How to Show the Implementer What to Do Next

Erik Mellen, Eversource, Westwood, Massachusetts
Susan Haselhorst, ERS, North Andover, Massachusetts
Isaac Wainstein, ERS, New York, New York

Abstract

After considerable effort and expense, and M&V impact evaluators deliver a verdict: a grade-like realization rate identifying the proportion of tracking savings that can be claimed by the program. When faced with a poor realization rate, the implementer must decide what to do next.

This paper presents a heuristic approach called the "allocation method" for systematically capturing the discrepancies revealed in the M&V process, aggregating the results, and reporting findings in a manner that can guide implementer action. This is followed by a presentation of a discrepancy summary developed using the cascade method a more computationally straightforward and traditional method, but one that can lead to incorrect aggregate results. The balance of the paper describes the design elements of the allocation method, including an analytical framework, an approach to systematic site characterization, and a method for synthesizing the results.

Introduction

At the conclusion of an impact evaluation, the implementer must decide on next actions. Should the implementer increase review staff, change savings algorithms, or fix the tracking system? Which of the changes are worth the expense? In a typical on-site measurement and verification (M&V) evaluation, there is a rich data set produced by M&V engineers that can inform that decision.

The engineers have recruited, visited, inspected, metered, and analyzed a statistically representative set of projects. The engineer knows where an error in tracking resulted in underestimating savings by 10% at one site and where an incorrect baseline caused an overestimation of savings by 80% at another, partially mitigated by a 20% understatement of hours of operation at the same site. How does one combine and synthesize these disparate site-specific discrepancies into a useful description of the factors driving the program realization rates?

The process of translating the site-specific observations into meaningful and quantitative programmatic findings is surprisingly challenging. It requires using an "implementer-centric" framework for categorizing discrepancies, employing a disciplined and consistent approach to assessing the category and magnitude of the errors found at each site, and making appropriate analytical choices for conveying the impact of the error program-wide.

This paper presents a heuristic approach, dubbed the "allocation method" as opposed to the "cascade method," for systematically capturing the discrepancies revealed in the M&V process, aggregating the results, and reporting findings in a useful way This is followed by a presentation of a discrepancy summary developed using the cascade method, a more computationally straightforward and traditional method but one that can lead to incorrect aggregate results. The balance of the paper describes the design elements of the allocation method, including an analytical framework, an approach to systematic site characterization, and a method for synthesizing the results.

Key Definitions

Before proceeding further, definitions are in order. In this paper, a “discrepancy” is the change in savings (kWh or MMBtu) associated with a difference between a program-reported and an evaluated parameter. A discrepancy parameter is a value, assumption, or process used to estimate savings. The discrepancy rate, similar to a realization rate, is the ratio of the evaluated to program-reported discrepancy parameters. The discrepancy category is the category under which a particular discrepancy will be reported to an implementer. Examples of the use of the terms follow:

- The evaluator’s metering of a lighting measure yielded 3,200 annualized hours of operation while the program had used the technical resource manual deemed hours (2,900) as the source of the hours of operation. The tracking savings for this measure was 1,000 kWh. The engineer categorized the discrepancy as a Deemed Assumption discrepancy. The discrepancy rate for the site is the ratio of the discrepancy parameters of hours of operation, 3,200/2,900 or 1.103, yielding a discrepancy of 103 kWh.
- The tracking savings for a measure was recorded as 1,705 kWh in the electronic tracking data, while the application documented savings of 1,507 kWh. The discrepancy of -198 kWh was categorized as Administrative discrepancy. This particular error appears to be a typographical data entry error (transposed digits), but another common source of errors in this category is a failure to update tracking systems with the most recent savings estimate.
- For a boiler replacement measure, the program had assumed the preexisting efficiency of 70% as the baseline efficiency. The evaluator determined the boiler was at end of life and therefore a code efficiency of 80% applied. The evaluator used the ex ante and ex post efficiencies as the discrepancy parameters to compute a discrepancy of -400 therms. The discrepancy category selected was End-of-Life Baseline.

The program discrepancy rate is the difference between the program realization rate (the savings that was achieved) and 1.0. A program with a realization rate of 70% has a discrepancy rate of 30%. The purpose of this paper is to present a methodology to help the implementer to better understand the actions they can take to improve the program by describing the factors that compose that 30%.

Most evaluated projects deviate from the program-reported basis in multiple ways. For any one type of discrepancy category, e.g., hours of operation, the application estimate may be lower (a negative discrepancy) or higher (a positive discrepancy) than the evaluated hours of operation.

Outcomes: What Does an Implementer Really Want?

To an implementer, a realization rate might appear to be a necessary evil that sums up the past in a single number. Except in the broadest sense, it is not a diagnosis; nor is it actionable. A forward-looking implementer wants to improve his/her program and to do that, needs to know the details of what went wrong and by how much in order to figure out how to make the program better in a cost-effective manner. An implementer’s wish list for a useful program diagnosis is likely to include the following features:

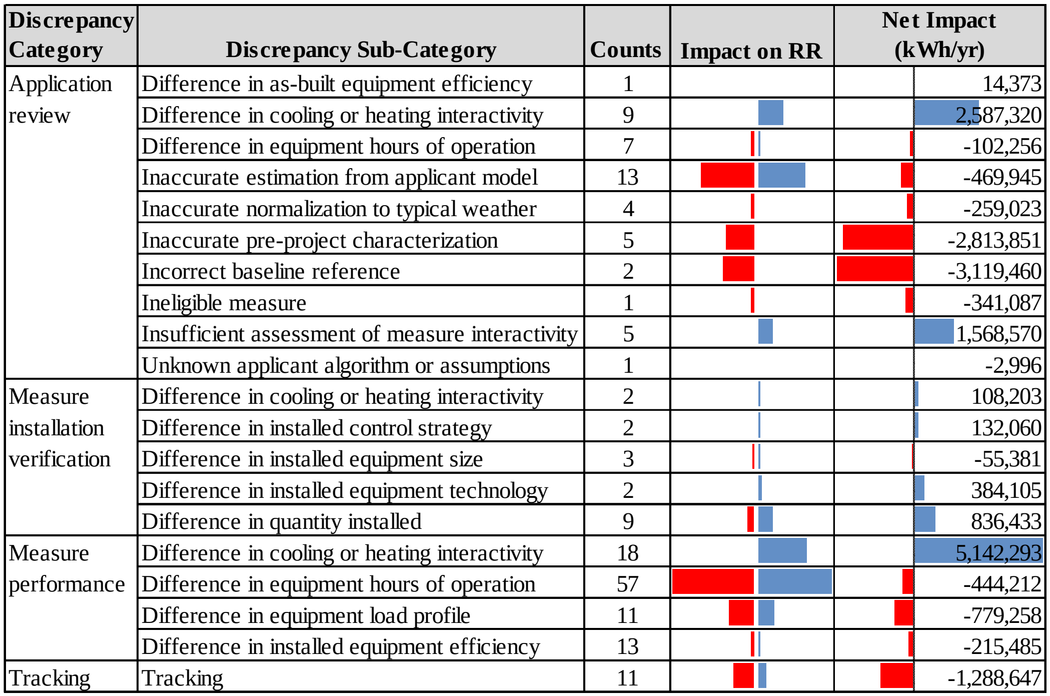
- An implementer manages a process and will therefore find value in having the discrepancy categories mapped to the phase of the implementation process where they occur.
- Granular and specific discrepancy categories are helpful – to a point. For example, it may be helpful to an implementer to distinguish between prescriptive and custom hours of operation discrepancies,

since the processes for estimating hours for prescriptive and custom hours are different. However, reporting too many categories can obscure underlying trends, which might become apparent with the appropriate aggregation of categories.

- The magnitude of the impact discrepancy category is a key finding. Larger discrepancies warrant more action from the implementer, all else being equal. As a corollary, the magnitudes of the discrepancies should indicate their relative importance to each other.
- When the implementer reviews a report of program discrepancies, they should be consistent with the evaluated savings and the program gross realization rate. In other words, the sum of the magnitudes of all the discrepancies should equal the discrepancy rate.
- Both the positive and negative components of a discrepancy category, as well as the net impact on the outcome, should be reported so that the range is apparent.
- The number of times the discrepancy was observed should be noted as an indicator, not only of the magnitude of the discrepancy, but also of its distribution through the program.

Table 1 presents a composite example of a program discrepancy summary produced from a large commercial and industrial (C&I) retrofit program. The study was based on the on-site M&V of 102 sites. The discrepancy categories are explicitly organized by where they occur within the implementation process. For example, it is in the measure installation verification phase that the error in quantity of units installed and the equipment sizing could have been observed and subsequently corrected. Often data of this kind is presented by order of magnitude, where the highest positive magnitude line item is listed first (in this case interactivity). That organization, however, does not facilitate answering the questions related to the process such as: Are my application review guidelines adequate for my review engineers? Do we have enough inspectors deployed to verify installations?

Table 1. Implementer Oriented Discrepancy Results



Discrepancy Category	Discrepancy Sub-Category	Counts	Impact on RR	Net Impact (kWh/yr)
Application review	Difference in as-built equipment efficiency	1		14,373
	Difference in cooling or heating interactivity	9		2,587,320
	Difference in equipment hours of operation	7		-102,256
	Inaccurate estimation from applicant model	13		-469,945
	Inaccurate normalization to typical weather	4		-259,023
	Inaccurate pre-project characterization	5		-2,813,851
	Incorrect baseline reference	2		-3,119,460
	Ineligible measure	1		-341,087
	Insufficient assessment of measure interactivity	5		1,568,570
	Unknown applicant algorithm or assumptions	1		-2,996
Measure installation verification	Difference in cooling or heating interactivity	2		108,203
	Difference in installed control strategy	2		132,060
	Difference in installed equipment size	3		-55,381
	Difference in installed equipment technology	2		384,105
	Difference in quantity installed	9		836,433
Measure performance	Difference in cooling or heating interactivity	18		5,142,293
	Difference in equipment hours of operation	57		-444,212
	Difference in equipment load profile	11		-779,258
	Difference in installed equipment efficiency	13		-215,485
Tracking	Tracking	11		-1,288,647

Note: The blue bars indicate positive savings whereas the red bars indicate negative savings.

The third column in Table 1 above displays the number of sites where the discrepancy category was observed. A discrepancy observed at most of the sites reveals a systemic problem, while a large magnitude discrepancy at a single site may be an outlier and require a different approach to resolve. The fourth column in the table displays the positive and negative discrepancies for each discrepancy category.

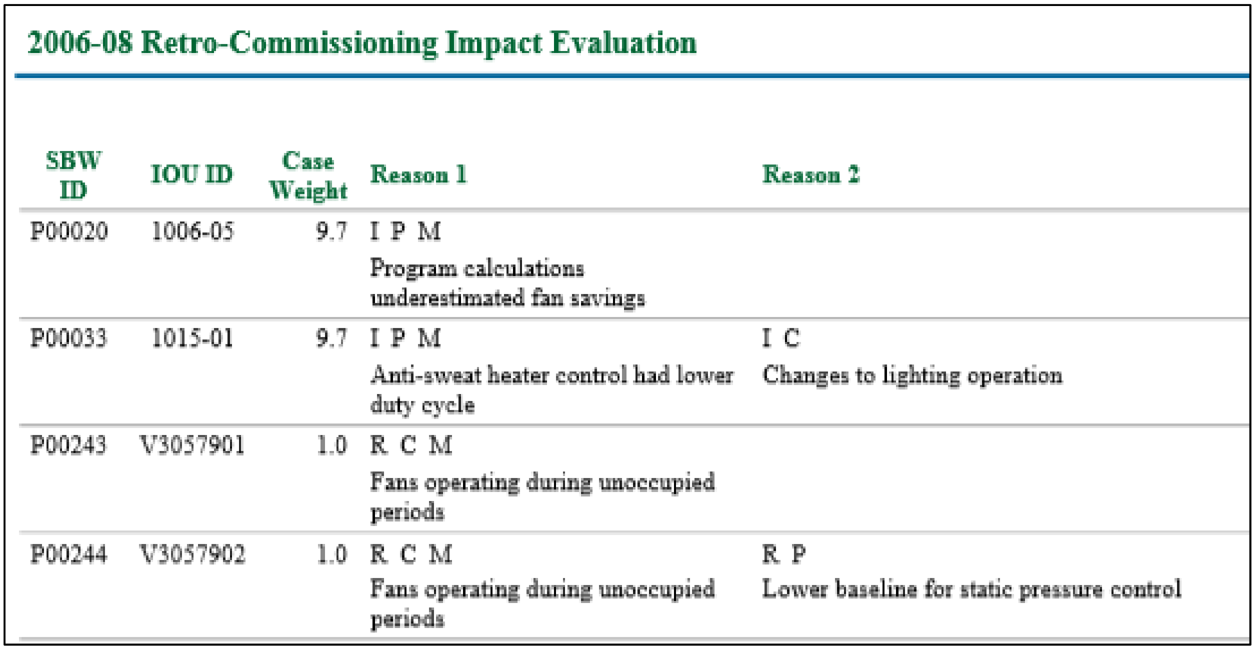
What follows is an illustrative scenario that demonstrates how an implementer might interpret Table 1 and the possible actions that he/she might take. This “dialogue” is a composite of actual conclusions and findings gleaned from multiple impact reports.

- The implementer examines the inspection phase first, relieved to see that his/her fiduciary responsibility to ensure that incented equipment is installed and operating has been fulfilled. The inspection processes appear to be adequate and in control, with no significant positive, negative, or net discrepancies.
- Moving to the measure performance phase, the hours of operation discrepancy category show a large positive and negative component across a large number of sites; however the good news is that the discrepancies largely net out. Hours of operation are difficult to forecast, so the widespread incidence of discrepancies is not unexpected. The implementer concludes that better estimates would require ex ante metering, which is time consuming and expensive and makes a mental note to explore this further with the impact evaluation team.
- The application review process could, however, be improved cost-effectively. One unhappily noted discrepancy category is the End of Life Reported as Retrofit, which captures those sites for which the applicant claims the existing equipment as baseline, the evaluator determines that the equipment was at the end of its life, and code or standard practice is the correct baseline. The implementer reluctantly agrees, after further examining the site reports, that the evaluator’s conclusions are correct. This category results in only a negative discrepancy. The implementer plans to provide the application-review engineers with additional training using the discrepancy sites as case studies.
- The implementer sees that interactivity is underestimated and appears as a discrepancy category in both the Application Review and Measure Performance categories. The evaluator explains that the lighting applications do not capture the cooling bonus at all and that it would be possible to better estimate that bonus at the application review stage. The interactivity category in the later stage arises in those cases where the applicant did estimate an interactive factor, but it did not match the evaluated interactive factors. The implementer does not see an easy way to incorporate the cooling bonus in either the application or the tracking system without muddling the basis for the incentives, which does not include cooling bonus. The implementer ponders how this might be included in program tracking totals, if not application totals.

The procedure for producing Table 1 is the focus of this paper; however, before proceeding to that topic, it is worthwhile to look at common alternative presentations of discrepancy data.

Alternate Methods

Evaluators usually identify, at least anecdotally, the sources of discrepancies discovered in an impact evaluation. In the following example, the evaluator diligently reported primary sources of discrepancy site by site, as illustrated in Figure 1, which is an excerpt from eight pages of tables. With this approach, the implementer can deepen their understanding of what went wrong at a particular site; however, it is difficult to identify trends or prioritize next steps from lists of discrepancies.



2006-08 Retro-Commissioning Impact Evaluation

SBW ID	IOU ID	Case Weight	Reason 1	Reason 2
P00020	1006-05	9.7	I P M Program calculations underestimated fan savings	
P00033	1015-01	9.7	I P M Anti-sweat heater control had lower duty cycle	I C Changes to lighting operation
P00243	V3057901	1.0	R C M Fans operating during unoccupied periods	
P00244	V3057902	1.0	R C M Fans operating during unoccupied periods	R P Lower baseline for static pressure control

Figure 1. Example of discrepancy lists

Typically, when sources of discrepancies are identified quantitatively, the evaluator cascades the discrepancy calculations, i.e., the magnitude of one discrepancy impact is dependent upon the magnitude of the prior discrepancies. In a cascade methodology, the order of the calculation of the discrepancy matters. The discrepancy calculated first will *appear* to have the most program impact, and subsequent discrepancies will appear to have less impact.

This is best illustrated by a simple example, as shown in Table 2. In the table, the Applicant column represents the assumptions and the basis of the tracking savings. The Evaluator column reports the observed quantity, hours, and delta watts observed by the evaluator. The ensuing savings and discrepancy are shown as well.

Table 2. Savings Calculation Example

Calculation	Applicant	Evaluator
Number of fixtures	100	60
Hours of operation	4,000	2,400
Delta watts (kW)	0.022	0.022
Site savings: Number of fixtures x Hours of operation x Delta watts	8,800 kWh	3,168 kWh
Site discrepancy		5,632 kWh
Site discrepancy rate		64%

In presenting the discrepancies identified in Table 2 in a cascaded model, the evaluator must choose the order in which the discrepancies will be calculated. Table 3 illustrates the principle using the example from Table 2. In the table on the left, the Installed Quantity category was selected first, while in the table on the right, it was selected second. In both tables, the first discrepancy has an impact of -3,520 kWh (40% of tracked savings of 8,800), while the second discrepancy is computed using not the tracking savings, but the tracking savings reduced by the first discrepancy (the quantity (8800-3520) x 40%). In the table on the left, the Installed Quantity appears to have the largest impact while on the right, the Hours of Operation category appears to have the largest impact even though both discrepancy categories contribute equally to the outcome.

Table 3. Impact of Order in a Cascaded Model

Discrepancy Category	Discrepancy Ratio	Discrepancy (kWh)
Program tracking savings	N/A	8,800
Installed quantity	40%	-3,520
Hours of operation	40%	-2,112
Program evaluated savings	36%	3,168

Discrepancy Category	Discrepancy Ratio	Discrepancy (kWh)
Program tracking savings	N/A	8,800
Hours of operation	40%	-3,520
Installed quantity	20%	-704
Program evaluated savings	52%	4,576

N/A = Not applicable

This, of course, can be explained, but it does obscure an intuitive grasp of the relative importance of each discrepancy category, particularly if there are more than a few. However, the methodology does have its place, as illustrated in the next two examples.

Table 4 presents discrepancy results from a small business direct install lighting program where the sources of the discrepancy are limited. It was possible to calculate the cascaded discrepancies with reported relative precision because of the fact that the calculation for each measure and site were exactly the same, and all the calculations were computed using a single data set. Thus, the parameter change for each discrepancy, starting with the calculation of the impact of the Installed Quantity discrepancy, could be recomputed for the entire sample including all the statistics. The result of the Installed Quantity calculation was the "baseline" for the Calculation Adjustment computation, which in turn became the baseline for the Operating Hours Adjustment discrepancy computation.

Table 4. Cascaded Discrepancy Report for a Small Business Direct Install Program

Discrepancy Category	Percentage of Change in Parameter	Discrepancy (MWh)	Discrepancy Impact -%	Relative Precision 90%	Description
Program tracking savings	N/A	109,000	N/A	N/A	N/A

Discrepancy Category	Percentage of Change in Parameter	Discrepancy (MWh)	Discrepancy Impact -%	Relative Precision 90%	Description
Installed quantity	6.1%	-6,600	N/A	2%	Number of units installed vs. tracked
Calculation adjustment	2.1%	-2,200	N/A	3%	Wattage, technology, HVAC differences
Operating hours adjustment	18.6%	-18,600	N/A	8%	Operating hours differences
Program evaluated savings	N/A	82,000	N/A	N/A	N/A

N/A = Not applicable

In the case where there are a few major sources of discrepancies, the cascaded approach can be accurately calculated. However, note that the range in discrepancies is lost and the multiple discrepancies of wattage, technology, and HVAC are collapsed into a single discrepancy in order to make the calculations manageable.

Discrepancy Allocation Model

The cascade method of presenting discrepancy results can be useful for programs where there are few sources of discrepancies or where the planned order of program remedies is known. However, for programs with many sources of discrepancies or where the program implementer's remediation has not been identified, the cascade method can easily lead to false conclusions about which factors are the largest contributors to the discrepancy rate.

The allocation method is particularly well suited to a C&I impact evaluation conducted through the evaluation of a sample of projects that undergo an engineering review. The source data is produced by the M&V site engineer responsible for the site – the person most in tune with the particular reasons for differences. The method is flexible enough to include any type of measure and also practical enough to implement with good planning. It is quantitative and unbiased. Both negative and positive discrepancies can be captured for each of a large number of individual discrepancy categories.

The fundamental difference between the two methods is that the cascade method calculates each discrepancy as it relates to a previous discrepancy, and the allocation method calculates each discrepancy independently, as if it were the only discrepancy impacting tracking savings, and then integrating the results.

An overview of the discrepancy allocation method appears in Figure 2.

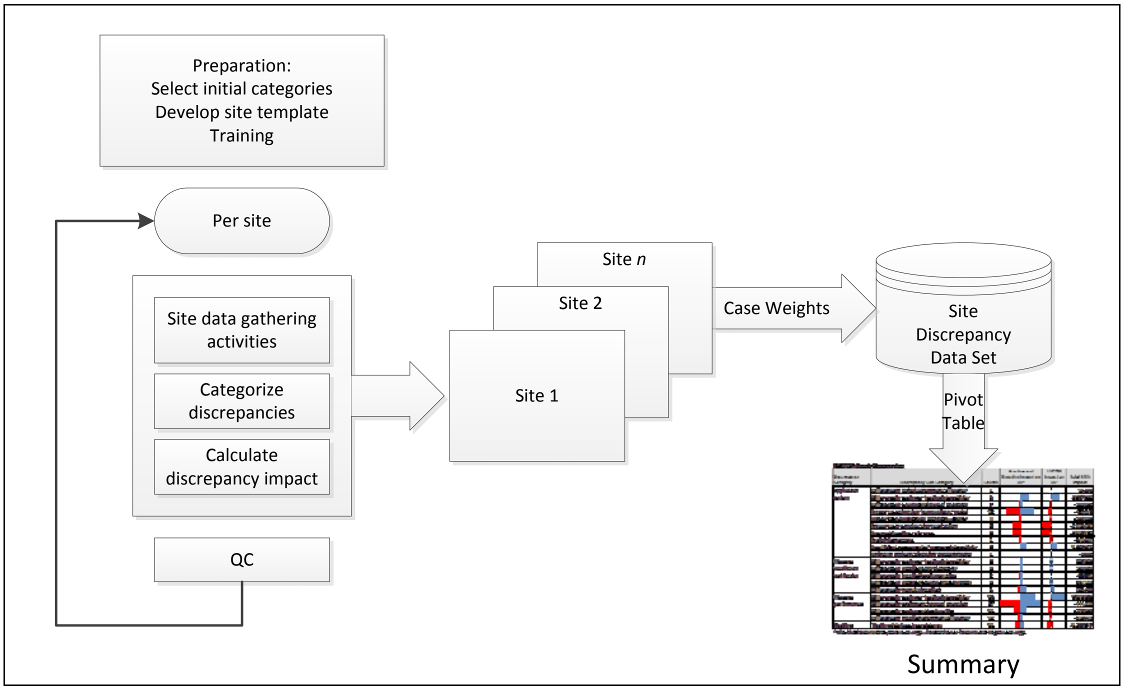


Figure 2. Discrepancy Implementation Process

The allocation method discrepancy analysis must be planned in advance, with the site discrepancy analysis conducted as an integrated part of producing the M&V site report. Attempting to do the analysis well after the site work has been done will be expensive and likely much less thorough. The planning work includes developing a discrepancy tool that is pre-populated with likely discrepancy categories and training the engineers in the purpose and use of the tool. The data set is built from the bottom up. For each site, the site engineer completes a "discrepancy tool," which is a spreadsheet template. This step leverages the engineer's intimate observations of conditions to tease out the major differences between the applicant and the evaluator assumptions and analysis of savings. The tool systematically compiles the discrepancies observed by the engineer for that site including the discrepancy category, associated discrepancy parameters, and estimates of the discrepancy impact on site tracking savings. The site discrepancy spreadsheet is QC'd by a senior engineer to ensure that the engineers are consistently classifying and calculating discrepancies.

Once the site-by-site data collection has been completed, the individual site discrepancy tool results are uploaded into a common data set, where the discrepancies and any site residual error can be weighted (using case weights) to the program level and analyzed and synthesized into the results. Ideally, the site residual errors will sum to zero, and the sum of the site discrepancy rates will equal the program discrepancy rate. Where this is not the case, the high influence sites are examined further to make sure the results are correct and that the site residual error is minimized.

When the interactive discrepancy impacts for a site are summed up, it is likely that the total of the discrepancies will not equate to the program discrepancy rate because of interactivity and other uncertainty. The final step in the process is to scale (or allocate) the magnitudes of the program level discrepancies to match the program discrepancy rate.

Site Discrepancy Process

For a typical M&V site, the engineer visits the site, reviews the application, installs metering, and conducts analysis to arrive at the evaluated site savings. In this process the engineer is likely to identify specific differences between the applicant assumptions and the evaluator findings. The site engineer's task is to identify and quantify the major sources of discrepancies and the related discrepancy parameters in a manner that explains the numerical difference between the evaluated value and the tracking value (the site discrepancy).

After cataloging the reasons for discrepancies, the next step is to calculate the independent impact of each discrepancy on the site tracking savings. The method for calculating the discrepancy impact depends upon the parameter and other analytical aspects of the project. The impact on tracking savings of a discrepancy can often be estimated assuming a linear relationship between the savings and the discrepancy parameter when that parameter is used multiplicatively in a savings equation. In Table 5, simple ratios of evaluated discrepancy parameter/tracking discrepancy parameter for the Hours of Operation and Installed Quantity produce the discrepancy rate. The product of this percentage and the tracking savings is the independent estimate of impact on savings. Both of these parameters are negative discrepancies as they reduce the savings. The third discrepancy, which is a positive discrepancy increasing savings, was estimated to produce an 8% impact on savings (assumed from the engineer's model of cooling interactivity).

Table 5. Example

Tracking savings	8800			Site	#100
Evaluated savings	4100				
Discrepancy Category	**Tracking Discrepancy Parameter**	**Evaluated Discrepancy Parameter**	**Discrepancy Rate**	**Discrepancy (kWh)**	**Residual Savings (kWh)**
Installed quantity	100	60	-40%	-3520	5280
Hours of operation	4000	3000	-25%	-2200	3960
Interactivity	100%	108%	8%	704	4277
Residual					-177

Table 5 includes a "residual savings" column that maintains a running estimate of savings as each discrepancy is cascaded into the savings estimate. The savings after factoring in the Installed Quantity is 5280 kWh. The Hours of Operation discrepancy has an independent discrepancy of 2200, but its cascaded (or interactive) effect on savings is -1,320 kWh, yielding a cumulative savings of 3,960 kWh. After the Interactivity Discrepancy is cascaded, the savings is 4,277 kWh. The residual error, or residual, is the difference between evaluated savings and the cumulative savings after the last discrepancy has been accounted. It is a goal of the engineer to minimize the residual.

For parameters that are inversely related to savings, such as efficiency, a slightly more complex calculation is required incorporating both the ex ante and ex post efficiency assumptions. For example, the magnitude of the savings impact is not properly represented by the ratio of the ex ante efficiencies or 75%/80%, which would yield a discrepancy of 6%, but rather the ratio of the change in efficiency calculated as: 1/ex ante baseline efficiency – 1/ex post efficiency, which is a much larger discrepancy rate. Building simulation models can be used to directly estimate the impacts of changes in parameters and the difference between two model runs used as the independent estimate of savings.

The task at the site level is to identify the key factors driving discrepancies such that the total site discrepancy is largely explained as indicated by a small residual error. Given that the applicant and evaluator methodologies can be very different and that savings are sometimes computed with complex models that are difficult to compare directly, it is rarely possible to identify exactly all the reasons that the program and evaluated savings diverge. Each site is likely to have an unexplained residual discrepancy, as shown in Table 6; however, the engineers strive to minimize this unknown residual by carefully considering the program's and evaluator's models of savings.

Residual error arises when the simple discrepancy impact models do not fully capture the complexity of either the impact or the interaction. The boiler example above uses a simple ratio of the rated combustion efficiencies. However, the modeled boiler efficiency is not constant across the firing range and the ex ante and ex post combustion curves are not parallel or linear; thus a simple ratio will approximate, but not exactly capture the actual or modeled difference. This second type of difference is retained in the analysis and is expected to be random and to cancel out programmatically and is one measure of how good the discrepancy analysis is on the whole.

In order to compile the discrepancies into a common database, the engineers will enter results in a consistent manner. A common site-specific spreadsheet tool can facilitate the engineering effort when it includes pick lists as a catalog of discrepancy reasons structured to tabulate the independent impacts of each discrepancy while monitoring the sum of the interactive results so the engineer may track the unexplained residual.

As a final point, the standard of rigor for computing discrepancies is more relaxed than the rigor required to compute the site realization rate. Sometimes the reasons for discrepancies cannot be discerned fully. For example, a billing analysis may provide a high level of confidence in a gas program measure evaluated savings, but not provide much insight into the reasons for a very large or small realization rate. Sometimes the analysis provides very good evidence for the magnitude of one source of a discrepancy at a site but still leaves a large residual error.

Aggregation and Allocation

After the site work and the discrepancy tool have been completed and QC'd for each site, the data from the sheets is uploaded to a single data set. The analytical component of the analysis is straightforward. The individual discrepancies and residual errors are weighted up using the site case weights. Pivot tables are then used to aggregate the discrepancy impacts for positive, negative, and net discrepancy impacts. The expectation is that the residual will net out to zero and can be ignored. Typically, several iterations ensue, where a senior engineer or analyst will examine the overall results, note trends, and go back to individual sites to refine the site results and re-aggregate them to program-level values.

It is in this iterative process that the evaluator can bring additional value to the implementer and where the allocation method shines because of its adaptability and flexibility. By examining the data for trends, the analyst may identify categories that might benefit from being split into subcategories (for example, splitting hours of operation discrepancies into prescriptive and custom hours of operation) or combined with other categories. Once the senior analyst is satisfied that the site results properly characterize the discrepancies and that the site residual error has been minimized, the final allocation method discrepancy rate is compared to the program discrepancy rate. By the nature of the method, which is based on a compilation of independently calculated factors that do not account for interactivity, the rates are not likely to be equal. As a final step, the magnitudes of the program level discrepancy categories are equally scaled to match the program discrepancy rate. This scaling factor has typically been in the order of a 5% adjustment in the experience of the writer.

Conclusions

A program realization rate alone provides frustratingly little actionable feedback to a program implementer or any insights into how program savings estimates can be improved. At the same time, valuable program performance data has been collected as part of the M&V work, ready to be tapped and analyzed with the right tools. Analysis and synthesis of this data can yield quantitative findings that can guide an implementer’s next steps.

It is all too common in the evaluation community, to conclude a complex program evaluation with site-by-site narrative descriptions of discrepancies or some other list-oriented summary of what went wrong. Lists of discrepancies, however, leave the implementer struggling to identify trends and priorities when that should be our job as evaluators, as we have the knowledge of the sites and the analytical skills to compile the information into actionable findings. The cascade method of analyzing discrepancies locks the analysis in to a few fixed discrepancy categories and an inflexible framework allowing the codependent impacts to be calculated in a fixed order. The results can be misleading as well, since the first discrepancy in the calculations will appear to have the most impact due to interactive effects.

The allocation method calculates each discrepancy independently, rather than in a cascaded fashion. This structure does not limit the number of discrepancy categories and has the flexibility to easily combine or split categories with little additional work. This method systematically extracts as much value as possible from the site findings and systematically combines them into a coherent and quantitative picture of the factors driving a program’s realization rate. The final results can be organized in a manner that suits the program implementer, whether ordered from largest to smallest magnitude or, as suggested in the first section, by the implementation cycle.

Regardless of the exact technique used by the evaluator, we recommend planning for reporting the discrepancies from the start of an evaluation to build the data collection tools necessary to capture the quantitative data from the details of the site work. This work should not be an afterthought; it is after all, the starting point for a program implementer intent on using the results of an impact evaluation for program improvement. This allocation method presented here is a useful model for organizing discrepancy data into a meaningful guide to what the implementers might do next.

Additional examples of studies using list, cascaded, and allocation methods are included in the “References” section.

References

Studies with List-Oriented Discrepancy Results

SBW Consulting, Inc. *Final Report 2006–2008 Retro-commissioning Impact Evaluation* prepared for California Public Utilities Commission Energy Division. February 8, 2010.

KEMA, Inc. and DMI, Inc. *Impact Evaluation of 2011–2012 Prescriptive VSDs* prepared for Massachusetts Program Administrators and Massachusetts Energy Efficiency Advisory Council, May 9, 2013.

Summit Blue Consulting. *Government Partnerships Programs Direct Impact Evaluation Report* prepared for the California Public Utilities Commission Energy Division. February 8, 2010.

Studies with Cascaded Discrepancy Results

ERS. *Con Edison EEPS Programs SBDI Final Impact Evaluation Report*, July 07, 2014.

KEMA, Inc. *Impact Evaluation of the Massachusetts Upstream Lighting Program Final Report* prepared for the Massachusetts Energy Efficiency Program Administrators and the Massachusetts Energy Efficiency Advisory Council. February 19, 2014.

ERS. *NYSERDA EmPower Program and National Fuel Gas Distribution Corporation's Low Income Usage Reduction Program Impact Evaluation (2010–2011)* prepared for the New York State Energy Research and Development Authority and National Fuel Gas Distribution Corporation.

Studies with Allocation Discrepancy Results

Itron, Inc. *2010–2012 WO033 Custom Impact Evaluation Final Report* prepared for California Public Utilities Commission. July 14, 2014 (details are of method are unknown to the author).

ERS. *Multifamily Performance Program Impact Evaluation (2009–2011)* prepared for the New York State Energy Research and Development Authority. February 2015.