistically valid sample to properly determine bias, rather than the standard practice of sampling at the program or portfolio level. Given that many implementers have multiple programs with sometimes over a hundred measures in each program, it is easy to see how this could result in exorbitant evaluation samples and budgets.

Because of this it is an unreasonable assumption that the evaluators will identify all needed TRM updates and collect all the necessary data without an increase in budget or change in scope. There will be times when particular errors or unreliable savings estimates will be identified through the evaluation, but this will likely be more by chance that the evaluation identified a particular issue, rather than an inherent capability of the standard evaluation. If the stakeholders do look to the evaluators to identify needed updates and gather the necessary data to make the updates, then this should be done separately from the typical impact evaluation activities, and allocated additional budget. This type of TRM update work often requires metering studies, calibrated simulation modeling, or other detailed engineering approaches not normally included in an Evaluation budget or scope.

While a regulator may have reservations about depending on the implementers to identify which measures need updating, it is probably an unnecessary concern. It is the implementers who best understand the details of each measure in the TRM and how well it represents the measures in their programs. As a result they often have the most insight as to which TRM protocols are inaccurate. In practice, identification of which TRM protocols need updating is best done by anyone who has insight into the accuracy of a measure. If the information comes from the implementer, evaluator, regulator, or the public, it is still valid and the stakeholders should not ignore it. There should be a process in place to prioritize which measures receive extra funding to get updated based on the overall impact and value of information, but every suggested update should be considered.

TRM Administrator

Initial TRM development plans rarely seem to include plans and budget for long-term upkeep and maintenance. Unfortunately, this often unanticipated expense has the potential to quickly doom a TRM to functional obsolescence.

In Illinois and Pennsylvania the original orders to adopt a TRM included budget and planning for the initial TRM, but did not include budgets and plans for maintaining and updating the TRM. In Pennsylvania this was eventually tasked to the Statewide Evaluator, although much of the work is also done by the utilities, along with their implementers and evaluators. This was not originally planned or budgeted for by any of the parties, and has required scope of work adjustments as well as shifting of budgets. There have also been requests by the stakeholders for budget approval of additional metering or engineering studies to update the TRM which were not originally planned. In Illinois, the need for a TRM administrator was recognized during the TRM development, but after the original budget was approved. Additional budget was eventually approved and a contract awarded to an independent TRM Administrator.

There are several options for who will be the Administrator. This could be handled by a designated and chartered body such as the RTF in the Pacific Northwest, a statewide evaluator hired by the regulator such as in Pennsylvania, the regulator itself, the implementers and their evaluators, or an independent contractor, such as in Illinois. There are advantages and disadvantages to each and what may be right for one jurisdiction may not be for another. If hiring an independent contractor is the selected path, however, it is necessary that the contractor spend a considerable effort understanding the regulatory framework of the

jurisdiction and work closely with the regulators, implementers, evaluators and public to ensure the TRM meets the needs of all and fits within the applicable regulations.

TRM maintenance and upkeep cannot be an afterthought. If not updated, it may quickly become inaccurate and unreliable. If there is not a plan and budget to maintain and update it, the TRM will likely fall by the wayside as an expensive tool that goes unused. How and when the TRM will be updated needs to be addressed, but this must be done in light of each particular jurisdiction's regulatory environment. TRM updates and upkeep can significantly affect implementers' program planning and ability to meet goals. If TRM measure savings are changed midstream in a program cycle, implementers may be at risk; but, ratepayers also bear risk if the TRM is not regularly updated to reflect 'real' savings. TRM development committees may clamor for research to update the TRM, while implementers may push back on frequent updates to reduce their risk. The risks of all must be weighed when making decisions on these issues.

Budgeting for a TRM

Finally, stakeholders must ask, "How much will it cost?" When approving the development of a TRM, it is the upfront direct cost to hire a TRM Developer that is usually considered, both by the regulator and implementers. While this direct cost may seem reasonable, it might not reflect the true cost to develop a TRM.

Once the costs of review for the regulators, implementers, evaluators, and the public are included, the true upfront cost can easily double, triple, quadruple or more. After it is developed there are the additional direct costs to maintain the TRM by some designated TRM administrator. There are costs to develop new measures, correct mistakes, and update existing measures, and there are additional costs to approve any changes through public comment processes.

The indirect costs to develop a TRM are also significant and can quickly add up. Indirect costs are borne by regulators who oversee the process of maintaining, updating and approving the TRM; by implementers who need to review the TRM during development and each update cycle, tailor tracking systems to match the TRM both initially and for each update cycle, update forms, and modify data collection requirements for TRM inputs; by evaluators who often are called upon to review the TRM during development and update cycles, to provide research into the reliability of TRM protocols in addition to their primary evaluation responsibilities, and to write new TRM measure protocols; and by the public who may also participate in the initial development process and update cycles.

These direct and indirect costs quickly add up and if not properly accounted for and planned for can overwhelm the initially perceived value of the TRM or derail its progress altogether. The total costs cannot be determined until the TRM objectives are clarified and all its impacts are understood, but the impacts of different policy and design choices can be considered and compared to gain a more solid understanding of what the TRM will cost. Relationships between implementation, evaluation, compliance, and the TRM are issues with significant cost ramifications. Further, the long term upkeep should also be included and planned for. To avoid cost becoming a TRM pitfall, all of these costs should be considered and planned for prior to embarking on the development of a TRM, not after.

Conclusion

Several states have recently entered the path of developing a statewide TRM. Some are a few years into it; some are just beginning. Experience has shown there are often unexpected pitfalls along the path to develop a statewide TRM. The best way avoid these is to learn from others who have already taken the path of developing a TRM. The experiences presented here from multiple states at various stages of TRM development, each with unique issues to deal with, can be used by others to guide their planning and

decision making process, either in choosing whether a TRM is right for them, and/or in the development process itself.

This paper is not intended to recommend or dissuade others from adopting a TRM as that decision should be based on the unique needs and regulatory environment of each jurisdiction. Regardless of the path chosen, these experiences provide several questions that should be answered and pitfalls to avoid in the decision, planning and development stages of a statewide TRM.

References

California Public Utility Commission. 2008. *California Database for Energy Efficiency Resources (DEER).* Retrieved May 23, 2013, from The California Energy Commission: http://www.deeresources.com/

Morgan Marketing Partners. 2013. *Michigan Energy Measures Database (MEMD).* Retrieved May 23, 2013, from Michigan Department of Licensing and Regulatory Affairs; Public Service Commisison: http://www.michigan.gov/mpsc/0,4639,7-159-52495_55129---,00.html

National Renewable Energy Laboratory (NREL) and The Cadmus Group, Inc. 2012. *Uniform Methods Project (UMP).* Retrieved May 23, 2013, from NREL Uniform Methods Project: http://ump.pnnl.gov/forum.php

Regional Technical Forum (RTF). 2012. *Guidelines for the Development and Maintenance of RTF Savings Estimation Methods.* Retrieved May 24, 2013, from Regional Technical Forum: http://rtf.nwcouncil.org/subcommittees/guidelines/RTF%20Guidelines%202013-04-16.pdf

State of Pennsylvania. 2008. Act 129. *Pennsylvania House Bill No. 2200, Act 129 of 2008.* Pennsylvania.