A Common Evaluation Framework for Comparison of Smart Grid Program Impacts Across Utilities

Stuart Schare, Navigant Consulting, Boulder, CO
Mike Sherman, Navigant Consulting, Burlington, MA

ABSTRACT

A variety of approaches have emerged in recent years to evaluate the impacts and cost-effectiveness of smart grid investments. However, if each utility conducting a smart grid pilot program selects its preferred evaluation approach, then comparison of results across programs can be difficult at best and misleading at worst. Utilities in Massachusetts are attempting to address this limitation by participating in a collaborative evaluation effort that may help determine which pilot structure has the greatest impact or provides the best return on investment. The Common Evaluation Framework describes an approach to impact analysis that each Massachusetts utility is expected to incorporate into its smart grid pilot evaluation. In order for a common approach to enable comparison across pilots, the evaluations must utilize common terminology and data formats and address consistent research objectives. This paper discusses the common data and methods established for the evaluation of pilot impacts across four unique pilots in Massachusetts, and it addresses the challenges and limitations of the collaborative evaluation approach.

Introduction

A variety of approaches have emerged in recent years to evaluate the impacts and cost-effectiveness of smart grid investments, including those in support of the U.S. Department of Energy's Smart Grid Investment Grant Program and the Bonneville Power Administration's regional smart grid deployment (Small, 2010; Gilbert, 2010). However, if each utility conducting a smart grid pilot program selects its preferred evaluation approach, then comparison of results across programs can be difficult at best and misleading at worst. Utilities in Massachusetts are attempting to address this limitation by participating in a collaborative evaluation effort that may help determine which pilot structure has the greatest impact or provides the best return on investment.

The Green Communities Act (GCA) required each distribution utility in the state to establish a smart grid pilot program utilizing "advanced ("smart") meters that provide real time measurement and communication of energy consumption [and] automated load management systems." Furthermore, the companies were to file proposals with the Department of Public Utilities (DPU) requiring "time of use [TOU] or hourly pricing for commodity service for a minimum of 0.25 per cent of the company's customers" (Commonwealth of Massachusetts, 2008).[1]

These plans described the components of the individual program pilots, including the underlying smart grid technologies, alternative dynamic rate structures, and the provision of information to customers through in-home displays and web portals. Among the three largest utilities in the state, the smart grid programs are each testing a variety of alternative technology and rate packages for residential customers, and each pilot is unique in its treatment of customer recruitment, dynamic rates, and load

[1] The purpose of the GCA is to provide for renewable and alternative energy and energy efficiency in the Commonwealth. A specific goal of the utility pilots is to reduce participants' "peak and average loads by a minimum of 5 percent." The Act does not specify which customer segments are to be included, but the common interpretation is that the 0.25 percent figure applies to the number of residential customers in each utility service territory.

control technology. Key highlights of the NSTAR, National Grid, and Unitil pilots, including their unique test groups, are provided below, and a detailed comparison of the smart grid offerings are provided in Tables 1a and 1b (Collaborative, 2011, Appendix A). For purposes of the Collaborative (discussed below), a "test group" is defined as "a segment of participants who are provided the same combination of enabling technology and rate structure."

- **NSTAR** – Participants are each provided with in-home energy displays (IHD) and connected to the utility via a home area network (HAN) and broadband gateway. The four unique test groups include TOU rates with critical peak pricing (CPP), another that also has smart thermostats for direct load control, a third offered a critical peak rebate (PTR), and a fourth with the IHD only.

- **National Grid** – The pilot includes eight test groups, defined by various combinations of in-home technologies, CPP rates, and PTRs.[2]

- **Unitil** – The pilot consisting of three test groups offered TOU rates, in-home technologies, and smart thermostats.

In its approval of the NSTAR pilot, DPU noted that "to fully leverage the valuable information that will come from each individual pilot, it is important that the pilots be conducted and evaluated in a coordinated manner." Consequently, DPU required establishment of "uniform statewide evaluation approaches and standards [to form] a framework that provides for the ability to compare results across all pilot programs proposed in the state" (MA DPU, 2010). The utilities and other stakeholders subsequently formed a collaborative committee consisting of representatives from each of the four Massachusetts electric utilities, the MA Attorney General's Office, the MA DPU, the MA Department of Energy Resources (MA DOER), and the Low Income Energy Assistance Network (LEAN). The Collaborative's broad objective was to develop a Common Evaluation Framework that would ensure consistency in the way that Smart Grid Pilot evaluations are conducted and thus facilitate comparability between the various smart grid pilots. As of spring 2011, both NSTAR and Unitil had begun recruitment of customers and were incorporating collaborative approaches into their survey work and evaluation planning.

The Collaborative structure and rules of procedure were not specified by the DPU, and the process is informal, with no official rules for voting among participants. The utilities generally consider the perspectives and recommendations of the other stakeholders, but the utilities have the ultimate discretion to determine their preferred evaluation approaches, subject to possible regulatory action by the DPU if Collaborative participants do not believe that the utilities are sufficiently incorporating stakeholder views.

Objectives of the Collaborative

The Collaborative's mission is to ensure that the quantitative and qualitative results of each utility Pilot are conducted and presented in a consistent manner wherever applicable, so as to enable reasonable comparisons across Pilots and more readily share what is learned among all parties. This common evaluation approach will help ensure that differences in results between the various pilots can be attributed to differences in the offerings rather than the analytic methodology. From the perspective of the DPU, the Collaborative's work is helping to maximize the value at a statewide level from the three similar, yet distinct, smart grid pilots being launched in 2011.

[2] As of April 2011, National Grid had withdrawn its pilot proposal and was not actively pursuing recruitment of customers.

Table 1a. Component Overview of Utility Smart Grid Pilot Programs and Test Groups

Company		Test Group	Eligibility*				Enabling Technology*						
			Geography		Form of Participation	Treatment of AC	Treatment of Broadband Internet**	Home Display	Web- Portal	Load Control	*Programmable Thermostat*	*Central Air Loads Control*	*Other Load Control Switch*	Other
								Access to technology is . . .						
NSTAR	1	Time of Use (TOU) and Critical Peak Price (CPP)	Residential - Two communities, plus a 3rd for low-income		Opt-in	Allowed	Required	Required	Required	Disallowed	-	-	-	Internet Gateway & ERT Bridge
	2	TOU + CPP w/ PCT or LCS				Required	Required	Required	Required	Required	*Allowed*	*Allowed*	*Disallowed*	Internet Gateway & ERT Bridge
	3	Critical Peak Rebate				Required	Required	Required	Required	Required	*Allowed*	*Allowed*	*Disallowed*	Internet Gateway & ERT Bridge
	4	Technology Only				Allowed	Required	Required	Required	Disallowed	-	-	-	Internet Gateway & ERT Bridge
National Grid	1	Critical Peak Price (CPP): no In-Home Tech.	Single Worcester community	Residential, Small C&I, Medium C&I	Opt-out	Allowed	Allowed	Disallowed	Required	Disallowed	-	-	-	-
	2	CPP: Level 1 In-Home Tech.					Allowed	Required	Required	Disallowed	-	-	-	-
	3	CPP: Level 2 In-Home Tech.					Allowed	Required	Required	Required	*Allowed*	*Allowed*	*Allowed*	-
	4	CPP: Level 3 In-Home Tech.					Required	Required	Required	Required	*Allowed*	*Allowed*	*Allowed*	Web/Mobile Device Tools Guaranteed
	5	Peak Time Rebate (PTR): no In-Home Tech.		All customers who opt out of CPPP or Hourly Pricing	Other		Allowed	Disallowed	Required	Disallowed	-	-	-	-
	6	PTR: Level 1 In-Home Tech.					Allowed	Required	Required	Disallowed	-	-	-	-
	7	PTR: Level 2 In-Home Tech.					Allowed	Required	Required	Required	*Allowed*	*Allowed*	*Allowed*	-
	8	PTR: Level 3 In-Home Tech.					Required	Required	Required	Required	*Allowed*	*Allowed*	*Allowed*	Web/Mobile Device Tools Guaranteed
Unitil	1	Simple Time of Use (TOU)	Residential - MA & NH		Opt-in	Required	Required	Disallowed	Required	Disallowed	-	-	-	-
	2	Enhanced Technology				Required	Required	Required	Required	Required	*Allowed*	*Allowed*	*Allowed*	Internet Gateway & ERT Bridge
	3	Smart Thermostat				Required	Required	Other	Required	Required	*Required*	*Required*	*Disallowed*	Paging network

* *Allowed* refers to test groups in which some participants may have the equipment or technology in question, but for which the equipment/technology is not required for participation.

** Two-way communications via broadband internet is required for some test groups and is optional for others.

Table 1b. Test Group Rates Structures

Company		Test Group	Rate Structure			
			Time of Use	Critical Peak Price	Peak Time Rebate	Other
NSTAR	1	Time of Use (TOU) and Critical Peak Price (CPP)	Yes	Yes	No	No
	2	TOU + CPP w/ PCT or LCS	Yes	Yes	No	No
	3	Critical Peak Rebate	No	No	Yes	No
	4	Technology Only	No	No	No	No
National Grid	1	Critical Peak Price (CPP)[3]: no In-Home Tech.	**Yes**	**Yes**	**No**	**No**
	2	CPP[4]: Level 1 In-Home Tech.				
	3	CPP[4]: Level 2 In-Home Tech.[4]				
	4	CPP[4]: Level 3 In-Home Tech.[4]				
	5	Peak Time Rebate (PTR): no In-Home Tech.	No	No	Yes	No
	6	PTR: Level 1 In-Home Tech.				
	7	PTR: Level 2 In-Home Tech.				
	8	PTR: Level 3 In-Home Tech.				
Unitil	1	Simple Time of Use (TOU)	Yes	Yes	No	No
	2	Enhanced Technology	Yes	Yes	No	No
	3	Smart Thermostat[7]	No	No	No	Yes

In developing a Common Evaluation Framework, the Collaborative established specific objectives for development of Framework products including the following (Collaborative, 2011):

1. **Shared research objectives** that define what the stakeholders hope to learn from the Smart Grid Pilots. These objectives address energy and peak load impacts, participant bill impacts, marketing, education, and participant satisfaction/interest/behavior.

2. **A pilot summary table** that defines the unique characteristics of each Pilot and that serves as a reference for comparison of each pilot and unique pilot segment with respect to eligibility/recruitment, technology provided, and rate structure. See Table 1 above.

3. **Common definitions** to be used by each utility and included in their evaluation plans. Because the infant concept of the smart grid lacks consistent vocabulary, the Framework established a common language to facilitate understanding and consistency across all parties.

4. **Guidelines for the type and format of data** for interval metering data, customer demographics, and household characteristics that should be collected for all Pilot participants. It is expected that each utility will collect additional customer data as they deem appropriate, but a minimum set of consistent data will facilitate post-pilot, cross-utility comparisons. For example, impact and process findings for each utility will be segmented according to a specific list of customer demographics which have been pre-defined to ensure that each pilot uses the same divisions for income groups, age divisions, home sizes, etc.

5. **Consistent approaches to collecting and analyzing *qualitative* results of the Pilot**. For example, survey questions intended to assess customer behavior and satisfaction with the program should be consistent across pilots to reduce bias that could be introduced when questions are phrased in different ways. Furthermore, each pilot is expected to conduct surveys at each of six distinct points in time, including pre-pilot, post-installation, after critical events, at the end of the pilot, when customers decline an invitation to participate, and when participants drop of out the pilot.

6. **Consistent approaches to analyzing the *quantitative* results of the Pilot**. While the Collaborative acknowledges that each evaluation will utilize unique statistical models based on the specific combinations of dynamic rates and technologies, minimum expectations have been established to facilitate cross-utility comparisons of quantitative pilot results.

7. **Outline for pilot evaluation reports** such that each report utilizes the same general structure and content.

The Framework document addresses each of the above items except for the report outline, which will be developed during future efforts of the Collaborative. At the heart of document is the approach for impact analysis, which draws on many of the other elements of the Framework. Highlights of this approach are discussed below.

Common Evaluation Framework for Impact Analysis

The Common Evaluation Framework describes an approach to impact analysis that each Massachusetts utility is expected to incorporate into its smart grid pilot evaluation. In order for a common approach to enable comparison across pilot, the evaluations must utilize common terminology and data formats and address consistent research objectives. The Framework document discusses these items in detail, but this paper focuses on the data and methods established for the evaluation of pilot impacts.

Impact Metrics

The impact evaluations are intended to identify the reductions in overall energy consumption and in peak demand resulting from various technologies, dynamic rates, and educational initiatives. The Framework calls for each company to estimate and present results for each of the five impact metrics defined in Table 2, as applicable, using the jointly established definitions of key terms.

Table 2. Common Metrics for Estimating Energy and Peak Demand Impacts of Smart Grid Pilots

Impact Metric	Description
Overall reduction in Energy Usage	Measured in kWh, this metric estimates the amount of a participant's reduction in energy consumption during certain pre-defined time periods
Demand reduction during Peak Periods	This metric is the average hourly reduction in energy consumption (kW) during a Peak Period, as defined for each pilot, when a utility typically experiences the highest demand for energy
Demand reduction during Critical Events	Measure of the demand reduction during a Critical Event, defined as "a period of time when the utility has notified participants of high system demand, [and] the participant may be subject to steep price differentials during a Critical Event, be eligible for a Peak Time Rebate, or be subject to direct load control of their air conditioning system"
Peak Demand Reduction during each of the hours of a Critical Event	Measures the demand reduction over time during a Critical Event, rather than one data point for the entire event
Change in energy use for the three hours immediately before and after a Critical Event	Expected to provide an estimate of the amount of energy shifted off peak due to either the "bounce-back effect" immediately after an event or "pre-cooling" immediately prior to an event

For the first three metrics, results will be presented for each of test groups in terms of the average of all participants within the group. Reductions in Energy Usage will be estimated both in kWh and as a percentage of consumption for the summer period (June to September), winter period (October to May), summer and winter peak periods (defined by each utility), and for the total pilot. Demand reductions during peak periods and critical events will be estimated in kW and the share of peak demand for the summer and winter period (Table 3)

Table 3. Demand Response Impact Metrics

	Overal Reduction									
	Summer		Winter		Summer Peak Hours		Winter Peak Hours		Total Pilot	
# of participants	%	kWh	%	kWh	%	kWh	%	kWh	%	kWh
#	%	kWh	%	kWh	%	kWh	%	kWh	%	kWh

Source: Collaborative, 2011, Appendix A

Results will also be broken out by key segmentations of the participants, as derived from the survey data. This segmentation is expected to include the following classifications of participants: a) all participants; b) low-income; c) high-income; d) low-use; e) high-use; f)low-income/low-use; g) low-income/high-use; h) participants with senior residents; i) small home; and j) large home. Each utility will collect demographic data in a consistent format defined by the Collaborative and will use common definitions of the above participant classifications such that comparisons of segmentation results across pilots will reflect differences in program structure and delivery rather than differences in definitions of terms.

In addition to evaluation of energy and demand impacts, the Framework calls for evaluation of participant bill impacts. Actual participant bills will be analyzed relative to the amount the participants would have been charged, absent their participation in the Pilot, for the year, as well as by month.

Impact Analysis Methodology

According to the Framework, energy and peak demand impact should be assessed using a regression model customized to the unique pilot and test group attributes of the participants being evaluated. Where appropriate, fixed effects regression modeling will be applied. With this modeling framework, all of the individual data are consolidated into a single panel (or longitudinal) data-set that is both cross-sectional (including many different individuals) and time-series (repeated observations for each individual).

Additional time-series variables will be included in each regression to control for variations in ambient temperature, weather, and whether a day is a weekend, holiday or weekday. The inclusion of weather and temperature variables implicitly performs weather normalization and obviates the need for explicit adjustments to the data to account for weather impacts. Essentially, the regression controls for weather effects and allows the analyst to forecast the effect that weather changes will have on the variable of interest (i.e., electricity consumption, income, age, or type of dwelling).

In theory, different evaluators could use the data from a given pilot and produce the same set of impacts using the prescribed methods. In practice, small variations in equally appropriate methods for cleaning the billing data, normalizing for weather, and specifying the regression model for a pilot's unique test groups would likely result in impact findings that are similar, but not identical. The impact analysis methodology described in the Framework is intended to be specific enough to ensure consistency of approach across pilots while remaining flexible enough to allow each utility to perform the analysis most appropriate to their offerings.

Challenges to Collaboration and Comparability

The stakeholders contributing to the collaborative evaluation process in Massachusetts have succeeded in defining a common analytic approach and common metrics for measuring and reporting

smart grid program impacts. However, the collaborative process has also presented challenges for program implementation and revealed limitations of the collaborative approach.

Implementation challenges include the following:

- **Unique utility interests.** Each utility brings to the Collaborative its own interests and preferences regarding a variety of topics from impact evaluation methodologies to the questions posed on participant surveys. The parties have proven remarkably flexible in accommodating one another's requests, but divergent opinions could pose difficult challenges for similar efforts elsewhere.

- **Non-coincident implementation timeframes.** At the time of the early Collaborative efforts, some utilities were approaching the participant recruitment phase, while others were still testing critical system integration functionality. If all pilots were to utilize the same base set of survey questions, then the guidelines for the pre-pilot survey had to be completed prior to any utility commencing operation of the pilot technologies or rates. As a consequence, the launch of one of the pilots was nearly delayed to accommodate the iterative nature of the multi-stakeholder process.

- **Accommodation of interest group concerns.** Interest groups have legitimate concerns, but as individual stakeholders in a multi-party process, they cannot always achieve all of their objectives. To the extent that the utilities are accommodating these interests relative to their own base proposals, the collaboration process forces interest groups to prioritize their concerns. For example, low income customers will be evaluated as a distinct subset of the participants in order to assess the impacts on and satisfaction of this specific group of customers; however, Spanish language surveys will not be required.

- **Length of participant surveys.** In order to accommodate the diverse interests alluded to above, the Collaborative tends to add survey questions without necessarily prioritizing to limit the number of questions. As a result, surveys may prove to be cumbersome to administer and may produce data that not all utilities will utilize in their evaluations.

The Collaborative approach may also in limitations to the applicability of some of the evaluation findings, including the following:

- **Demographic breakouts may not provide significant results,** leaving impacts on some groups uncertain. For example, the Collaborative agreed to document and perform cross-tabulations and impact analysis on a variety of demographic subgroups such as low income renters and households with elderly residents. However, without requirements for a sample size large enough to accommodate analysis of these customer segments, the analysis may produce inconclusive results that add little to the collective knowledge.

- **Differences in program designs may limit comparability of specific program attributes.** The appearance of commonality in evaluation of the various programs cannot ensure that alternative program components can be compared. For example, if two pilots utilize a different set of different prices for peak and off-peak rates, a simple comparison of energy consumption impacts using the common evaluation framework might suggest that differences in impacts are

the result of the different sets of prices. However, the differing impacts may well be driven by other variations between the programs that are not captured in the regression equations, such as the length of the peak period or the manner in which customers are recruited to the pilot.

Thus far the Collaborative participants have been able to reach consensus on all issues despite the challenges discussed above. For example, stakeholders were willing to accelerate decisions regarding pre-pilot survey questions in order to accommodate one utility that was to begin recruitment. The stakeholders have largely deferred resolution of the challenges for applicability and comparability of findings, but the Collaborative participants generally accept that the process is likely to provide some, but not all, of the desired outcomes.

Conclusions

The Collaborative framework promises to lend a minimum level of comparability in findings across the Massachusetts' utilities smart grid pilot programs. By ensuring common terminology analytic approaches, and evaluation metrics, framework has limited the number of variants across programs that could contribute to differences in findings. However, there is no guarantee that the pilot evaluations will produce reliable indicators of what might happen in a large-scale deployment or that differences in findings for a given parameter (e.g., peak prices across two pilots) are actually attributable to the apparent differences in the value of the parameter itself.

Some important issues are not discussed explicitly in the framework, such as how to ensure a representative sample that produces results that can be extrapolated to the customer population. This is a difficult goal to achieve, as the utilities needed to limit geography to keep costs and logistics manageable, and they do not all recruit customers in an identical manner. National Grid, for example, was pursuing an opt-out recruitment approach for a limited geographic area, while NSTAR is recruiting only customers who actively choose to enroll. Perhaps any differences in impact can be attributable in part to the alternative recruitment approaches. But these are not two otherwise identical pilots; technology and rates are different, so there are multiple parameters that could explain any measured differences in impacts. The only way to avoid this uncertainty would have been to do a centrally planned experimental design across all of the Massachusetts utilities. However, this was not called for by the GCA and would have required an even greater level of coordination and collaboration on the part of the utilities and the other stakeholders. This top-down approach would have been fundamentally at odds with the notion that each company design a plan that they believe has the greatest chance of success given their corporate strategies and customer bases.

Future activities for the Collaborative include developing additional survey instruments, such as for obtaining post-pilot customer feedback, and reviewing initial evaluation results for thoroughness and for their implications for state energy efficiency and smart grid policy. While the Collaborative is not a panacea for the challenges of program evaluation, it has established a number of guidelines that are expected to produce a measure of consistency and comparability across programs. Notably, the common terminology, data formats, analytic methods, and impact metrics are a significant step toward the state's goal of learning from the various smart grid pilots to develop a long-term policy for leveraging new technology to make better use of the state's energy resources and electrical infrastructure.

References

[Collaborative 2011] Massachusetts Smart Grid Collaborative Technical Subcommittee, Massachusetts Smart Grid Collaborative "Common Evaluation Framework," unpublished, March 2011.

Commonwealth of Massachusetts, *An Act Relative to Green Communities, Senate No. 2768,* approved by the Governor, July 2, 2008.
http://www.malegislature.gov/Laws/SessionLaws/Acts/2008/Chapter169

Department of Public Utilities, Commonwealth of Massachusetts, Order in D.P.U 09-33, RE: Petition of NSTAR Electric Company for approval pursuant to St. 2008, c. 169, § 85 of a Smart Grid Pilot Program, March 15, 2010.

[EPRI 2010] Electric Power Research Institute, *Methodological Approach for Estimating the Benefits and Costs of Smart Grid Demonstration Projects,* Final Report, January 2010.

Gilbert, Erik *et al*, "The Role of Smart Grid in Enhancing Energy Efficiency and Demand Response: An Economic Assessment of a Regional Smart Grid Deployment in the Pacific Northwest," *ACEEE Summer Study*, July, 2010.

NSTAR Electric Company, *NSTAR Smart Grid Pilot Plan Filing,* Exhibit NSTAR-1, filed in Docket 09-33, March 31, 2009.

Acknowledgements

The authors wish to acknowledge the contributions of many individuals from the Massachusetts investor-owned utilities and their consultants, the stakeholder organizations, and the DPU. The following individuals and organizations have provided particularly significant input to the collaborative process and to the Framework document:

- Matt Siska, GDS Associates
- NSTAR Electric Company
- National Grid
- Unitil
- Western Massachusetts Electric Company
- Attorney General of Massachusetts
- Massachusetts Department of Energy Resources
- Massachusetts Department of Public Utilities
- Low Income Weatherization and Fuel Assistance Program Network