Final Report

Impact Evaluation of New York Small Business Services Energy Efficiency Program, 2010 and 2011 Program Years

National Grid USA
Prepared by KEMA, Inc.
September, 2014

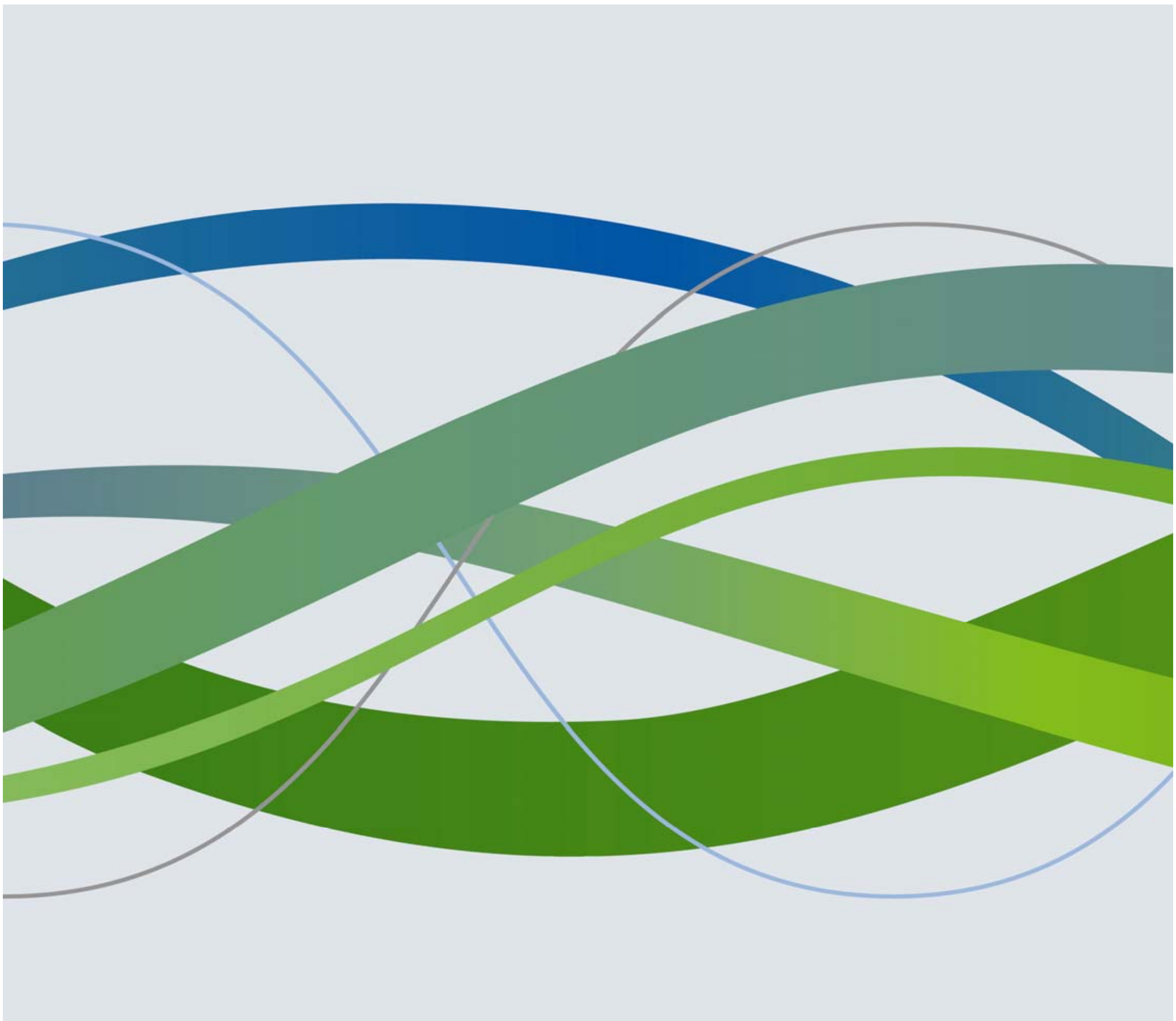


Table of Contents

In 2012 KEMA became part of Det Norske Veritas (DNV), a global provider of services for managing risk. In 2013, a merger between DNV and Germanischer Lloyd (GL) took place. Driven by our purpose of safeguarding life, property and the environment, DNV GL enables organizations to advance the safety and sustainability of their business. Operating in more than 100 countries, our 16,000 professionals are dedicated to helping our customers in the maritime, oil & gas, energy and other industries to make the world safer, smarter and greener.

In the Energy industry, DNV GL delivers world-renowned testing and advisory services to the energy value chain including renewables. Its expertise spans onshore and offshore wind power, solar, conventional generation, transmission and distribution, smart grids, and sustainable energy use, as well as energy markets and regulations. DNV GL's 3000 energy experts support clients around the globe in delivering a safe, reliable, efficient, and sustainable energy supply. For more information please visit www.dnvgl.com/energy.

Table of Contents

Table of Contents

List of Exhibits

Table of Contents

1. Executive Summary

This document summarizes the results of an evaluation of National Grid's 2010 and 2011 Small Business Services Energy Efficiency Program ("SBS") in the Niagara Mohawk territory. The SBS Program provides direct installation of energy-efficient measures for small commercial and industrial ("C&I") customers (defined as customers with an average monthly demand of less than 100 kW).

The primary objective of this evaluation is to quantify the gross and net annual energy and summer demand impacts of lighting measures installed through the SBS program. The savings and factors of interest to the study includes the coincident summer on-peak factor, connected kW savings and realization rate, kWh savings and realization rate, percent on peak kWh, summer demand HVAC interactive effect factor and kWh HVAC interactive effect factor. The study was designed to provide two primary means of assessing gross impacts and an additional task to assess net-to-gross ("NTG") impacts. All gross energy savings and net factor sampling work was designed to achieve ±10.0% at the 90% confidence level.

1.1 2010 and 2011 Program Activity Summary

Table ES 1 below provides a summary of the 2010 and 2011 SBS Program population. This summary breaks out savings for prescriptive lighting with and without controls, other lighting, and all other measures. Combined, the two years of activity tracked 245,219MWh of annual energy savings, accumulated from 111,936 MWh in 2010 and 133,283MWh in 2011.

We have grouped the prescriptive lighting[1] with and without controls as individual line items to show the relationship between the 2010 sample frame[2] for the Measurement &Verification ("M&V") study to the overall population. Overall, the combined lighting measures and controls savings are the focus of a majority of program savings, with 93% of total savings installed at 92% of accounts across years. The prominence of lighting in this program was a primary consideration in deciding to focus on lighting in much of the evaluation work undertaken on the SBS Program. It should be noted that this study focuses on lighting without controls, although these results have a direct bearing on all prescriptive lighting, including those installed with controls. As such, these results will impact the savings estimates of the vast majority of program savings if incorporated into the New York Technical Resource Manual ("TRM") and/or if used to refine tracking savings estimates by National Grid. National Grid is initiating a study on lighting controls at this time that will round out the study of savings associated with lighting. This pending controls study also plays a key role in executing some of the recommendations from this study.

[1] Prescriptive lighting measures were identified in the Measure Description field as being either CFL, Fluorescent, Custom Lighting, HPS, LED or PSMH.

[2] The sample frame for the M&V site work is the prescriptive lighting without controls activity in 2010, which is 64,265 MWH, after removing control savings installed in 2011 on 2010 projects.

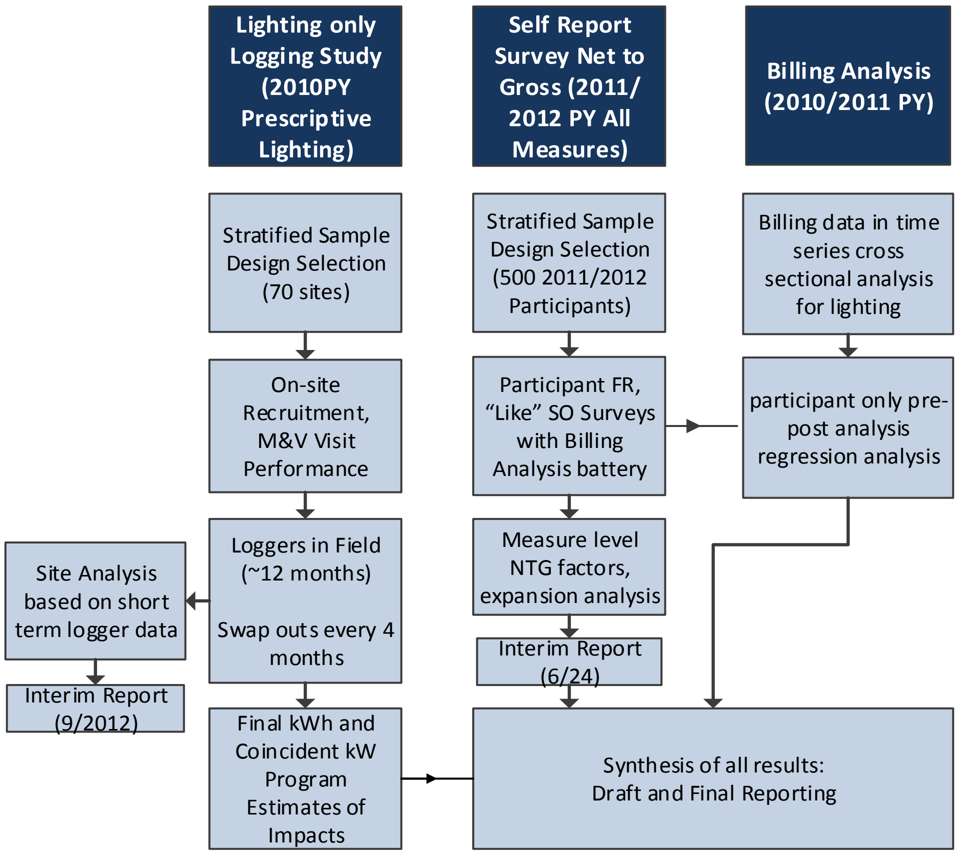
Table ES 1: 2010 and 2011 SBS Program Activity

End Use	kWh Savings	
	kWh	%
2010 Program Year		
Prescriptive Lighting without Controls	64,498,471	57.6%
Prescriptive Lighting with Controls	36,946,355	33.0%
Other Lighting	1,659,890	1.5%
HVAC, Motors, Refrigeration	8,831,028	7.9%
Total	111,935,744	100.0%
2011 Program Year		
Prescriptive Lighting without Controls	85,525,632	64.2%
Prescriptive Lighting with Controls	39,845,725	29.90%
Other Lighting	13,887	0.0%
HVAC, Motors, Refrigeration	7,897,597	5.9%
Total	133,282,841	100.0%
Total of Both Years		
Prescriptive Lighting without Controls	150,024,103	61.6%
Prescriptive Lighting with Controls	76,792,080	30.60%
Other Lighting	1,673,777	1.0%
HVAC, Motors, Refrigeration	16,728,625	6.8%
Total	245,218,585	100.0%

1.2 Key Study Activities

Figure ES 1 below presents the overall evaluation plan, which includes the three primary evaluation activities. These are an M&V logger study on 2010 prescriptive only lighting participants, an assessment of NTG factors from 2011 through July of 2012 participants, and a two stage billing analysis on lighting savings among 2010 and 2011 participants. Although the gross and net impacts examined are not consistently from the same set of program years, the technical manual calculation driving the savings estimates has been consistent during this time and program changes in SBS have been minimal from 2010 through 2012. This helps assure that gross savings estimates and realization rates from disparate years (e.g., 2010 vs. 2011), remain comparable. The study design was also developed to keep savings estimates from both the billing analysis and M&V defined as gross, which helps assist in their use as independent efforts that can be combined in a triangulating fashion.

Figure ES 1: Overall Evaluation Plan



The logging study was performed at a statistically selected sample of 70 on-site visits with a year of time of use lighting logging and included an interim results deliverable after the first 4 months of metering. The billing analysis utilized weather normalized participant billing data in a pre/post analysis in a series of models to explore the most appropriate model for estimating gross program savings that would be comparable to the M&V based gross savings estimates. The NTG effects were determined through the use of 484 SBS program participant surveys that covered 1,055 program installed measures. As part of the NTG survey, questions were asked that were designed to better understand the billing analysis results as well as for incorporation into the billing analysis model.

1.3 Key Study Results

There are three primary sets of results presented here. These include the results from the On-site M&V and re-engineering work, the billing analysis, and net to gross assessment.

On-site M&V

Table ES 2 summarizes the energy savings results of the on-site M&V analysis for the SBS Program. Results are presented by adjustment factors to demonstrate which computational inputs were influential in the difference between the tracking system estimate of savings and our gross on-site estimate. These factors are further described in Section 3.1.This table presents these results by nature of discrepancy in two ways. The first is by providing the incremental ratio, which shows the percent difference between each level of adjustment. The second is cumulative ratio, which shows the percent difference between the each level of adjustment and the tracking estimate. Showing the results in this manner allows one to see the relative change from one adjustment to another as well as the overall accumulated change relative to the tracking estimate.

The annual energy savings gross realization rate for prescriptive lighting without controls was found to be 77.2% without HVAC interactive effects included and 80.3% with interactive. The relative precisions for the final impact estimate with interactive is ±9.1%. The primary driver of the realization rate was the decrease in operating hours observed through the on-site M&V, which caused a 15% drop in energy savings relative to that captured in the previous quantity adjustment.

Table ES 2: Summary of 2010 PY Prescriptive Lighting Gross Energy Results and Realization Rates

Prescriptive Lighting Without Controls (n=70)	kWh	Incremental Ratio		Cumulative Ratio	
		Value	Relative Precision at 90% Confidence (±)	Value	Relative Precision at 90% Confidence (±)
Tracking	64,265,935	N/A	N/A	N/A	N/A
Documentation Adjustment	64,187,086	99.9%	0.6%	99.9%	0.6%
Technology Adjustment	63,873,139	99.5%	0.5%	99.4%	0.7%
Quantity Adjustment	61,163,048	95.8%	1.7%	95.2%	1.8%
Operational Adjustment	49,591,059	81.1%	8.9%	77.2%	9.0%
HVAC Interactive Adjustment	51,594,070	104.0%	1.1%	80.3%	9.1%
Gross Realization Rate	51,594,070	80.3%	9.1%	80.3%	9.1%

Table ES 3 is structured the same as Table ES 2, but provides results specific to demand impacts. The connected kW gross realization rate is 95.7% with a relative precision of ±2.0% at the 90% confidence level. The realization rate was primarily affected by lower quantities of installed lighting (observed on-site) than what was recorded in the tracking system. However, we note that this adjustment was still minor in nature (<4%) and occurred over three years.

Table ES 3: Summary of Prescriptive Lighting Gross Demand Results and Realization Rates

Prescriptive Lighting Without Controls (n=70)	kW	Incremental Ratio		Cumulative Ratio	
		Value	Relative Precision at 90% Confidence (±)	Value	Relative Precision at 90% Confidence (±)
Tracking	18,437	N/A	N/A	N/A	N/A
Documentation Adjustment	18,370	99.6%	0.8%	99.6%	0.8%
Technology Adjustment	18,340	99.8%	0.7%	99.5%	1.0%
Quantity Adjustment	17,639	96.2%	1.7%	95.7%	2.0%
Gross Realization Rate	**17,639**	**95.7%**	**2.0%**	**95.7%**	**2.0%**

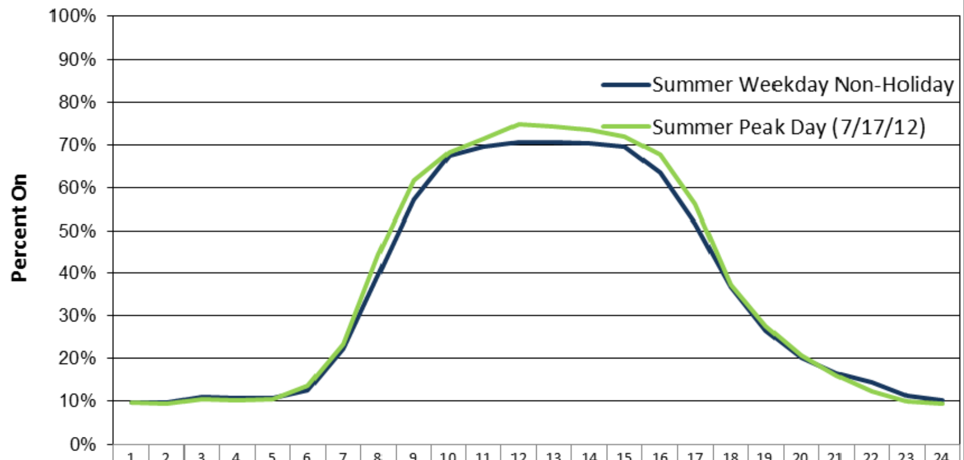
Table ES 4 shows the remaining factors of interest from the on-site M&V work for prescriptive lighting measures. The estimated coincidence factor from this study is 0.566, which compares to a factor of 1.0 from the TRM. The evaluated values for the kW HVAC interactive effect factor are 1.104 for summer kW and 1.04 for annual kWh. These reflect the percent of gross summer peak kW and gross kWh savings that are due to interactive effects. The final three rows show the average weighted hours-of-use from the lighting loggers, the % on peak energy savings and the non-electric HVAC interactive effect in MMBtus per kWh saved.

Table ES 4: Other Savings Factors

Other Ratio Estimation Results (n=70)	Value	90% Confidence Interval (+/-)
Summer kW Factors		
Coincidence Factor	56.6%	15.3%
HVAC Interactive Effect Factor	110.4%	2.7%
KWh Factors		
HVAC Interactive Effect Factor	104.0%	1.1%
Average Hours of Use	2,708	329
% On Peak KWh	72.0%	6.1%
Non-Electric		
Heating HVAC Interaction Effect (MMBtu/kWh)	-0.0023	0.0020

Figure ES 2 presents the weighted percent of on time for non-holiday summer weekdays and for the NY peak day in 2012 (July 17). The summer peak day profile represents lighting operation during system peak and is largely the same as the regular summer weekday profile, primarily due to lighting not being a weather sensitive measure. The 56% noted in hour 17:00 (5 p.m.) is our coincident factor estimate based upon the metering performed in this study.

Figure ES 2: Summer Weekday and Peak Day Lighting Profiles



	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24
Summer Weekday Non-Holiday	10%	9%	11%	11%	11%	13%	22%	40%	57%	67%	70%	71%	71%	70%	70%	63%	52%	37%	27%	20%	16%	14%	11%	10%
Summer Peak Day (7/17/12)	10%	9%	10%	10%	10%	14%	23%	45%	62%	68%	71%	75%	74%	74%	72%	68%	56%	37%	28%	21%	16%	12%	10%	9%

__Billing Analysis__

This section presents a discussion of results from the billing analysis portion of the evaluation. This includes the presentation of three different approaches to the overall billing analysis result: site-level modeling with and without a comparison group, and a pooled approach.

Table ES 5 presents the evaluated gross savings using the three billing analysis approaches. Each approach was employed in an effort to get past the limitations of the billing analysis process. The first approach uses the former 2010 participants as the comparison group for the 2011 participants. This difference-of-differences approach corrects for possible trends that exist outside of the program, such as an economic growth. This approach requires an adequate comparison group. The second approach provides evaluated gross savings without the comparison group. The disadvantage of this approach is that there is no control for external changes. The third approach is the Pooled Time-Series Cross-Sectional model. The pooled model takes a different approach that controls for monthly and post-installation effects in a single model.

The overall realization rate from the savings estimate with a control group for prescriptive lighting without controls is 20%, while the realization rate for all lighting measures is 23%. Without the comparison group, we find that the overall lighting realization rate improves to about 31%, which may indicate spillover effects among program non-participants. The realization rate from the pooled model-is 33% for 2011, which is very similar to the results without the comparison group. While it is difficult to say that the results from one approach are better than the others, we are inclined to consider the comparison group model results as being most appropriate. This is due to the composition of the group and because, in general, the approach addresses certain kinds of bias better than the model can with the other approaches.

Table ES 5: 2011 Billing Analysis Results Summary

Level of Analysis	Number of Accounts	Tracking Savings per Account (kWh)	Estimated Savings per Account (kWh)	Realization Rate	90% Confidence Bound
Evaluated Gross Savings with Comparison Group					
All Lighting	3,706	25,888	6,029	23%	8%
Prescriptive Light Without Control	2,547	23,201	4,644	20%	11%
Evaluated Gross Savings without Comparison Group					
All Lighting	3,706	25,888	7,997	31%	2%
Prescriptive Light Without Control	2,547	23,201	5,668	24%	2%
Evaluated Gross Savings with Pooled Model					
All Lighting	3,706	25,888	8,601	33%	5%

Net to Gross (NTG)

This section presents a discussion of results from the self-reported NTG portion of the evaluation. Table ES 6 presents these results, including direct attribution, spillover and overall NTG. The attribution rate is provided in the first series of rows in the table and represents the percent of savings installed through the program that were attributable to the program. Attribution is calculated as the inverse of the free-ridership rate. An attribution of 94 percent indicates that program free-ridership is 6 percent, indicating that the vast majority of savings resulting from this program would not have occurred without the program's influence. All but one measure group had attribution rates higher than 90 percent.

The spillover rate is presented in the middle of the table and represents the savings installed outside of the program by program participants as a result of their participation in the program. It is expressed as a ratio of savings attributable to program influence. The spillover assessed in this study is termed "inside like spillover", which is a subset of the larger spillover adjustment. Overall, the program "inside like spillover" rate was 2 percent.

The combined NTG ratio places attribution and spillover into a single ratio. The overall NTG for the program is 96 percent. Individual measure group ratios range from 78 percent for non-HVAC motors and drives to 97 percent for prescriptive lighting without controls.

Table ES 6: Net to Gross Rate by End-use Category Summary

End-use Category	n	Direct Attribution Rate	Relative Precision at 90% Confidence
Attribution Rate			
Prescriptive Lighting without Controls	668	94.8%	1.3%
Prescriptive Lighting with Controls	313	93.3%	2.6%
Refrigeration	54	93.6%	3.4%
Non-HVAC Motors and Drives	22	78.1%	36.2%
Overall Attribution	**1,057**	**94.2%**	**1.2%**
Spillover (Inside Like[3])			
Prescriptive Lighting without Controls	668	2.2%	54.7%
Prescriptive Lighting with Controls	313	0.6%	89.3%
Refrigeration	54	1.2%	92.8%
Non-HVAC Motors and Drives	22	0.1%	175.2%
Overall Spillover	**1,057**	**1.7%**	**47.6%**
Combined			
Prescriptive Lighting without Controls	668	97.0%	1.7%
Prescriptive Lighting with Controls	313	93.9%	2.6%
Refrigeration	54	94.8%	3.7%
Non-HVAC Motors and Drives	22	78.2%	36.1%
Overall Net-to-Gross	**1,057**	**95.9%**	**1.4%**

Using the gross realization rate and NTG estimate for prescriptive lighting without controls, we are able to calculate overall net savings. Table ES 7 shows these results. Note that although this table only presents the net savings for prescriptive lighting without controls, the findings of this study will allow National Grid to revise savings for all prescriptive lighting installed in the SBS Program.

Table ES 7: Prescriptive Lighting without Controls Net Savings

Parameter	Electric Energy (kWh/yr)
Tracked Savings	64,265,935
Gross Savings Realization Rate (RR)	80.3%
Evaluation Net-to-Gross Ratio (NTG)	97%
Net Savings	50,057,379

[3]Spillover savings is defined as the energy savings associated with energy efficient equipment installed by consumers who were influenced by an energy efficiency program, but without direct financial or technical assistance from the program. "Inside like spillover" refers to spillover savings of the same measure type as that installed through the program, installed at the same location.

1.4 Conclusions and Recommendations

The NTG results from this study are stable, with an overall estimate of nearly 96% based upon self-reports. This impact study also included the performance of on-site M&V work with re-engineering analysis to derive results that have traditionally not been achievable through other methods (e.g., the determination of coincidence factors, kW and kWh Interactive factors, connected kW and kWh realization rates, load shapes, and average hours of use). A billing analysis was also undertaken for purposes of deriving a kWh realization rate and to attempt a deconstruction of the realization rate into various components, such as hours-of-use realization rate, interactive effect factor, etc. This two pronged approach was expected to provide National Grid with all of the factors needed to refine and calculate future tracking system estimates through the on-site with M&V effort, along with an accompanying attempt to assess the ability to derive more granular results from a billing analysis. As it turns out, the results from the two methods of assessing gross savings for 2011 lighting measures are divergent and statistically different. Specifically, the M&V based assessment of prescriptive lighting without controls has a realization rate of 80% while the savings for the same measure group and year with comparison group is roughly 20%.

Had the two evaluation activities employed in this study provided consistent overall findings, it would have been feasible to perform a synergized, nuanced analysis. However, since the results are divergent, we are left to assess the veracity of each savings estimate. Ultimately, while many steps were taken in both the M&V and billing analysis to mitigate bias within state-of-the-art evaluation methods, we are recommending that National Grid utilize the on-site M&V results from this study and perform additional exploration of the two greatest weaknesses in the billing analysis before more fully embracing these findings. The two weaknesses are incomplete billing data and uncertainty around pre-existing equipment operation. Two recommendations to address these weaknesses are:

1. We recommend that the presence of meters and accounts be fully assessed at all sites included in the upcoming lighting controls study. DNV GL is planning to perform pre and post metering at roughly 70 sites among small businesses in New York for National Grid, which will necessitate multiple visits at each. Given the critical importance of the inclusion of all consumption data in a billing analysis, this upcoming project provides an opportunity to explore the incidence of multiple accounts at a project, compare this information to that in the tracking system and assess what those findings might mean for this and other billing analyses on the small business population.
2. We recommend that the operation of the pre-existing lighting equipment be quantified as part of the pre-metering visits associated with the study referenced above. This effort would provide insight into the likelihood of whether a lack of functioning equipment in the pre-existing condition is negatively impacting the billing analysis results. This effort should be fully

documented and include an assessment of what those findings might mean for this and other billing analyses on the small business population.

Other Conclusions and Recommendations

The following are recommendations based upon the findings from this study.

- The reliability of a billing analysis is largely dependent upon the availability and relationship of consumption data to program treated spaces. Despite a significant effort to acquire all accounts associated with each participant in the population as well as the on-site sample, we suspect the inability to do this consistently at either level limited the ability to produce a confident billing analysis result. We recommend that a renewed effort be undertaken to have program vendors more thoroughly collect meters and/or accounts that serve program treated areas in the interest of making a billing analysis a more feasible option for consideration when evaluating the small business program in the future.
- Currently, the TRM stipulates using a factor of 0.90 for calculating NTG in reporting program savings. We recommend consideration of the NTG results from this report for the SBS Program. There are two ways to apply NTG factors from this report. The first would be to apply this factor to tracking savings and utilize the net realization rate from this report. Alternatively, if changes to the TRM are made such that it is believed that the tracking savings become more similar to those estimated as gross savings in this report, then using the combined NTG ratio from this report may be more appropriate for consideration.
- The lighting hours-of-use assumptions in the TRM appear to be high in general. However, only the automotive facility TRM assumed hours fall outside of the 90% confidence interval result. As such, while there is evidence from this study that the TRM hours might be overstated, we encourage National Grid and the DPS to compile added evidence from other New York Program Administrators ("PAs") to be sure this trend holds across other territories before making TRM revisions on this matter[4].
- The TRM currently assumes a coincident factor of 1.0 for commercial indoor lighting measures, which we regard as very high. The estimated coincidence factor from this study is 0.566. We recommend consideration of this value for small business direct installation lighting applications or at least encourage National Grid and the DPS to compile additional data from other New York PAs.
- Currently, National Grid uses HVAC interactive estimates of 1.19 for summer kW and 1.07 for annual energy savings. These values are proxies that represent average factors from the table of HVAC interactive factors by Building Type in the TRM. This study suggests that National Grid might consider the use of a kW HVAC Interactive Effect Factor of 111.1%, and a kWh

[4] National Grid uses hours of use vendor estimates at the site and location level at the time of participation, however such adjustments may help other PAs that rely more heavily on the Technical Resource Manual.

interactive factor of 104.2% or again work with DPS to combine this data with that from other New York PAs to inform these values.

- We recommend that future studies of energy impacts of the small business program use an error ratio of 0.5 for lighting for-sites metering sample sizing. The error ratio we recommend for targeting demand savings is a 0.2 error ratio.
- If not being performed already, we recommend that National Grid work with Program vendors to ensure that the location of measures installed can be acquired for evaluation purposes.

2. Overview

This document summarizes the results of an evaluation of National Grid's 2010 and 2011 Small Business Services Energy Efficiency Program (SBS) in the Niagara Mohawk territory. The SBS Program provides direct installation of energy-efficient measures for small commercial and industrial ("C&I") customers. The small business sector (defined as customers with an average monthly demand of less than 100 kW) is a heterogeneous group of end users that represent a sizeable portion of the National Grid commercial electric customer base in their Niagara Mohawk territory.

The SBS program is largely considered a resource acquisition program. The following table summarizes the documented SBS program goals and achievements as captured in the 2011 annual report[5]. In 2011, the SBS Program overachieved on the number of participants (achieving 139% of goal). Also, the program overachieved net first year annual and peak savings achieving 132% and 161% of the goals, respectively. The total utility cost for the SBS 2011 program year was $38,584,695, which represents 136% of the funding budgeted. Although the program budget was overspent, the relationship between spending and energy savings achieved remained largely intact for the SBS Program.

Table 1: SBS Goals versus Achievement

2011 Parameter	Small Business		
	Achieved	Goal	% Achieved
Participants	8,291	5,978	139%
Net[6] First-year Annual MWh	133,382	105,859	132%
Net Peak kW	39,748	24,731	161%
Utility Cost (without Shareholder Inc.)	$38,584,695	$28,277,353	136%

2.1 Purpose and Overview of the Study

The primary objective of this evaluation is to quantify the gross and net annual energy and summer demand impacts of lighting measures installed through the SBS program. The savings and factors of interest to the study includes the coincident summer on-peak factor, connected kW savings and realization rate, kWh savings and realization rate, percent on peak kWh, summer demand HVAC interactive effect factor and kWh HVAC interactive effect factor. The study was designed to provide two primary means of assessing gross impacts and a third task to assess NTG impacts. All gross energy savings and net factor sampling work were designed to achieve ±10.0% at the 90% confidence level.

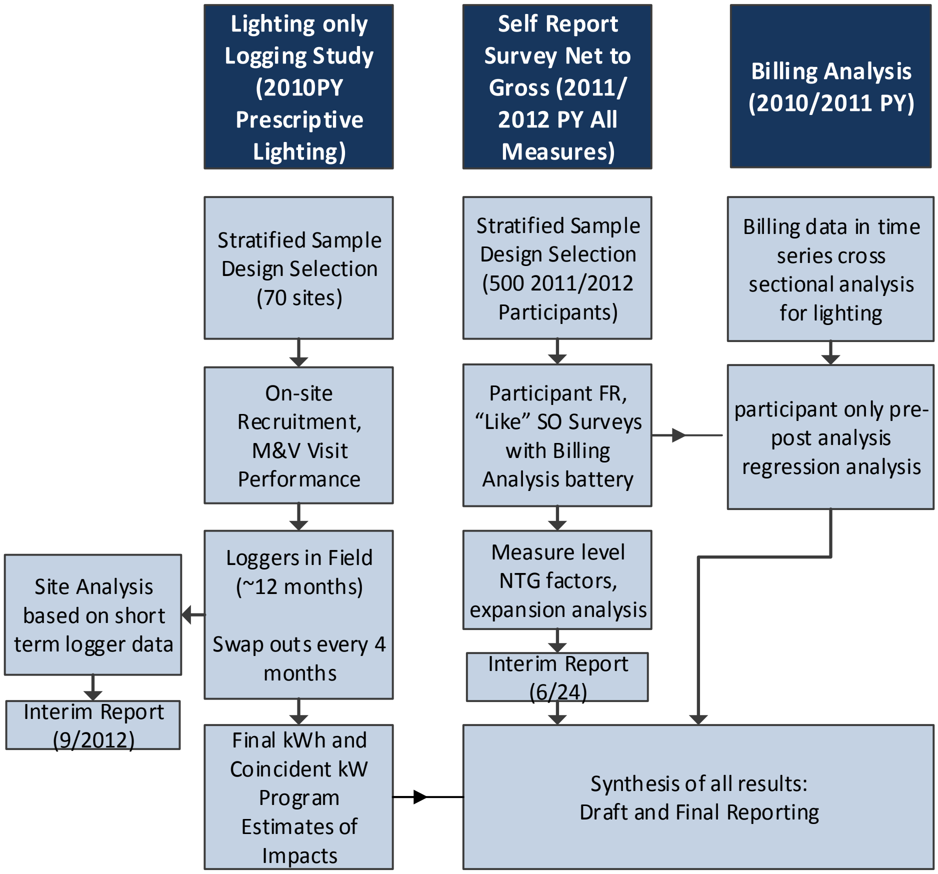
[5] 2011 Energy Efficiency Programs, Annual Report, March 30, 2012, p 53, 54 & 103.

[6] Net savings in this table includes Spillover net of free-ridership

Figure 1 below presents the overall evaluation plan, which includes the three primary evaluation activities. This includes an M&V logger study on 2010 prescriptive only lighting participants, an assessment of NTG factors from 2011 through July of 2012 participants, and a two stage billing analysis on lighting savings among 2010 and 2011 participants. Although the gross and net impacts examined are not consistently from the same set of program years, the technical manual calculations driving the savings estimates have been consistent during this time and program changes in SBS have been minimal from 2010 through 2012. This helps assure that gross savings estimates and realization rates from disparate years (e.g., 2010 vs. 2011), remain comparable. The study design was also developed to keep savings estimates from both the billing analysis and M&V defined as gross, which helps assist in their use as independent efforts that can be combined in a triangulating fashion.

The logging study was performed at a statistically selected sample of 70 on-site visits with a year of time-of-use lighting logging and included interim results deliverable after the first 4 months of metering. The billing analysis utilized weather normalized participant billing data in a pre/post analysis in a series of models to explore the most appropriate model for estimating gross program savings that would be comparable to the M&V based gross savings estimates. The NTG effects were determined through the use of 484 SBS program participant surveys that covered 1,055 program installed measures. As part of the NTG survey, questions were asked that were designed to better understand the billing analysis results as well as for incorporation into the billing analysis model. More detail on the specific methodologies employed is provided later in this report and as appendices.

Figure 1: Overall Evaluation Plan



2.2 Program Population Summary

Table 2 below provides a summary of the 2010 and 2011 SBS Program population. This summary breaks out savings for prescriptive lighting with and without controls, other lighting, and all other measures. Combined, the two years of activity tracked 245,219MWh of annual energy savings (accumulated from 111,936 MWh in 2010 and 133,283MWh in 2011).

We have grouped the prescriptive lighting[7] with and without controls as individual line items to show the relationship between the 2010 sample frame for the M&V study to the overall population. Savings related to prescriptive lighting measures without controls accounts for 58% of tracked savings in 2010, 64% in 2011, and 61% across both program years. Overall, lighting measures and controls are the focus of a majority of program savings, with 93% of total savings installed at 92% of accounts across years. The

[7] Prescriptive lighting measures were identified in the Measure Description field as being either CFL, Fluorescent, Custom Lighting, HPS, LED or PSMH.

prominence of lighting in this program was a primary consideration in deciding to focus on lighting in much of the evaluation work undertaken on the SBS Program. It should be noted that this study focuses on lighting without controls, although these results have a direct bearing on all prescriptive lighting, including those installed with controls. As such, these results will impact the savings estimates of the vast majority of program savings if incorporated into the TRM, and/or if used to refine tracking savings estimates by National Grid. National Grid is initiating a study on lighting controls at this time that will round out the study of savings associated with lighting.

Table 2: 2010 and 2011 SBS Program Activity

End Use	kWh Savings	
	kWh	%
2010 Program Year		
Prescriptive Lighting without Controls	64,498,471	57.6%
Prescriptive Lighting with Controls	36,946,355	33.0%
Other Lighting	1,659,890	1.5%
HVAC, Motors, Refrigeration	8,831,028	7.9%
Total	111,935,744	100.0%
2011 Program Year		
Prescriptive Lighting without Controls	85,525,632	64.2%
Prescriptive Lighting with Controls	39,845,725	29.90%
Other Lighting	13,887	0.0%
HVAC, Motors, Refrigeration	7,897,597	5.9%
Total	133,282,841	100.0%
Total of Both Years		
Prescriptive Lighting without Controls	150,024,103	61.6%
Prescriptive Lighting with Controls	76,792,080	30.60%
Other Lighting	1,673,777	1.0%
HVAC, Motors, Refrigeration	16,728,625	6.8%
Total	245,218,585	100.0%

As discussed earlier, the on-site M&V element of this evaluation focused on 2010 prescriptive lighting measures that have been installed in projects without lighting control installations. This group is comprised of 2,762 accounts, or just fewer than 66% of all participating accounts. However, a final adjustment had to be made to this participant group to ensure the final sample did not have lighting controls installed from the 2011 program year. After cross referencing the account numbers of 2010 participants that installed prescriptive lighting without controls with 2011 accounts that had installed controls through the SBS Program, we removed five accounts totaling 232,535 kWh from the sample frame. This sub population of accounts became the sample frame for the logger study task and totals 64,265,935 kWh of savings.

3. Methodology

3.1 Prescriptive Lighting On-sites with Metering

To estimate gross savings for prescriptive lighting, we performed a time-of-use study at a statistically selected sample of SBS program participants. The Energy Efficiency Protocol ("EEP") guidelines for statistical standards in evaluation work in New York are set at estimating gross savings at ±10% at 90% confidence. DNV GL used Model-Based Statistical Sampling ("MBSS") methodologies to inform a sample design that met these targets. The final sample design was based on the performance of 70 site visits[8].

The final stratified sample design is provided in the table below. The first three columns in the table provide the stratum number, the cut point used to allocate sites to each strata, and the number of projects in each stratum. The final three columns show the total tracking savings in each stratum, the number of sample points that were randomly selected from each stratum, and the final case weights. Our sample unit was a "project" that included all program activity at an account level, as opposed to using application level information as the sample unit. This approach provides an opportunity to cover all lighting activity at each site visited as opposed to only evaluating portions of activity that might be reflected in a single application from a site.

Table 3: 2010 Prescriptive Lighting Final Sample Design

Stratum	Max Savings	Projects	Total Savings	Sample	Final Case Weights
1	12,643	1,366	9,350,029	12	113.83
2	24,294	629	11,132,279	15	41.93
3	41,098	397	12,480,998	14	28.35
4	82,067	248	14,058,569	14	17.71
5	578,173	117	17,244,060	15	7.80

To help minimize non-response bias, DNV GL sent out advance letters to a primary and backup sample (70 primary sample points and approximately 70 backup sample points).These advance letters were sent out on National Grid letter head, introduced the study and DNV GL, explained the on-site process, and assured the customer that DNV GL was a qualified vendor and-encouraging the customer to participate in the site work if they were contacted. The customer was provided a phone number and National Grid

[8] In exploring error ratios for use in designing the sample, the highest error ratio we could find from similar lighting studies we had performed is 0.45.Since the SBS program has not been evaluated for gross impacts recently, we used of an error ratio of 0.5, which is marginally poorer than what we have experienced.

contact if they had any questions or otherwise wanted to authenticate the study. The backup sample points were used to replace primary sample points that had either refused or had been called five times or more and had not responded to left messages.

Recruitment for this sample was performed by experienced DNV GL phone recruiters. The final response and refusal rates experienced are provided in Table 4 below. The response rate calculated in the table below includes all customers that were eligible and refused the on-site or were eligible and unable to be reached (non-contact). In developing this table, we have remained consistent with the definitions and calculation of Response Rate [19] and Refusal Rate [110] described in the *Standard Definitions: Final Dispositions of Case Codes and Outcome Rates for Surveys*.[11] Response Rate 1 is calculated to be 67%. Refusal Rate 1 is calculated to be 11%. We consider a Response Rate 1 of 67% and a refusal rate of 11% to be reasonable.

Table 4: Final On-site Recruitment Response and Refusal Rates

Disposition Description	Number/Rate
Complete	70
Refused - Eligible	11
Non-contact - Eligible	24
Total Contacts	106
Response Rate 1	66.7%
Refusal Rate 1	10.5%

The response rate is an indicator of potential bias associated with sample-specific estimates of population parameters. We cite it as an indicator since an assessment of the extent of bias due to non-response really rests upon any differences there might be between those customers that allowed the visit and those that we were unsuccessfully able to schedule. Since the recruitment was performed for an impact study in the midst of an economic down turn, we might expect the greatest risk of bias to this study being a sample of non-contacts due to business closings. We examined all non-contacts in our sample and they all appear to be instances where the business was present (e.g., dispositions where we left messages that were not responded to (voicemail, other), where no one picked up the phone, or we spoke with someone who passed us on to another individual). As such, we do not believe the response and refusal rates experienced in this study have resulted in any particular bias in the M&V results.

[9]Response Rate 1, or the minimum response rate, is the number of visits divided by the number of visits plus the number of refusals and non-contacts.

[10]Refusal Rate 1 is the number of refusals divided by the visits plus the non-respondents (refusals and non-contacts).

[11]http://www.aapor.org/AM/Template.cfm?Section=Standard_Definitions2&Template=/CM/Content Display.cfm&ContentID=3156

Data collection performed at all on-sites included physical inspection and inventory, interview with facility personnel, observation of site operating conditions and equipment, and the installation of metering for roughly 12 months. At each site, the DNV GL team performed a facility walk-through that focused on verifying the post-retrofit or installed conditions of each program installed lighting measure in addition to HVAC equipment information to assess interaction. At sites with insufficient project paperwork, we engaged program vendors to gather additional documentation on measure locations. The NY DPS evaluation plan guidance suggests that the National Action Plan for Energy Efficiency ("NAPEE") Model Energy Efficiency Impact Evaluation Guide provides generally acceptable evaluation approaches for New York. The method for the prescriptive on-sites with metering adheres to International Performance Measure and Verification Protocol ("IPMVP") Option A as discussed in the document cited above[12]. Specifically, the method employed uses a combination of stipulated factors (wattage) as well as measurements of key factors (i.e., quantity and hours of use) to calculate savings in an engineering model. A full description of the monitoring, verification and analysis methodology is provided in Appendix A.

Figure 2 below provides a graphical presentation of the lighting logger installation at the on-site visits. The Y axis shows the number of sites with loggers installed while the X axis shows the months of installation. We began installations in mid-November of 2011, with all but one site performed by the end of January, 2012. There were two logger swap outs that occurred during the course of the 12 month metering period. Most loggers were removed between December 2012 and January 2013. All were removed by the end of March, 2013.

Figure 2: Site Logger Install Profile

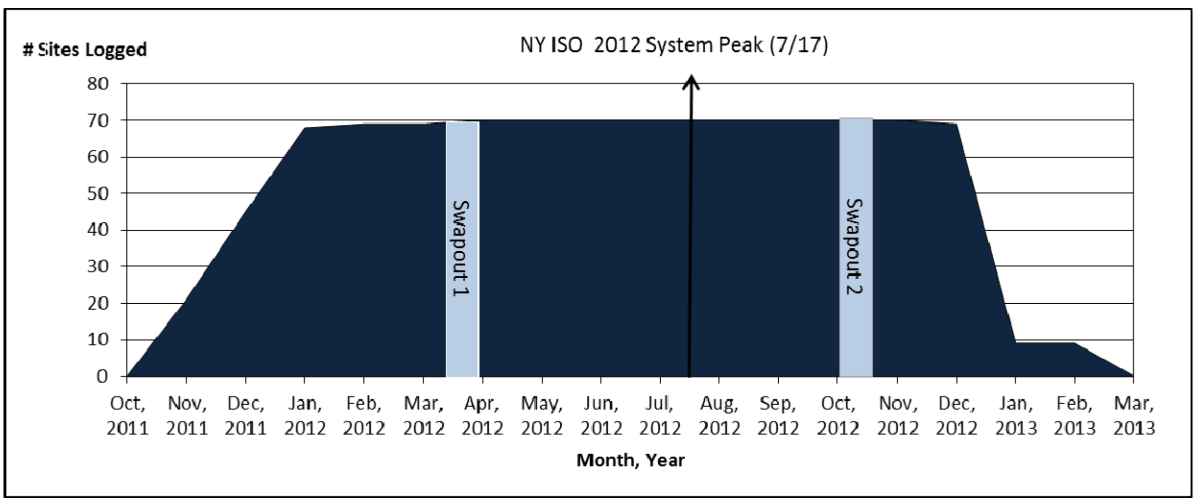


Table 5 shows the final status of the loggers installed and their associated records (hourly on-off transitions). A total of 657 loggers were installed in this study on program installed fixtures.

[12] http://www.epa.gov/cleanenergy/documents/suca/evaluation_guide.pdf, Section 4.

Approximately 3.3% of loggers failed, which is reasonable for a study of this nature. Another 1.4% exceeded its storage capacity and 1.7% were unable to be found[13]. These events resulted in the need to estimate or otherwise impute 13% of all time-of-use records. This estimation process was performed through one of three methods. When possible, we utilized data from a redundant logger (1.3%); otherwise we imputed it from other loggers at the site (6.3%) or utilized results for the same logger location from a different time period with similar characteristics, i.e. weekday non-holiday (92.4%).

Table 5: Logger and Record Final Disposition

Final Disposition	No. of Loggers	% of Loggers	No. of Records	% of Records
Logger Failure	22	3.3%	95,437	7.4%
Logger Deployment Issue	1	0.2%	6,145	0.5%
Logger Not Found	11	1.7%	46,300	3.6%
Loggers with Usable Data	623[14]	94.9%	1,139,838	88.6%
Loggers with Complete Data	614	93.5%	1,119,690	87.0%
Loggers with Partial Data	9	1.4%	20,148	1.6%
Total	657	100.0%	1,287,720	100.0%

An 8,760 hourly spreadsheet analysis was developed for each logger series[15] and was used to estimate hourly energy use and diversified coincident peak demand for all lighting sites. A spreadsheet engineering model was used to develop all savings estimates and factors of interest for each sampled site. A typical meteorological year (TMY3) dataset of ambient temperatures for the location closest to each site was used for all interactive savings analyses. The measure level analysis was performed in a manner that allowed the determination of impacts at each primary discrepancy that might cause the gross savings to differ from the tracking savings. Performing the analysis in this manner also allows us to report impacts at each level of adjustment. Each of these adjustments, or discrepancies, is described below:

- Documentation Adjustment: The Documentation Adjustment reflects any change in savings due to discrepancies in project documentation. Evaluators recalculated the tracking estimates of savings using all quantities, fixture types/wattages, and hours documented in the project file. All tracking system discrepancies and documentation errors are reflected in this adjustment.
- Technology Adjustment: The Technology Adjustment reflects the change in savings due to the identification of a different lighting technology (fixture type and wattage) at the site than represented in the tracking system estimate of savings.

[13] These loggers may have been inadvertently removed by the customer or stolen.

[14] This total is the sum of the loggers with both complete and partial data.

[15] By logger series we mean the series of loggers used to log a fixture over the course of the year.

- Quantity Adjustment: The Quantity Adjustment reflects the change in savings due to the identification of a different quantity of lighting fixtures at the site than presented in the tracking system estimate of savings.
- Operational Adjustment: The Operational Adjustment reflects the change in savings due to the observation or monitoring of different lighting operating hours at the site than represented in the tracking system estimate of savings.
- HVAC Interactive Adjustment: The HVAC Interactive Adjustment reflects changes in savings due to interaction between the lighting and HVAC systems among the sampled sites. Generally, these impacts cause a heating penalty and a cooling credit. This adjustment reflects impacts from electric heating and/or cooling. Other fuels were also calculated and reported on a per kWh saved.

All savings results were then expanded to reflect program level impacts through ratio estimation procedures. Please see Appendix A for a detailed discussion of the full analysis methodology and Appendix C for a detailed discussion of the expansion analysis.

3.2 Billing Analysis

Two different billing analysis approaches were performed. The first was a pooled, time-series, cross-section approach that weather-normalizes and measures savings in a single model that includes multiple months of data for all accounts/sites. The second was a site-level modeling approach, which is performed in two stages. The first stage of the site-level approach applies a more flexible, site-level weather normalization process. The second-stage estimates the savings on the pre-/post- differences in weather-normalized consumption. While we provide a discussion about the two methods here, we have also included the detailed formulas and equations in Appendix E.

Both billing analysis approaches produce estimates of gross savings. The pooled, time-series, cross-sectional approach estimates pre-post change in consumption without a comparison group. The baseline is pre-period consumption. The regression approach corrects for general trends and events that affect the whole population. There is nothing in this process that can be construed as controlling for what would have happened in the absence of the program which would be the case in a net savings estimate.[16]

The site-level approach estimates a pre-post change in consumption for participants and then corrects it with an estimate of pre-post change in consumption in the designated comparison group. In this approach, it is the comparison group that controls for trends and events that may be confounding the participant estimate of consumption change. The comparison group is also specifically designed to produce a gross savings result as it does not control for what would have happened in the absence of the

[16] This discussion specifically focuses on free-ridership as a driver of net versus gross savings. The issue of spillover is addressed below where we discuss the comparison group.

program with respect to installation of program-related measures. To that end, the comparison group is designed to be as similar as possible to the participants in terms of general characteristics, and approximately reflective of non-program change in the participant population. The use of prior participants serves this dual purpose because, as prior participants, they are similar to current participants in the key characteristic of program participation, and also as prior participants, we can rule out additional activity with program-related measures during the post-period after their earlier participation.[17] These two characteristics are what make prior or subsequent participants the best choice for a comparison group.

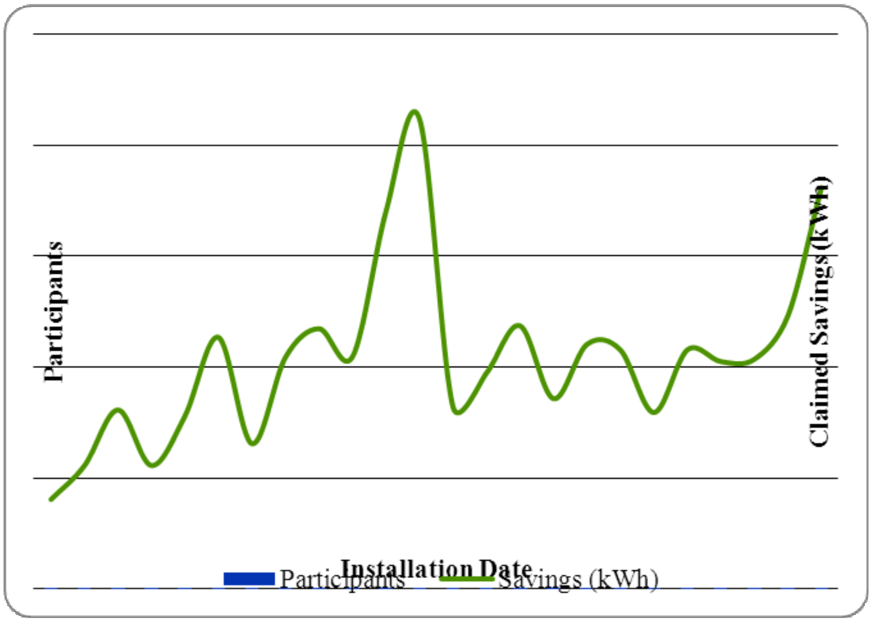
Installation Timing and Billing Data

The billing analysis focused on installations from January 2010 through December 2011. This data window facilitated a comparison group drawn from previous participants as well as approximately 12 months of pre- and post-installation data for both current participant and comparison group members. National Grid provided tracking for the two program years, as well as billing data from 2008 through 2013.

Table 2 earlier in this report provides an overview of the program tracking data for the time period of this analysis. An important aspect of the program participation data that is not covered in the overview is the distribution of participation over time. Figure 3 presents the number of participants and the claimed savings by date. This perspective is important for the billing analysis as it shows that estimated savings are based on installations that occurred across a range of installation dates through the year, supporting the use of prior and subsequent participants as comparison group members.

[17] For a billing analysis to produce a true net savings estimate, one would need to construct a comparison group with non-participants who installed program-related measures without a program rebate at the same time as their matched participant site. In this scenario, the non-program change captured by the comparison group would include program-related measure installation outside of the program.

Figure 3: National Grid SBS Participation Analysis



DNV GL received billing data in three batches. The first batch contained billing data for accounts from February 2007 through June 2012. The second batch included billing data for accounts in the small business tariff, covering the period of January 2012 through March 2013. The third batch, explained in more details in Section 1.4.3, contained billing data for 209 additional accounts.

The billing data was primarily screened for suspicious zero data. Accounts with zero reads were removed from the billing analysis, as they frequently indicated start-up or shut-down behavior that would confound the estimate of savings. Extremely high spikes in the billing data are addressed in subsequent data cleaning. All billing data preparation activities were applied consistently to both the pre- and post-installation periods and to both the participant and comparison group billing data.

Weather Data

The evaluation team used two kinds of weather data in this study: actual weather data and typical weather data. Actual weather data was retrieved from the Integrated Surface Hourly (ISH), a database that is actively maintained by the National Oceanic and Atmospheric Administration (NOAA). The ISH database undergoes a series of quality checks and is widely used. For typical weather data we used the Typical Meteorological Year (TMY3) database compiled by NREL.

The billing analysis used data collected from January 2009 through July 2013, for the following weather stations:

- Albany: USAF 725180
- Buffalo: USAF 725280

- Syracuse: USAF 725190

This weather data was screened for extreme and missing values. Once we determined that the quality of the data was satisfactory, we aggregated the hourly readings of temperature to daily averages.

Comparison Group

A comparison group was used in the analysis approach given the availability of early SBS participants from January 2010 through December 2010. Only consumption data from a time period after this group's program participation was used for the comparison group part of the billing analysis. Comparison group members were effectively matched with analysis group participants that installed measures a year later. This meant comparison group sites had a full year of data to match the analysis participant group's pre-installation period and a similar amount of data as the analysis group's post-installation period. Each comparison group site enters into the billing analysis model with a theoretical install date of 1 year after measure installation. That is, the comparison group effectively measures the change in consumption from the first year post participation to the second-year post-participation. The claim is that this difference reflects the overall, weather-normalized systematic change in this period. The participant evaluated gross savings estimate accounts for this difference.

The use of a comparison group in the billing analysis provides a savings estimate that should be considered gross savings. While we acknowledge that, despite the specific construction of the comparisons group, there may still be a partial net adjustment occurring in the savings, the use of a comparison group of past participants largely minimizes the net effects in the savings estimate. Their recent participation in the program makes it less likely that they would undertake any other program-related efficiency improvements outside of the program that might be included in the billing analysis results. This issue is discussed at length in the Uniform Methods Project chapter on whole building billing analysis for retrofit programs[18].

Spillover is a possible concern with the use of previous program participants as the comparison group. Under the present construction, spillover would reduce the savings estimate to the full possible extent if it occurred exactly 12 months after the original program participation. That is, the savings produced by the spillover would manifest as a downward movement in consumption for the comparison group which would be removed from the participant estimate of consumption change. The spillover effect would reduce to zero as it moved further from the artificial pre-post change point assigned to the comparison group site. Spillover for this evaluation is estimated at approximately two percent. This would be the

[18] Section 4, Chapter 8: Whole-Building Retrofit with Consumption Data Analysis Evaluation Protocol in The Uniform Methods Project: Methods for Determining Energy Efficiency Savings for Specific. http://www.nrel.gov/docs/fy13osti/53827.pdf

maximum downward bias on the gross savings estimate and would drop to zero, for instance, if spillover savings occurred contemporary with the sites previous program participation.

3.2.1 Challenges in Billing Analysis Evaluated Gross Savings

As part of performing the billing analysis, we investigated possible errors in the evaluated gross savings, which can threaten the validity of the realization rate ("RR") result. We looked at incomplete billing data, pre- and post-installation energy consumption, and mismatched comparison groups as part of this process.

Incomplete Billing Data

Initially, only the billing data for the primary accounts identified in the tracking data were available for the billing analysis. In the early stages of the billing analysis, the evaluation team identified lighting projects that saved more than 100% of the annual energy used by the meter to which they were assigned. With further investigation, we discovered that for some such projects the installed lighting fixtures were served by meters associated with multiple accounts. Incomplete tracking data is a problem because the full site-level savings are included in the denominator of the site-level RR, while the evaluated gross savings are estimated from only a single account. If there are multiple affected accounts, the overall effect is an artificial reduction in the RR.

Table 6 provides an example. This example site has two accounts, each of which realized a reduction in consumption with the installation of program measures. The program gross savings from the tracking database are 14,000 kWh/year. In the complete data scenario, both accounts are recognized as part of the site, with a pre-program energy consumption of 25,000 kWh/year and a post-program energy consumption of 11,000 kWh/yr. The realized savings (14,000 kWh/year) equals the tracking savings and the realization rate is 100%. In the incomplete data (which was available to the billing analysis), the consumption for only the primary account is present. The difference in energy use between the pre-program and post-program consumption is 6,000 kWh/year, for a realization rate of only 43%.

Table 6: The Impact of Incomplete Billing Data to Realization Rate

Inputs	Account #1	Account #2	Site
Complete data (actual savings)			
Tracking savings	14,000		14,000
Pre- program consumption	10,000	15,000	25,000
Post- program consumption	4,000	7,000	11,000
		Savings	14,000
		Realization rate	100%
Incomplete data (available to the billing analysis)			
Tracking savings	14,000		14,000
Pre- program consumption	10,000		10,000
Post- program consumption	4,000		4,000
		Savings	6,000
		Realization rate	43%

This simple scenario can be easily altered to include scenarios from more than two accounts. Across any possible scenario, missing billing data can only reduce the estimated savings in the numerator while maintaining the same estimate in the denominator. Evaluated gross savings can range from full savings to zero savings-producing realization rate of from 100% to 0%. An extreme example of a 0% realization rate would happen if the site savings were attached to the first-floor account number but all of the retrofitting occurred on other floors of the building service by other accounts.

<u>Pre- and Post-Installation Energy Consumption</u>

A common concern with the billing analysis approach is that non-program changes somehow systematically bias a result. With limited information as to what is occurring at each site, it is impossible to control for other major changes that might increase or decrease the consumption. A basic assumption underlying a billing analysis approach is that these effects, positive and negative, offset each other. These effects will offset if the participant group has positive and negative non-program changes that are approximately equal (expansions vs. closures, increases in machinery vs. other energy efficiency measures, etc.). They will also cancel out if a comparison group is used in the analysis and its consumption expands or contracts at a similar rate to the participant group.

The reason these kinds of biases are important to consider for an analysis like this is that they occur at the scale of consumption, not savings. For example, if the program gross savings for an analysis group were approximately 20% of site-level annual consumption, any effect of bias in consumption will be amplified fivefold in the savings estimate. For example, a 5% overestimation of post-installation consumption will reduce the RR by 25%.

The primary concern is when the measure of interest is a small part of a larger project. If the lighting retrofit coincided with a major expansion, then the savings due to the increased efficiency of the new lighting could be lost amidst the overall increase in consumption. Alternatively, if the lighting retrofit coincided with the installation of a new, high efficiency cooling system, the reduction in consumption could be well beyond that, which is explained by the lighting retrofit alone.

While speculative, the nature of the SBS program recruitment process should make either of these scenarios a lower-probability outcome. The program is designed to recruit hard-to-reach businesses that are not already participating in other programs and are challenged regarding upfront capital costs. This makes it less likely that businesses are integrating SBS installations into some larger project that could change the consumption dynamic of the site.

It is important to note that using a comparison group does not control for this particular concern. The comparison group controls for the effects of general economic trends, however, it does not control for other site consumption changes that took place separate from the lighting implementation. Because the comparison group is drawn from prior participants, to the extent that those additional changes (positive or negative) in consumption take place, they do so at the time of installation, prior to the period under consideration for this evaluation.

Since it is so difficult to rule out some bias under the circumstances of the billing analysis, it is worth considering the potential scale of such a bias. The evaluation team has identified no hypothetical scenarios that would affect more than a relatively low percentage of sites. This would require an extreme bias within this small group to affect the overall result. There is limited evidence for this in the model diagnostics and savings distributions.

Weather Normalization

There is one additional caveat that is always salient for billing analysis-based evaluations. The weather-normalization aspect of billing analysis is essential for the comparison of pre- and post-installation consumption on an equal footing. Weather normalization is achieved based on a relatively simple regression-based model structure that disaggregates consumption into up to three kinds of consumption: base load, heating and cooling. The latter two are identified based on the correlation of consumption with heating and/or cooling degree days.

While this approach is powerful and widely used on residential billing data, the model structure is simple and may not adequately address drivers of commercial activity like changes in schedules or commercial-specific technologies like ventilation systems, etc. To the extent that this is the case, which is difficult to gauge, it is primarily a concern when weather during the evaluation experience is substantially different in the pre- and post-installation periods. The more extreme the actual weather, the greater potential for error in the normalized estimates. The more different the weather in the pre- and post-installation periods, the more likely the normalization is affecting the estimated pre- to post- installation change.

For this evaluation, cooling degree days increased substantially through the analysis period from the unusually cool summer in 2009 through to an extremely warm summer in 2012, with more than doubled the cooling degree days. Under this weather scenario, if a model under-estimated site-level cooling, then the weather–normalized estimates of consumption would be biased upward in each subsequent year, in turn downwardly biasing any efficiency related savings[19]. It is difficult to assess whether this is the case either at the site level or in aggregate. However, the nature of the weather that occurred through the analysis period makes it important to include as a potential challenge to the billing analysis.

3.2.2 Mitigation Strategies

In addition to some of the challenges noted earlier, the evaluation team is also aware of recent small business evaluations with both M&V site work and billing analyses that have provided divergent results, although these efforts are not publicly available at this time. As part of this study, multiple steps were taken to help mitigate this possibility. These activities include the items bulleted below and are further discussed in the results section of this report.

- We ran the billing analysis on the site level sample from the M&V study to facilitate a direct comparison of results using the two methods.
- We included additional questions in the NTG survey that were asked of 2011 participants to help further isolate program impacts from other business or economic changes.

In addition to these items, we took great care to work with National Grid to ensure the integrity of the both the tracking and billing data used in the analysis. This included the following data checks.

<u>Checks on Tracking Data</u>

DNV GL and National Grid checked the tracking data in multiple ways and multiple times as part of this process. First, we verified that the total tracking savings matched the total savings claimed by the program. We also submitted the complete list of accounts with savings to National Grid, to confirm that the values were correct. Finally, at a later stage of the project, we submitted a sample of 4 accounts to National Grid to double-check that the values were still accurate. Although National Grid found small discrepancies in the assumed installation dates, the tracking savings were confirmed, and the differences in dates were deemed too small to affect the billing analysis.

---

[19] In other words, if we underestimate cooling and the second year in the analysis is hotter, adjusting the cooling load would not allow us to appropriately downscale the post consumption to "normal". This would under-estimate savings. If the pre-period was hotter, the opposite could happen and the model would over-estimate savings

Checks on Billing Data

Another possibility for introducing bias into the realization rate is using incorrect billing data for the program participants. Should this happen, the consequences could be significant because participants could be assigned billing data reflecting no program changes. Here again, we performed several checks to verify that the data used in the study was correct. The first check was to confirm with National Grid that we were provided with a complete extract of the billing data. National Grid confirmed that they provided us with a database containing billing data for all accounts in the proper tariff code.

The next check was to ensure that the matching between the tracking accounts and the billing accounts was performed properly. Internally, we verified that the account numbers in both datasets were the same after removing leading zeroes that appeared in certain cases. The matching was also confirmed with National Grid's assistance on two occasions. In the first instance, we provided plots of billing data and savings for a random sample of accounts that National Grid spot checked. In the second instance, we provided National Grid with a list of accounts containing both the tracking savings and the raw, disaggregated billing data associated with those accounts. National Grid confirmed that the billing data for those accounts were matched correctly and also had the right values for the billing fields (start date of the reading, end date of the reading, number of days in the billing period, and the kWh for each billing cycle).

Checks on Additional Accounts

The next potential source of downward bias for the realization rate is that the analysis might be missing billing data associated with the program savings. This issue was discussed earlier in Section 3.2.1. Due to the importance of the nature of incomplete billing data in the participating population, we took multiple steps to help draw together the most appropriate analysis dataset possible.

After receiving the initial list of participating accounts, DNV GL worked with National Grid to identify multiple accounts that might be located at the same site listed on the application, for each customer. This included the identification of customers behind the accounts, and then extraction of all other billing accounts for those customers. With a more refined dataset, DNV GL completed an account matching effort and marked all accounts that shared the same approximate address by street name, city, state, and zip code. Those accounts with the same customer ID and the same approximate address were flagged as having the same site ID.

Following this mapping of accounts, a third list of additional accounts was provided by National Grid for the analysis. National Grid queried its database system for all applications that had additional accounts listed with paid dates in 2010 and 2011. In all, National Grid provided a final list of 311 additional accounts that were also included in the analysis.

In addition to this effort to add accounts, while on-site, we gathered information on the relationship between meters and the treated areas by asking field staff to look for and write down any additional meters or billing accounts that could be found at the same premise. We were able to perform this effort at 37 sites that were also in the billing analysis, although at many of these sites we encountered challenges to identifying additional meters, including known meter locations being inaccessible (some maintained by offsite property managers), contacts who did not know meter location(s), and general uncertainty over whether all meters were observed following full walkthroughs. Out of the 37 sites examined, we found additional meters or accounts for inclusion in the analysis at 10 sites. National Grid then provided the corresponding billing accounts for the additional meters in that list.

3.3 Net to Gross

We employed a self-report NTG approach to estimating free and "inside like spillover." This approach utilized a core algorithm that is consistent with the standardized approach that National Grid has exercised for assessing NTG in Massachusetts. DNV GL also calculated NTG based on an alternative method that we have seen implemented in NY State. This survey was performed with 484 participants from 2011 through mid-year, 2012.

The NTG sampling approach below was implemented to select the measure types and number of targeted completes. The final sampling methodology employed a stratified ratio estimation model that placed participants into measure types of interest and then into strata by measure size, measured in kWh savings. The methodology then estimated appropriate sample sizes based on the expected variance of the net kWh to the tracking kWh (the error ratio). We looked at two precision estimates when developing the sample design.

- For the first estimate, we assumed an error ratio of 0.7. This error ratio is based on previous experience in evaluations where there are fairly diverse measure types and the expected gross RR is close to 100%. Using this error ratio, we estimated the relative precisions on the NTG ratio, labeled "Expected Relative Precision" in Table 6.
- For the second estimate, we assumed a "worst case" estimate of error ratio of 1.0, labeled "Conservative Relative Precision." Using this error ratio, we estimated the relative precision on the NTG ratio if the variance in observation was substantially different from previous NTG studies.

Table 6 shows the estimates of precision by sampling measure category. The relative precision at the overall program level was designed to be less than 10% at the 90% confidence interval under both scenarios. We also planned the precision for Tube Lighting to be less than 10% and the precision for each of the other sampling measure categories to be less than 20%.

Table 7: Designed Precision Estimates by Sampling Measure Category

Sampling Measure Category	Sample Measures	2011/2012 Population Measures	Expected Relative Precision at 90% Confidence	Conservative Relative Precision at 90% Confidence	Percent 2011/2012 Program Savings (kWh)
Fluorescent Tube Lighting	250	7,719	7%	10%	70%
LED	100	2,805	12%	17%	12%
Lighting Controls	40	1,582	18%	26%	3%
Other Lighting	60	7,184	16%	22%	9%
Non Lighting	50	659	16%	23%	6%
Overall	**500**	**19,949**	**6%**	**8%**	**100%**

As stated earlier, the NTG approach used in this study was consistent with that used for National Grid in other territories. Since this work was scoped, however, the NY DPS has issued guidelines concerning the estimation of NTG based upon self-reports (Appendix H). We reviewed the Appendix H document provided by the NY DPS as part of this evaluation and have confirmed that the approaches used in our NTG assessment comply with that guidance. Specifically, the methods we used include:

- A NTG battery that included setup questions to ensure the correct respondent was surveyed and to help orient the respondent to topics of interest in the instrument.
- An analysis approach that was performed twice by different staff as a redundancy effort to assure the accuracy of the estimate. DNV GL has also exercised an approach that handles the timing variable slightly differently to determine the sensitivity of that input to using it with different methods.
- A NTG battery that included the use of consistency checks to help ensure that contradictory answers from respondents are identified and clarified.
- A NTG battery performed at the measure level to guide respondents to provide more accurate and specific information about timing, efficiency and quantity questions that would otherwise be captured with general NTG questions.

The NTG surveys were performed by Tetra Tech, an energy services company with direct experience in NTG survey design and data collection of this nature. The final response and refusal rates experienced in the NTG survey are provided in the table below. Response Rate 1 is calculated to be 39%.Refusal Rate 1 is calculated to be 15%.

Table 8: NTG Survey Disposition

Status	Number (n=1,249)
Bad phone numbers	24
Language Barrier	9
Completed surveys	484
Called out (>20 attempts)	5
Refused	189
Dropped Interview	3
Already completed survey	3
Ineligible[20]	77
Active sample - quota closed	354
Active sample - quota not closed	96
Response Rate 1	39%
Refusal Rate 1	15%

The final NTG survey sample is provided in the table below. A total of 498 surveys were completed overall, representing a total of 1,057 measure types. A total of 668 were performed on prescriptive lighting without controls, which was the primary measure type of interest in the on-site M&V work.

Table 9: NTG Completes by Measure

Analysis Category	n
Prescriptive Lighting without Controls	668
Fluorescent Tube Lighting with Controls	118
Lighting Controls	93
Cooler Measures	54
LED	33
CFL	29
Exit LED	23
Motors/Drives - Non-HVAC	22
Other Lighting	17
Overall	**1,057**

The Massachusetts standard approach is the core approach used in this study for assessing NTG for this study. This approach is well documented[21], has been designed to address many of the concerns that can

[20] Ineligibility includes those who did not recall participation, did not recall the measure installation, and were retired or no longer in business.

accompany self-reported NTG methods and is consistent with the approach taken in other National Grid territories such as Rhode Island and Massachusetts. The questions and algorithms used to determine the measure level net factors are provided in detail in Appendix D.

As part of this exercise, we also decided to assess free ridership via an alternate method based upon concerns we are aware of in NY regarding how various factors can be used to impute the final estimate. The core issue of concern is that while efficiency, timing and quantity are each unique elements to explore for free ridership, timing can be considered conditional upon efficiency and quantity. As such, the multiplication of all factors may systematically produce a lower free ridership score (and higher NTG estimate). In the alternate algorithm method, we multiply efficiency and quantity as a single initial free ridership score, then average that score with the timing rating. We present a comparison of these approaches at the end of the NTG results section. Otherwise, all results are founded upon what we refer to as the core (MA) method.

[21] http://www.ma-eeac.org/Docs/8.1_EMV%20Page/2011/2011%20Commercial%20&%20Industrial%20Studies/Mass%20Cross-Cutting%202010%20CI%20FR-SO%20Final%20Report.pdf

4. Results

This section is broken up into four primary sections. It includes a section on the results of the time-of-use metering and verification performed on prescriptive lighting installed in the program without controls, a section on the prescriptive and overall lighting billing analysis, a section on NTG results, and a section that seeks to interpret all results as a whole.

4.1 Prescriptive Lighting On-sites with Metering

This section of the study focused exclusively on prescriptive lighting installed without controls installed in the 2010 SBS Program. There are six specific results of interest from this study activity, which are listed below. We provide program level realization rates for two of these results (kWh and connected demand (kW)).We also provide peak demand (kW) coincidence and interactive factors at the time of summer peak as well as a kWh interactive factor, average hours of use, and percent on-peak savings. Where appropriate, we provide precisions around each result at the 90% confidence interval. A detailed description of all savings factors is presented in Appendix B.

- Summer kW Coincidence Factor –This is the percentage of the connected kW savings coincident with the summer peak period, which is defined as hour ending 5 pm on the hottest day, excluding holidays - (in this case, July 17, 2012 at hour ending 5 p.m.).
- Summer Demand (kW) HVAC Interactive Effect Factor – This is an adjustment factor applied to the gross connected kW savings that are due to interactive effects during the summer on-peak period.
- Connected kW Realization Rate – This is the ratio of the evaluated connected kW savings divided by the tracking connected kW savings. This realization rate is broken out by various adjustments, including documentation, technology and quantity.
- Average Hours of Use – Average annual hours of use for lighting measures based on lighting logger data.
- kWh Realization Rate – This is the ratio of the evaluated gross energy savings divided by the tracking gross energy savings. This realization rate is broken out by various adjustments, including documentation, technology, quantity, hours of use and interactive.
- kWh HVAC Interactive Effect Factor – This is this is an adjustment factor applied to the gross kWh savings that are due to interactive effects.
- Percent On-Peak Energy Savings – The percentage of energy savings that occur during on-peak hours defined as non-holiday, weekdays between 7 am and 11 pm.

Figure 4 presents a scatter plot of evaluation results versus tracking savings for annual energy savings (kWh). A one-to-one reference line is plotted as a dashed line on the diagonal of the figure. In addition, the final realization rate is plotted as a solid dark line reflecting the average savings-weighted realization rate of all sample points. We also show the final gross realization rate of 80.3% as a line.

Figure 4: Scatter Plot of Evaluation Results for Annual kWh Savings

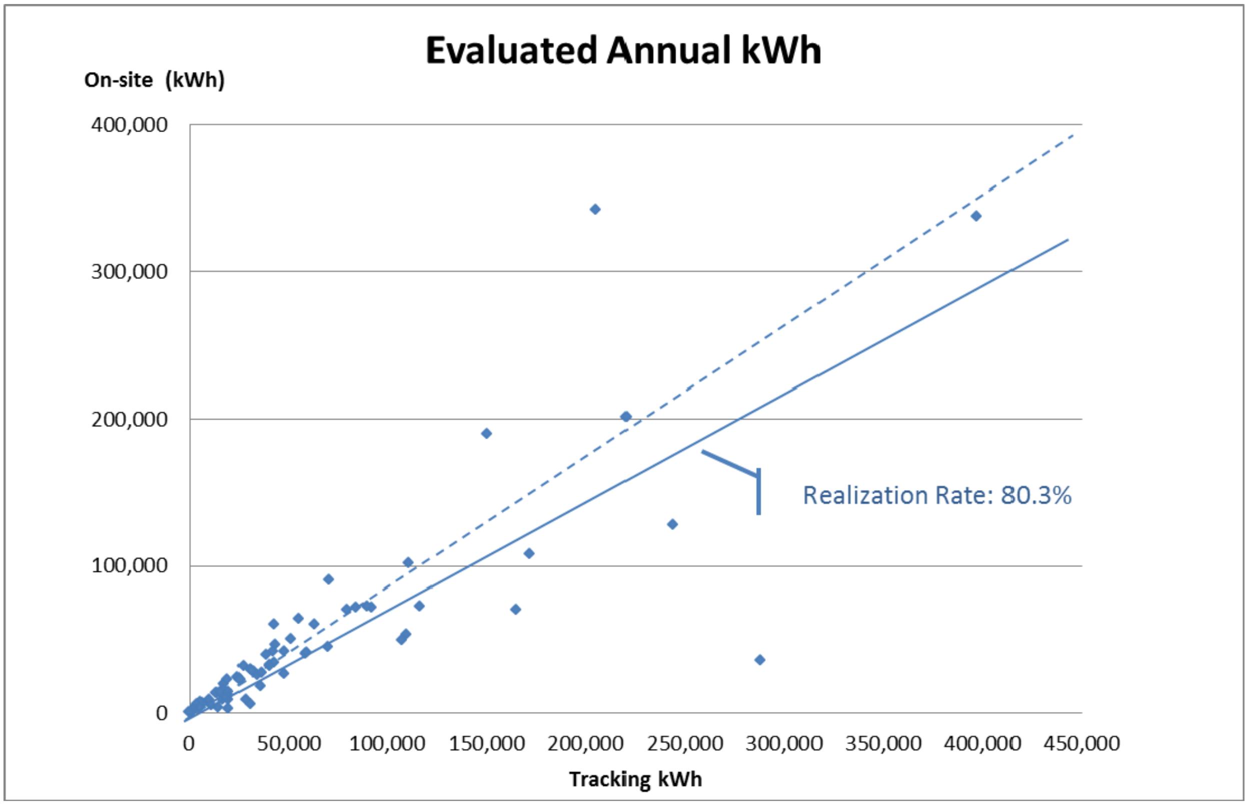


Table 10 summarizes the results of this analysis for the SBS Program. This table presents these results by nature of discrepancy in two ways. The first is by providing the incremental ratio, which shows the percent difference between each level of adjustment. The second is cumulative ratio, which shows the percent difference between the each level of adjustment and the tracking estimate. Showing the results in this manner allows one to see the relative change from one adjustment to another as well as the overall accumulated change relative to the tracking estimate.

The annual energy savings gross realization rate for prescriptive lighting without controls was found to be 77.2% without HVAC interactive effects included and 80.3% with interactive. The relative precisions for the final impact estimate with interactive is ±8.6%. The primary driver of the realization rate was the decrease in operating hours observed through the on-site M&V, which caused a 15% drop in energy savings relative to that captured in the previous quantity adjustment. In studies of a similar nature, we have found that participant self-reported hours, such as those used to estimate the tracking savings for this program, do tend to be overstated in general. This may explain some of the observed change in operating hours. The final error ratio was found to be 0.44.

Table 10: Summary of Prescriptive Lighting Gross Energy Results and Realization Rates

Prescriptive Lighting Without Controls (n=70)	kWh	Incremental Ratio		Cumulative Ratio	
		Value	Relative Precision at 90%Confidence (±)	Value	Relative Precision at 90% Confidence (±)
Tracking	64,265,935	N/A	N/A	N/A	N/A
Documentation Adjustment	64,187,086	99.9%	0.6%	99.9%	0.6%
Technology Adjustment	63,873,139	99.5%	0.5%	99.4%	0.7%
Quantity Adjustment	61,163,048	95.8%	1.7%	95.2%	1.8%
Operational Adjustment	49,591,059	81.1%	8.9%	77.2%	9.0%
HVAC Interactive Adjustment	51,594,070	104.0%	1.1%	80.3%	9.1%
Gross Realization Rate	51,594,070	80.3%	9.1%	80.3%	9.1%

Table 11's format is the same as Table 10, but provides results around demand impacts. The connected kW gross realization rate is 95.7%, with a relative precision of ±2.0% at the 90% confidence level. The primary driver of this realization rate was lower installed quantities of lighting observed on-site than were recorded in the tracking system. However, we note that this adjustment was still minor in nature (<4%) and occurred over three years. The final error ratio was found to be 0.13.

Table 11: Summary of Prescriptive Lighting Gross Demand Results and Realization Rates

Prescriptive Lighting Without Controls (n=70)	kW	Incremental Ratio		Cumulative Ratio	
		Value	Relative Precision at 90% Confidence (±)	Value	Relative Precision at 90% Confidence (±)
Tracking	18,437	N/A	N/A	N/A	N/A
Documentation Adjustment	18,370	99.6%	0.8%	99.6%	0.8%
Technology Adjustment	18,340	99.8%	0.7%	99.5%	1.0%
Quantity Adjustment	17,639	96.2%	1.7%	95.7%	2.0%
Gross Realization Rate	**17,639**	**93.9%**	**2.0%**	**95.7%**	**2.0%**

Table 12 shows the remaining factors of interest from the on-site M&V work for prescriptive lighting measures. The estimated coincidence factor from this study is 0.566, which compares to a factor of 1.0 from the TRM. Although with slightly different definitions, the estimate of 0.566 is in the middle of a recent New Hampshire Small Business lighting study that estimated 0.47 and a recent Massachusetts Small Business lighting study that estimated 0.66.

The evaluated values for the kW HVAC interactive effect factor are 1.104 for summer kW and 1.04 for annual kWh. These reflect the percent of gross summer peak kW and gross kWh savings that are due to interactive effects. Currently, National Grid uses estimates of 1.19 for summer kW and 1.07 for annual kWh. These values are proxies intended to represent an average from the table of HVAC interactive factors by Building Type in the TRM. The final three rows show the average weighted hours of use from the lighting loggers, the % on peak energy savings, and the non-electric HVAC interactive effect in MMBtus per kWh saved.

Table 12: Other Savings Factors

Other Ratio Estimation Results (n=70)	Value	90% Confidence Interval (+/-)
Summer kW Factors		
Coincidence Factor	56.6%	15.3%
HVAC Interactive Effect Factor	110.4%	2.7%
KWh Factors		
HVAC Interactive Effect Factor	104.0%	1.1%
Average Hours of Use	2,708	329
% On Peak KWh	72.0%	6.1%
Non-Electric		
Heating HVAC Interaction Effect (MMBtu/kWh)	-0.0023	0.0020

4.1.1 Lighting Profiles

Figure 5 presents the weighted percent on time for non-holiday weekdays and weekends. These profiles are weighted by controlled kW from all lighting loggers installed. The pattern of use of lighting during the weekend as well as the week is roughly the same, with hours beginning to ramp up in the 6 a.m. hour and beginning shut down in the 4 p.m. hour. The rate of operation is roughly 24-28% higher during the week on core day hours as compared to the weekend, although the lowest points of each pattern are consistently 9% to 10%.

Figure 5: Weekday and Weekend Lighting Profiles

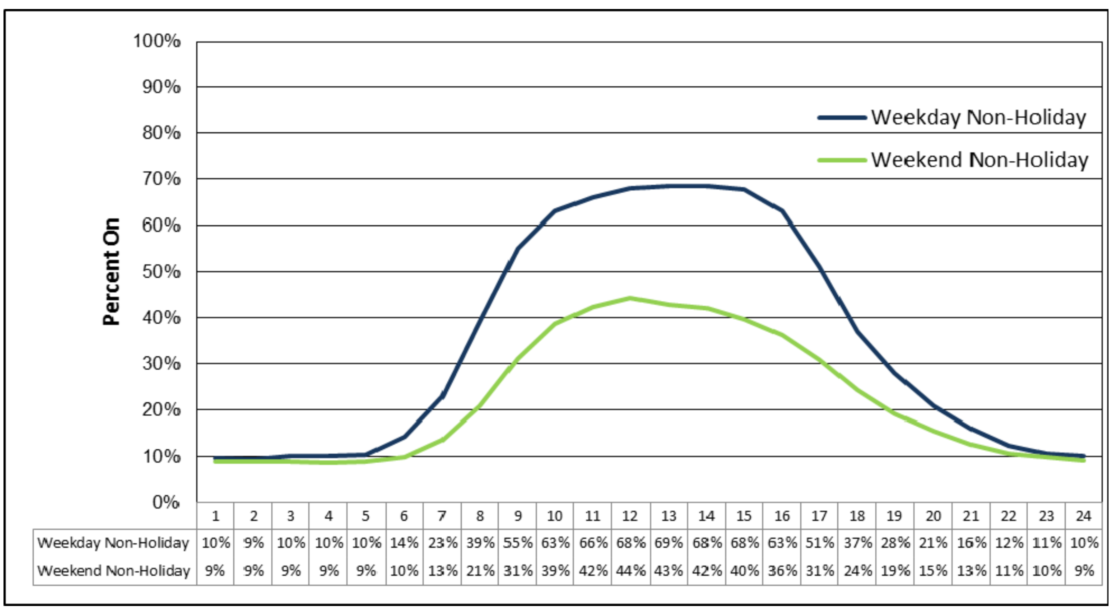
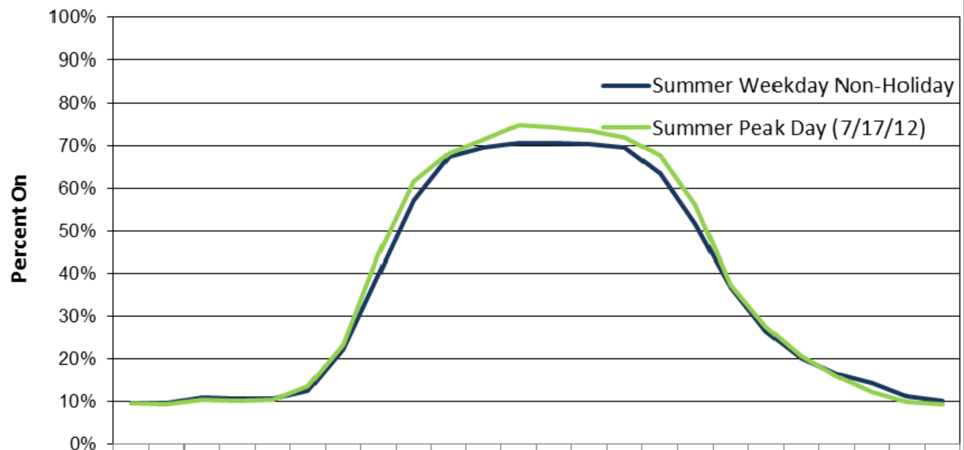


Figure 6 presents the weighted percent on time for non-holiday summer weekdays and for the NY peak day in 2012 (July 17). The summer peak day profile represents lighting operation during system peak and is largely the same as the regular summer weekday profile, primarily due to lighting not being a weather sensitive measure. The 56% noted in hour 17:00 (5 p.m.) is in fact the coincident factor presented earlier.

Figure 6: Summer Weekday and Peak Day Lighting Profiles



	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24
Summer Weekday Non-Holiday	10%	9%	11%	11%	11%	13%	22%	40%	57%	67%	70%	71%	71%	70%	70%	63%	52%	37%	27%	20%	16%	14%	11%	10%
Summer Peak Day (7/17/12)	10%	9%	10%	10%	10%	14%	23%	45%	62%	68%	71%	75%	74%	74%	72%	68%	56%	37%	28%	21%	16%	12%	10%	9%

Hour Ending

4.1.2 Hours of Use

The long term time-of-use metering that occurred in this study provided the opportunity to compare the ability of short term metering to estimate annual hours of use to actual hours of use. Specifically, at the time of performing the interim report of M&V findings, we expanded the first wave of metering performed to reflect annual operation based upon the data itself as well as customer reports on seasonality and other cyclic business behaviors that might affect lighting operation. This was assessed on a case by case basis, with only a couple of sites needing adjustment to estimate the full year of operation. Table 13 shows the comparison of the first three month metering period to the full 12 month metering period for the three facility types with the most sample points available to produce reliable estimates. In the three month metering period, we had 68 sites with interim data as two sites were unable to be revisited in time for that report.

Overall, the final 12 month estimate of lighting hours was 94.5% of that estimated in the three month interim report. Of the 68 sites in the 3 month sample, 36 (or 53%) had lower operating hours in the 12 month actual hours than those annualized after the first logger wave. Another 43% had higher schedules and 4% had the same fixed schedule in both metering periods.

Table 13: Annualizing Three versus Twelve Months of Metering

Facility Type	Count of Facility Type		3 Month Hours of Use	12 Month Hours of Use	Percent	Relative Precision at 90% Confidence	Absolute Precision +/-Hours
	3 Month	12 Month					
Automotive Facility	8	8	2,338	2,469	105.6%	±15%	369
Manufacturing Facility	20	20	2,890	2,825	97.8%	±23%	582
Retail	17	17	3,258	3,463	106.3%	±17%	587
Other	23	25	2,734	2,487	91.0%	±26%	659
Overall	**68**	**70**	**2,864**	**2,708**	**94.5%**	**±12%**	**329**

Table 14 presents a comparison of the 12 months hours of use estimate as derived from the long term metering to the hours of operation assumptions in the TRM. These results suggest that overall the TRM is overestimating hours of use by roughly 39%. However, in looking at the facility level results, only the automotive facility TRM estimate falls outside of the error bound around the 90% confidence interval result.

Table 14: Comparison of Twelve Month Metering to NY TM HOU Assumptions

Facility Type	Count of Facility Type	NY TM Hours of Use	12 Month Hours of Use	Percent	Relative Precision at 90% Confidence	Absolute Precision +/-Hours
Automotive Facility	8	4,056	2,469	60.9%	±15%	369
Manufacturing Facility	20	2,857	2,825	98.9%	±23%	582
Retail	17	3,748	3,463	92.4%	±17%	587
Other	25	4,496	2,487	55.3%	±26%	659
Overall	**70**	**3,796**	**2,708**	**71.3%**	**±12%**	**329**

Table 15 presents a comparison of the three month estimate of the summer coincident factor to the final 12 month estimate as derived from the long term metering that logged over the peak period. The three month estimate was based upon operation at hour ending 5 pm on the same weekday that the historical peak fell on as gathered from the time of use data gathered in the first wave of metering (December – March). The actual final frequency weighted 12 month estimate was roughly 6% higher than the three month estimate taken during the winter. This suggests that lighting operation at the coincident peak hour for small business may be driven more by business hours of operation and perhaps seasonal activity than daylight or other external factors.

Table 15: Comparison of Three versus Twelve Month Summer Coincident Factor Estimate

Facility Type	Count of Facility Type	3 Month CF Estimate	12 Month CF Estimate	Percent
Automotive Facility	8	72.0%	80.7%	112.0%
Manufacturing Facility	20	42.5%	48.5%	114.2%
Retail	17	82.9%	82.4%	99.5%
Other	25	34.1%	44.7%	131.2%
Overall	**70**	52.7%	59.1%	112.1%

4.2 Billing Analysis

This section presents a discussion of results from the billing analysis portion of the evaluation. The results include three different approaches to an overall billing analysis result: site-level modeling with and without a comparison group, and the pooled approach. In addition, results from a billing analysis of the on-site sample provide both a direct comparison with the on-site engineering results as well as evidence of the effect of locating more complete site-level data.

4.2.1 Evaluated Gross Savings with Comparison Group

Table 16 presents the evaluated gross savings under the approach of using the former 2010 participants as the comparison group for the 2011 participants. This difference-of-differences approach corrects for possible trends that exist outside of the program, such as an economic growth. This approach requires an adequate comparison group. This approach also has the limitation that it is not possible to report results for the 2010 on sites, since on-sites were not performed on 2011 participants. The overall realization rate for prescriptive lighting without controls is 20% while the realization rate for all lighting measures is 23%. One notable pattern is that the realization rates are quite different across regions: the realization rate is the lowest for participants in Buffalo at 12%, while it is the highest for Albany at 44%.

Table 16: Evaluated Gross Savings with Comparison Group

Level of Analysis	Number of Accounts	Tracking Savings per Account (kWh)	Estimated Savings per Account (kWh)	Realization Rate	90% Confidence Bound
All Lighting	3,706	25,888	6,029	23%	8%
Prescriptive Light Without Control	2,547	23,201	4,644	20%	11%
Albany	885	25,831	11,375	44%	9%
Buffalo	1,277	29,373	3,400	12%	19%
Syracuse	1,544	23039	7,219	31%	10%

4.2.2 Evaluated Gross Savings without Comparison Group

For comparison, in Table 17 we also provide evaluated gross savings without the comparison group. The disadvantage of this approach is that there is no control for external changes. Without the comparison group, we find that the overall lighting realization rate improves to about 31% for 2010 and 2011, which may indicate spillover effects among program non-participants. In general, the realization rates are similar between the two program years evaluated.

Table 17: Evaluated Gross Savings without Comparison Group

Year	Level of Analysis	Number of Accounts	Tracking Savings per Account (kWh)	Estimated Savings per Account (kWh)	Realization Rate	90% Confidence Bound
2010	All Lighting	3,744	26,815	8,444	31%	2%
	Prescriptive Light No Control	2,441	23,685	5,606	24%	3%
	Albany	841	21,321	9,882	46%	5%
	Buffalo	1,658	31,189	7,119	23%	4%
	Syracuse	1245	24,700	9,237	37%	3%
2011	All Lighting	3,706	25,888	7,997	31%	2%
	Prescriptive Light No Control	2,547	23,201	5,668	24%	2%
	Albany	885	25,831	11,154	43%	4%
	Buffalo	1,277	29,373	5,584	19%	3%
	Syracuse	1544	23,039	8,184	36%	3%

4.2.3 Evaluated Savings from Pooled Model

In Table 18, we present results from the Pooled Time-Series Cross-Sectional model, as another form of validating the previous estimates. The pooled model is discussed in more detail in Appendix E. This method takes a different approach that controls for monthly and post-installation effects in a single model. The realization rates from the pooled model are about 31% for 2010 and 33% for 2011, which is very similar to the results without the comparison group.

Table 18: Evaluated Gross Savings with Pooled Model

Year	Level of Analysis	Number of Accounts	Tracking Savings per Account (kWh)	Estimated Savings per Account (kWh)	Realization Rate	90% Confidence Bound
2010	All Lighting	3,744	26,815	8,446	31%	6%
2011	All Lighting	3,706	25,888	8,601	33%	5%

4.2.4 Savings for On-sites with and without Additional Accounts

Clearly, the billing analysis results are lower than might be expected based upon the on-site M&V findings and relative to our longstanding experience of how a small business program of this nature is likely to perform. We performed several activities to mitigate potential shortcomings in the billing analysis, as discussed earlier. However, we believe there are three weaknesses that remain that may be downwardly biasing the billing analysis results. The first is the possibility that significant quantities of pre-existing equipment were not operational or functioning at the time of replacement. This would be a "correct" downward bias. In the recommendations section of this report, we recommend that an effort to assess this possibility be incorporated into a lighting controls study that is being undertaken shortly.

The second possibility has been the focus of other evaluators addressing similar results in similar evaluations. These evaluators point to high vacancy rates as possibly explaining the low realization rate. To the extent that program participation coincides with an increase in business activity, the billing analysis will include that increased consumption in addition to the efficiency-related savings in the estimate of pre-post change. An increase in overall consumption would counteract efficiency-related savings at the site level and could downwardly bias the aggregate estimate of savings. A billing analysis estimate is always vulnerable to the potential for bias from non-program change. Those changes may occur in both directions, and the method assumes that, beyond the trends and events controlled for by the technical methods, these effects are approximately zero. It is on this basis that billing analysis has been widely used for decades. The explanation of low billing analysis results for a specific program at this time as a consequence of vacancy rates relies on two key assumptions: That program participants are more likely to be increasing consumption due to business processes and those increases are correlated with the program participation timeframe. These assumptions would need to hold true not just in comparison to historical norms. They would also have to hold true in comparison to previous participants in their second year after participation. That is, to the extent that the increase in consumption continued into the second year post program, then the prior participant comparison group would, in fact, adjust for some or all of that increase. The evidence provided to support these claims, in the form of vacancy rates and the like, is suggestive at best. We acknowledge that increased occupancy or production may be one factor in low realization rates produced with the billing analysis for this evaluation. We are less inclined to think that it can explain the full decrease in estimated savings relative to onsite savings.

The third possibility is that despite all efforts, we might not have successfully gathered the consumption for all program treated areas in the billing analysis. Complete consumption data is a threshold need for a billing analysis and can introduce a fatal flaw in the analysis if not sufficiently comprehensive. This is a tractable issue that we were able to explore for sites that were in our M&V sample, at which we sought additional meters and accounts serving treated spaces while on-site. This allows us to examine this issue through a pre/post normalized analysis of consumption data on the subset of participants where we did on-sites. Due to the nature of this assessment, we are not able to include a comparison group. However, based on earlier results, the adjustment to savings from the comparison group were marginal.

Sixty three of the seventy on-site sample participants had billing data that was sufficient for inclusion in this analysis. Among these, field staff was able to collect additional information and/or confirm meter and account numbers at 37 sites. For the full sample of on-site visit sites also in the billing analysis, as well as two additional subsets of the on-site sample, we examined savings based on pre/post normalized consumption. In Table 19, we show the weighted savings from the different analyses.

The top row in the table represents the 63 on-sites also in the billing analysis. The middle row presents 37 on sites also included in the billing analysis that we refer to as investigated added accounts. At these 37 sites, we were able to confirm the billing analysis meter was present at the site, sought out other meters and added them into the analyzed consumption when possible. As noted earlier, there were many sites where we did not have access to all areas where meters might have been located. We also identified three sites with additional meters that we were unable to identify in the billing data and map into our analysis. The last row in the table presents only those 10 on-sites where additional metered or account data was added to the billing analysis. These are the sites where we were able to identify and map in additional consumption data after all efforts had been undertaken.

The sample sizes in this table are small, by virtue of increasingly stringent criteria for inclusion. So while we proceed with caution in interpreting this information, there seems to be some evidence that as additional consumption data is identified and integrated into consumption based savings estimates, those estimates increase. The approximate improvement in realization rate increases roughly a third, from 26% to 36% for those sites where we were able to add in observed additional meters versus the on-site group as a whole. When we limit our analysis to just the 10 sites for which we added in all observed meters and known data, the billing analysis result rises to a 77% realization rate.

This examination of the consumption and on-site data in this table is suggestive that incomplete billing data may be affecting the ability of the billing analysis to produce reliable savings estimates (as discussed in Table 6 earlier). While this study was able to gather additional meters as part of our on-site work, it was not a primary focus of our site work and we do not feel it was comprehensive enough to be definitive in this assessment. However, as a result of this effort, we are left to wonder if issues in account matching to treated space are confounding our final billing analysis work despite our best efforts to gather all relevant consumption data. In the recommendations section of this study, we recommend that as part of a pre/post M&V study of lighting controls on SBS participants, we examine this issue in greater detail to better understand how this issue might affect billing analyses on this population.

Table 19: Evaluated Gross Savings for On-sites with and without Additional Accounts

Level of Analysis	Number of Participants	Tracking Savings per Account (kWh)	Onsite Savings per Account (kWh)	Onsite Realization Rate	Pre/Post Normalized Savings per Account (kWh)	Pre/Post Realization Rate
Original Onsite Sample	63	22,487	17,859	79.4%	5,912	26%
Investigated Added Accounts	37	19,200	15,237	79.4%	6,950	36%
Only Added Accounts	10	16,313	11,547	71.0%	12,628	77%

4.2.5 Supplemental Survey Questions

As part of our net-to-gross survey work, DNV GL asked supplemental survey questions with program participants to assess if respondents reported changes in certain aspects of their business within a year before or after the time they participated the program. These questions were asked of all NTG respondents from 2011 and the first half of 2012, regardless of whether they were included in the final billing analysis. We asked these questions to help understand any changes that might have impacted participants' electric use at a high level that might help us understand the billing analysis results or perhaps put them into better context. Table 20 below summarizes the responses. The first three change types can be expected to bias the billing analysis down and suggest that square footage, business activity and energy using equipment have all experienced a general net increase among participants. The fourth item (use due to economy) has the potential to bias results upward, however, the use of a comparison group can be expected to net out general economy related trends such as that suggested among respondents. While not definitive, it does seem that in general one might expect these findings to be placing only marginal downward pressure on the billing analysis result as the net changes are relatively small.

Table 20: Reported Change in Activity within a Year of participation in Program

Reported change type (n=497)	Increase	Decrease
Facility square footage	1.4%	0.2%
Business activity	6.8%	3.4%
Energy using equipment	6.8%	3.0%
Energy usage due to economy	1.0%	6.4%
Number of employees	4.0%	4.4%
Hours of operation/shifts	3.0%	3.0%
Some other area of business	1.4%	
No change/Did not know	44.1%	

Respondents were also asked about other changes or one-time events that might have caused a significant impact on energy use during this same period. We took those reported changes and classified them as items that would increase or decrease their consumption. Just over 6% of NTG survey respondents reported changes or one-time events that might have caused a significant impact on their energy use. Table 21below summarizes the types of modifications reported by the respondents. Most changes were witnessed due to addition of usable space, although significant changes were not observed overall.

Table 21: Reported Changes or One-time Events Affecting Energy use

Reported change type (n=497)	Impact on Energy Consumption	Percent Reporting Change
Changes that would be causing upward pressure on realization rate		
Change in equipment leading to decreased energy consumption (ex: replacing existing windows with energy efficient windows, piece of machinery taken out of production line etc.)	Decrease	1.21%
Deletion of usable space (closure of a facility, deletion of office space etc.)	Decrease	0.80%
Changes that would be causing upward pressure on realization rate		
Addition of usable space (ex: addition of warehouse, office space etc.)	Increase	2.01%
Change in equipment leading to increased energy consumption (ex: installation of new fabricating equipment, production equipment etc.)	Increase	0.40%
Changes that would be causing upward or downward pressure		
Climatic factors (ex: extreme events such as hurricane Irene and Sandy, significant climatic changes in temperature, humidity etc.)	Increase or decrease	1.01%
Miscellaneous/Other (ex: tenant based changes, fluctuations in industrial testing facility etc.)	Increase or decrease	0.80%

Although these supplemental questions point to some possible patterns among the participants, we do not feel they are conclusive with respect to how much reported changes in activity or one-time events among participants are affecting the billing analysis. While some questions suggest an increase in energy consumption, others suggest a decrease, all of which makes it difficult to determine the net effect of these reported trends to the analysis.

4.3 Net to Gross

This section of the report presents the NTG results. Recall, we calculated two free ridership rates as part of this study. The first we refer to as the core approach, which is consistent with the Massachusetts standardized approach. The second is the alternative approach, which averages the computation of the initial efficiency and quantity free ridership estimate with timing as compared to multiplying all factors together.

We also note that the spillover cited in these results is “inside like” spillover, meaning that it includes only program attributable savings that are like the measures installed within the facility in which they were installed. The nature of this spillover quantification is that it can be regarded as a conservative estimate as it does not include other spillover possibilities such as spillover associated with different measures in different locations, nor non-participant spillover.

Overall, the results of the NTG analysis estimate that 94 percent of kWh savings from the program were attributable to the program. In addition, the program resulted in two percent spillover. Together, the overall NTG ratio is 96 percent.

Table 22: Overall SBS Program NTG Ratios

	kWh				
			90% Confidence Interval		
Overall Program Ratios	**n**	**NTG Ratio Components**	**+/-**	**Lower Bound**	**Upper Bound**
Attribution Rate	1,057	94%	1%	93%	95%
Spillover Rate	1,055	2%	<0.1%	1%	3%
Combined Net-to-Gross	**1,055**	**96%**	**1%**	**95%**	**97%**

The attribution rate represents the percent of savings installed through the program that were attributable to the program. Attribution is calculated as the inverse of the free-ridership rate. An attribution of 94 percent indicates that program free-ridership is six percent. All but one measure group had attribution rates higher than 90 percent. The SBS program targets small businesses, which is typically viewed as a sector that does not undertake efficiency on its own due to its cost and lack of individuals dedicated to efficiency within their organizations. The attribution rate indicates that the vast majority of savings resulting from this program would not have occurred without the program’s influence. Table 23 below presents the direct attribution by measure group. These rates are stable among the larger measures, with results in a tight range between 93% and 95%. Non-HVAC Motors and Drives has a poorer attribution rate, but due to its small sample size, it also has a particularly wide precision associated with it.

Table 23: Direct Attribution Rate by End-use Category

End-use Category	n	Direct Attribution Rate	Relative Precision at 90% Confidence
Prescriptive Lighting without Controls	668	94.8%	1.3%
Prescriptive Lighting with Controls	313	93.3%	2.6%
Refrigeration	54	93.6%	3.4%
Non-HVAC Motors and Drives	22	78.1%	36.2%
Overall	**1,057**	**94.2%**	**1.2%**

The spillover rate in this study represents the savings installed outside of the program by program participants as a result of their participation in the program. It is expressed as a ratio of savings attributable to program influence. The spillover assessed in this study is termed "inside like spillover"[22], which is a subset of the larger spillover adjustment. Overall the program "inside like spillover" rate was two percent. Installing measures from the Prescriptive Lighting without Controls and Refrigeration measure groups resulted in program participants installing additional measures outside of the program that totaled to an overall estimate of two percent of attributable program savings, respectively.

Table 24: Spillover by End-use Category

End-use Category	n	Direct Attribution Rate	Relative Precision at 90% Confidence
Prescriptive Lighting without Controls	668	2.2%	54.7%
Prescriptive Lighting with Controls	313	0.6%	89.3%
Refrigeration	54	1.2%	92.8%
Non-HVAC Motors and Drives	22	0.1%	175.2%
Overall	**1,057**	**1.7%**	**47.6%**

The combined NTG ratio places attribution and spillover into a single ratio. The overall NTG for the program is 96 percent. Individual measure group ratios range from 78 percent for non-HVAC motors and drives to 97 percent for prescriptive lighting without controls.

[22]Spillover savings is defined as the energy savings associated with energy efficient equipment installed by consumers who were influenced by an energy efficiency program, but without direct financial or technical assistance from the program. Inside like spillover refers to spillover savings of the same measure type as that installed through the program, installed at the same location.

Table 25: Combined Net-to-Gross by End-use Category

End-use Category	n	Direct Attribution Rate	Relative Precision at 90% Confidence
Prescriptive Lighting without Controls	668	97.0%	1.7%
Prescriptive Lighting with Controls	313	93.9%	2.6%
Refrigeration	54	94.8%	3.7%
Non-HVAC Motors and Drives	22	78.2%	36.1%
Overall	**1,057**	**95.9%**	**1.4%**

As stated earlier, we performed a core free ridership approach that was consistent with our original work scope as well as an alternate approach consistent with the manner in which we have recently handled the computation of free ridership in New York on a similar study. For the measures examined in this program, the results between the two approaches are similar. One reason behind this is because this methodology change only affects measures with partial free ridership. In addition, since free ridership is so low in general, respondents with no free ridership tendencies are simply unaffected by the method.

Table 26: Comparison of Core versus Alternate NTG Rates

		Core Free ridership Method		Alternate Method	
End Use Measure Category	n	NTG Ratio	Relative Precision at 90% Confidence	NTG Ratio	Relative Precision at 90% Confidence
Prescriptive Lighting without Controls	668	97%	2%	96%	2%
Prescriptive Lighting with Controls	313	94%	3%	93%	3%
Refrigeration	54	95%	4%	94%	4%
Motors	22	78%	36%	78%	36%
Overall	1,057	96%	1%	95%	2%

Using the gross realization rate and NTG estimate for prescriptive lighting without controls, we are able to calculate overall net savings. Table 27 shows these results. Note that although this table only presents the net savings for prescriptive lighting without controls, the findings of this study will allow National Grid to revise savings for all prescriptive lighting installed in the SBS Program.

Table 27: Prescriptive Lighting without Controls Net Savings

Parameter	Electric Energy (kWh/yr)
Tracked Savings	64,265,935
Gross Savings Realization Rate (RR)	80.3%
Evaluation Net-to-Gross Ratio (NTG)	97%
Net Savings	50,057,379

5. Conclusions and Recommendations

While the net-to-gross estimates from this study appear stable at an overall net-to-gross estimate of nearly 96%, the two methods of assessing gross savings for 2011 lighting measures are divergent and statistically different. Specifically, the M&V based assessment of prescriptive lighting without controls has a realization rate of 80%, while the savings for the same measure group and year from the billing analysis with a comparison group is roughly 20%.

This study design included the performance of on-site M&V work with re-engineering analysis to derive results that have traditionally not been achievable through other methods, including the determination of coincidence factors, kW and kWh Interactive factors, connected kW and kWh realization rates, load shapes, and average hours of use. A billing analysis was also undertaken for purposes of deriving a kWh realization rate and to attempt a deconstruction of the realization rate into various components such as an hours-of-use realization rate, interactive effect factor, etc. This two pronged approach was expected to provide National Grid with all of the factors needed to refine and calculate future tracking system estimates through the on-site with M&V effort, along with an accompanying attempt to assess the ability to derive more granular results from a billing analysis.

Had the two evaluation activities employed in this study provided consistent overall findings, it would have been feasible to perform a synergized, nuanced analysis. However, since the results are divergent, we are left to assess the veracity of each savings estimate. Ultimately, while we took many steps in both the M&V and billing analysis to mitigate bias within state of the art evaluation methods, we are recommending that National Grid utilize the on-site M&V results from this study and perform additional exploration of the two greatest weaknesses we see in the billing analysis work before more fully embracing those findings. The two recommendations we have in this regard are:

1. We recommend that the presence of meters and accounts be fully assessed at all sites included in the upcoming lighting controls study. DNV GL is planning to perform pre and post metering at roughly 70 sites among small businesses in New York for National Grid, which will necessitate multiple visits at each. Given the critical importance of the inclusion of all consumption data in a billing analysis, this coming project provides an opportunity to explore the incidence of multiple accounts at a project, compare this information to that in the tracking system and assess what those findings might mean for this and other billing analyses on the small business population.
2. We recommend that the operation of the pre-existing lighting equipment be quantified as part of the pre metering visits associated with the study referenced above. This effort would provide insight into the likelihood of whether a lack of functioning equipment in the pre-existing condition is negatively impacting the billing analysis results. This effort should be fully documented and include an assessment of what those findings might mean for this and other billing analyses on the small business population.

Using the gross realization rate and NTG estimate for prescriptive lighting without controls, we are able to calculate overall net savings. Table 28 shows these results. Note that although this table only presents the net savings for prescriptive lighting without controls, the findings of this study will allow National Grid to revise savings for all prescriptive lighting installed in the SBS Program.

Table 28: Prescriptive Lighting without Controls Net Savings

Parameter	Electric Energy (kWh/yr)
Tracked Savings	64,265,935
Gross Savings Realization Rate (RR)	80.3%
Evaluation Net-to-Gross Ratio (NTG)	97%
Net Savings	50,057,379

Other Conclusions and Recommendations

The following are recommendations based upon the findings from this study.

- The reliability of a billing analysis is largely dependent upon the availability and relationship of consumption data to program treated spaces. Despite a significant effort to acquire all accounts associated with each participant in the population as well as the on-site sample, we suspect the inability to do this consistently at either level limited the ability to produce a confident billing analysis result. We recommend that a renewed effort be undertaken to have program vendors more thoroughly collect meters and/or accounts that serve program treated areas in the interest of making a billing analysis a more feasible option for consideration when evaluating the small business program in the future.
- Currently, the TRM stipulates using a 0.9 factor as a stand in for NTG in reporting program savings. We recommend consideration of the NTG results from this report for the SBS Program. There are two ways in which one might apply NTG factors from this report. The first is if the factor is applied to tracking savings, under which case the net realization rate from this report can be used. Alternatively, if changes to the TRM are made such that it is believed that the tracking savings become more similar to those estimated as gross savings in this report, then using the combined NTG ratio from this report may be more appropriate for consideration.
- The lighting hours of use assumptions in the TRM appear to be high in general, although only the automotive facility TRM assumed hours falls outside of the 90% confidence interval result. As such, while there is evidence from this study that the TRM hours might be overstated, we

encourage National Grid and the DPS to compile added evidence from other New York PAs to be sure this trend holds across other territories before making TRM revisions on this matter[23].

- The TRM currently assumes a coincident factor of 1.0 for commercial indoor lighting measures, which we regard as a high value. The estimated coincidence factor from this study is 0.566. We recommend consideration of this value for small business direct installation lighting applications or at least encourage National Grid and the DPS to compile additional data from other New York PAs.
- Currently, National Grid uses HVAC interactive estimates of 1.19 for summer kW and 1.07 for annual energy savings. These values are proxies that represent average factors from the table of HVAC interactive factors by Building Type in the TRM. This study suggests that National Grid might consider the use of a kW HVAC Interactive Effect Factor of 111.1%, and a kWh interactive factor of 104.2%, or again work with DPS to combine this data with that from other New York PAs to inform these values.
- We recommend that future studies of energy impacts of the small business program use an error ratio of 0.5 for lighting for-sites metering sample sizing. The error ratio we recommend for targeting demand savings is a 0.2 error ratio.
- If not being performed already, we recommend that National Grid work with Program vendors to ensure that the location of measures installed can be acquired for evaluation purposes.

[23] National Grid uses hours of use vendor estimates at the site and location level at the time of participation, however such adjustments may help other PAs that rely more heavily on the Technical Manual.

A. On-site Measurement, Verification and Analysis Methodology Appendices

A key task in the on-site engineering assessment is the installation of measurement equipment to aid in the development of independent estimates of savings. This appendix reviews that process for the sampled prescriptive lighting without controls on-site sample.

The type of measure influences the measurement strategy used. Lighting, the technology of interest to this study, is a time dependent measure type that runs at a constant load. Mathematically, hour-of-day and day-of-week are usually the most relevant variables in the energy savings analysis of these measures. The primary metering equipment used for this study is time of use lighting loggers, which were installed for a year.

The following section outlines the generic methodology for time-dependent lighting measures.

Monitoring. The time-of-use (TOU) loggers used in this study provide measured hours of use. These small devices use specialized sensors – photocells in the case of lighting measures – to sense and record the dates and times that a device turns on and off. This TOU data will be used to support the evaluation in two key ways:

1. To develop summer peak coincidence factors, and
2. To develop annual hours of use.

The measure scope influenced the appropriate number of loggers and systems monitored for each site. Factors that drove the number of installed loggers include the number of unique schedules at the site, and the anticipated level of variation among the schedules within a particular space type.

Verification. A detailed inventory was performed for each installed measure. This inventory included a verification of the quantity and technologies installed from the program, as well as customer reported operating hours for specific equipment, pre-existing equipment types and location of installed equipment. Methods of control were also be examined and inventoried at this time. Other variables that were gathered and analyzed include the types of heating and cooling systems serving the areas of the installed measure for the calculation of interactive HVAC effects.

Analysis. This section serves as a detailed example that illustrates the calculation of all savings and adjustment factors. DNV GL modified a single line item from a site in a similar study to serve as an example of the calculation methods. The table below presents a summary of all savings parameters for this particular example.

Table 29: Calculation Example Result Summary

Parameter	Annual KWH	Difference %	Connected kW	Difference %
Gross (TRACKING) kWh/Connected kW Savings	3,690	N/A	0.74	N/A
Adjustment - Documentation Change	0	0%	0.00	0%
Adjustment - Technology Change	0	0%	0.00	0%
Adjustment - Quantity Change	-410	-11%	-0.08	-11%
Adjustment - Operation Change	543	15%	N/A	N/A
Non-Interactive Savings	3,823	104%	0.66	89%
Adjustment - Cooling Interaction	314	9%		
Adjusted Gross (ONSITE) Savings	4,136	112%		

Parameter	On-Peak Summer kW	Difference %	On-Peak Winter kW	Difference %
Connected Demand Savings	0.66	N/A	0.66	N/A
Adjustment - On-Peak Coincidence	-0.12	-18%	0.00	0%
Non-Interactive Savings	0.54	82%	0.66	100%

Parameter	On-Peak Summer kW	Difference %	On-Peak Winter kW	Difference %
Non-Interactive Savings	0.54	N/A	0.66	N/A
Adjustment - HVAC Interaction	0.14	27%	0.00	0%
Adjusted Gross (ONSITE) Savings	0.68	127%	0.66	100%

Table 30 presents the pre-retrofit condition for this space as outlined in the application documentation. The pre-retrofit condition included (18) 2F40SSS fixtures rated at 94 watts each. The application also assumed 5,000 annual operating hours.

Table 30: Tracking Pre-Retrofit Condition

Qty	Lighting Fixture Code	Fixture Type	Fixture Description	W/Fixt	Hours of Operation per Year
18	2F40SSS	2L4’ STD/STD	Four Foot T12 Systems	94	5,000

Table 31 represents the proposed condition according to the tracking system. In this case, the pre-retrofit fixtures were to be replaced with (18) 2F32EEE fixtures rated at 53 watts each. The hours of operation in the proposed condition were also 5,000 annual operating hours.

Table 31: Tracking Proposed Condition

Qty	Lighting Fixture Code	Fixture Description	Fixture Type	W/Fixt	Hours of Operation per Year
18	2F32EEE	2L4' T8EE/ELEE	Four Foot T8 HP/RW Systems	53	5,000

The first step of the savings analysis was to recreate the savings calculations based upon project documentation. This was done to isolate any documentation adjustments.

Documentation Adjustments

Documentation adjustments reflect any change in savings due to discrepancies in project documentation. Evaluators recalculated the tracking estimates of savings using all quantities, fixture types/wattages, and hours documented in the project file. All tracking system discrepancies and documentation errors are reflected in this adjustment. The documentation adjustments are calculated according to the following formulae:

DOC KWH ADJ = Recreated Tracking kWh Savings – Tracking kWh Savings = 3,690 - 3,690 = 0 kWh

DOC KW ADJ = Recreated Tracking kW Savings – Tracking kW Savings = 0.74 – 0.74 = 0 kW

Hours of Use and Coincidence

The first on-site task was establishing the customer's holiday and vacation/shutdown schedule. Table 32 shows the input for the site holiday analysis. In this particular case, the site contact informed the evaluating engineer that the facility was closed during 6 major holidays. He also stated that the facility does not have any long shutdowns.

Table 32: Input for Site Specific Holidays

Holiday	Date	Site Observed Holidays
New Year's Day	1/1/2012	Y
Martin Luther King Day	1/16/2012	N
Presidents Day	2/20/2012	N
Good Friday	4/6/2012	N
Memorial Day	5/28/2012	Y
Independence Day	7/4/2012	Y
Labor Day	9/3/2012	Y
Columbus Day	10/8/2012	N
Veteran's Day	11/11/2012	N
Thanksgiving Day	11/22/2012	Y
Day After Thanksgiving	11/23/2012	N
Christmas Eve	12/24/2012	N
Christmas Day	12/25/2012	Y

To determine the annual operating hours from monitoring lighting logger data, engineers examine the hourly percent run time across the entire monitoring period.   For this study, lighting logger data was adjusted for the daylight savings time change that occurred within the monitoring period.

For the three month logger data analysis, an 8x24 profile (Monday through Friday plus Holiday by hour-of-day) is generated using a computer program to represent the average percentage of time that the fixture operated during the monitoring study. Table 33presents the profile of the logger used for this example.

Table 33: Logger Profile Summary

Hour Ending	Sun	Mon	Tue	Wed	Thu	Fri	Sat	Hol
1	33.4	36.0	48.5	45.4	45.5	52.7	44.8	57.3
2	33.9	36.0	41.8	47.0	41.4	49.3	45.8	50.9
3	32.0	36.0	49.5	38.6	41.8	38.9	40.5	50.0
4	32.0	36.3	41.1	35.6	35.3	37.1	36.5	50.0
5	57.1	39.4	39.7	36.2	33.5	37.4	49.4	50.0
6	33.8	56.7	54.1	52.8	83.6	49.7	35.3	50.0
7	33.6	75.3	89.1	66.3	94.3	65.9	38.6	50.0
8	35.0	98.7	99.8	99.5	99.0	98.8	46.9	51.7
9	36.6	100.0	100.0	99.9	100.0	100.0	51.1	56.8
10	37.9	100.0	100.0	100.0	100.0	100.0	58.2	50.0
11	34.7	100.0	100.0	100.0	99.8	100.0	52.8	50.0
12	36.6	99.6	100.0	100.0	99.9	100.0	53.3	50.0
13	35.7	98.4	100.0	100.0	99.5	100.0	44.7	50.0
14	34.6	99.1	100.0	100.0	98.8	99.9	42.7	50.0
15	33.5	100.0	100.0	100.0	100.0	100.0	47.8	50.0
16	36.6	94.1	91.6	94.0	91.9	90.2	42.6	50.0
17	34.2	71.7	66.2	64.2	65.5	61.0	41.8	50.0
18	36.1	80.6	83.8	83.3	80.1	83.3	36.8	50.0
19	37.2	94.4	95.2	95.5	95.1	94.8	35.3	50.0
20	33.7	94.7	88.9	92.9	93.6	96.6	34.5	50.0
21	32.0	98.3	96.2	94.5	96.5	96.5	36.9	50.0
22	33.1	95.6	91.5	88.4	86.9	73.0	34.8	50.0
23	32.2	76.5	70.6	67.7	60.7	65.3	34.0	50.0
24	32.7	51.2	48.4	46.1	46.2	54.9	42.2	50.0

For the 12 month logger data analysis, the actual 8,760 hourly percentage of time that the fixture operated during the monitoring study was used directly. This analysis concluded that this fixture operates 5,827 hours per year, of which 67% of these operating hours occur coincide with the defined on-peak period definition. The on-peak summer and winter coincidence factors are 82% and 100%, respectively.

Non-Interactive On-Site Savings

Table 34 represents the on-site installed condition as found the evaluation team. For this example, the evaluator identified (16) 2F32EEE fixtures, which was two fewer fixtures than in the project documentation. A schedule identification number ("1" in this example) maps the hours of operation and the summer and winter coincidences into this spreadsheet.

Table 34: On-Site Installed Condition

Qty	Lighting Fixture Code	Fixture Description	Fixture Type	W/Fixt	Schedule Number	Hours of Operation per Year	On-Peak Summer Coincidence	On-Peak Winter Coincidence
16	2F32EEE	2L4' T8EE/ELEE	Four Foot T8 HP/RW Systems	53	1	5,827	82%	100%

The on-site pre-retrofit condition, presented in Table 35, was established through review of project documents, discussion with facility personnel, and observational inference. This lighting fixture savings analysis presumes that the operating hours did not change between the pre- and post-retrofit conditions and that the installed condition had two less fixtures than were captured in the tracking system, which results in the assumption there were also two less fixtures in the pre retrofit condition.

Table 35: On-Site Pre-Retrofit Condition

Qty	Lighting Fixture Code	Fixture Description	Fixture Type	W/Fixt	Hours of Operation per Year
16	2F40SSS	2L4' STD/STD	Four Foot T12 Systems	94	5,827

Table 36 presents the adjusted gross on-site savings for this example.

Table 36: Adjusted Gross On-Site Savings

Connected kW Savings	kW Summer Savings	kW Winter Savings	kWh Savings
0.656	0.536	0.656	3,823

Heating and Cooling Interaction

Heating and cooling interaction was calculated for each line item where applicable based on the specific HVAC systems serving the space. When lighting equipment converts electrical energy to light, a significant amount of that energy is dissipated in the form of heat. Energy efficient lighting measures convert more electrical energy to light and less to heat. Since installing energy efficient lighting adds less heat to a given space, a complete estimation of energy savings considers the associated impacts on the heating and cooling systems or "interactive effects."

The interactive effects take into account the effect of the energy efficient lighting measures on their corresponding heating and cooling systems. Energy efficient lighting serves to reduce the heat gain to a given space and accordingly reduces the load on cooling equipment. But this reduced heat gain has the added consequence of increasing the load on the heating system.

As part of the on-site methodology, evaluators interviewed facility personnel to ascertain the cooling and heating fuel, system type, and other information with which to approximate the efficiency of the HVAC equipment serving the space of each lighting installation. The DNV GL team expresses HVAC system efficiency in dimensionless units of Coefficient of Performance (COP), which reflects the ratio of work performed by the system to the work input of the system. Table 37 details the COP assumptions for general heating and cooling equipment types encountered in this study. Where site specific information yields improved estimates of system efficiency, these were used in place of the general assumptions below.

Table 37: General Heating and Cooling COP Assumptions

Cooling System Type	COP		Heating System Type	COP
Packaged DX	2.9		Air to Air Heat Pump	1.5
Window DX	2.7		Electric Resistance	1
Chiller <200 Ton	4.7		Water to Air Heat Pump	2.8
Chiller >200 Ton	5.5			
Air to Air Heat Pump	3.9			
Water to Air Heat Pump	4.4			
Refrigerated Area (high temp)	1.4			
Refrigerated Cases (low temp)	1.9			

Interactive effects are calculated only at sites where heating or cooling systems are in use. Leveraging the 8,760 profile of hourly demand impacts, the DNV GL team computes electric interactive affects during the hours that lighting and HVAC are assumed to operate in unison.

DNV GL utilizes Typical Meteorological Year 3 (TMY3) hourly dry-bulb temperatures for the weather station closest to each facility as the balance point criteria in this analysis. For each hour in a typical year, DNV GL computes HVAC interaction according to the following equations:

Cooling kW Effects = 80% * Lighting kW Savings / Cooling System COP

Heating kW Effects = -80% * Lighting kW Savings / Heating System COP

The 80% values represent the assumed percentage of the lighting energy that translates to heat which either must be removed from the space by the air conditioning system or added to the space by the heating system during the aforementioned HVAC hours. This assumption is consistent with those established and employed in previous impact evaluations of custom lighting measures. Also, heating factors are negative because heating interaction erodes gross lighting savings, while cooling interactive boosts it.

B. Description of Results and Factors

This section presents a listing of realization rate and savings factors that were produced as part of this study. Each entry contains a description of that savings variable. At the end of this section, we provide the savings algorithms that can be used to help guide the application of these study results to the National Grid Tracking system savings estimates.

B.1 Realization Rates

Annual KWh – This result is the gross annual kWh realization rate *not* including additional savings due to HVAC interactive effects. This realization rate is the evaluation gross annual kWh savings (excluding HVAC) divided by the tracking gross annual kWh savings.

Annual KWh with HVAC – This result is the gross annual kWh realization rate including additional savings due to HVAC interactive effects. This realization rate is the evaluation gross annual kWh savings (including HVAC) divided by the tracking gross annual kWh savings.

Connected KW – This result is the gross connected kW realization rate, which includes any documentation, quantity, and technology adjustments. This realization rate is the evaluation gross connected kW savings divided by the tracking gross connected kW savings.

B.2 Savings Factors

Summer Coincidence Factor – Diversity x Coincidence. This is the percentage of the connected kW savings coincident with the summer peak period.

Summer kW HVAC Interactive Effect – This is an adjustment factor applied to the gross connected kW savings that are due to interactive effects during the summer on-peak period.

Summer KWh HVAC Interactive Effect – This is this is an adjustment factor applied to the gross kWh savings that are due to interactive effects

Summer On Peak KWh % – This is the percentage of energy savings that occur during on-peak hours.

Table 38: Summary of Results and Factors

Tracking	
(a)	Annual kWh
(b)	Connected kW
(c)	Average Hours of Use
Evaluation	
(d)	Annual kWh
(e)	KWh HVAC Factor
(f)	On-Peak % Annual kWh
(g)	Connected kW
(h)	Summer kW Coincidence Factor
(i)	Summer kW HVAC Factor
(j)	Average Hours of Use
Realization Rates	
(k)	Annual kWh (Excluding HVAC)
(l)	Annual kWh (Including HVAC)
(m)	Connected kW
(n)	Average Hours of Use

Savings Algorithms	
Evaluated Annual kWh Savings	(a) x (k) x (e)or(a) x (l)
Evaluated Connected kW	(b) x (m)
Evaluated Summer Peak kW Reduction	(b) x (m) x (h) x (i)

C. Ratio Expansion

This appendix provides the specific ratio estimation computations DNV GL employed to develop estimates of evaluation verified gross and net impacts.

Ratio Estimation

DNV GL used the statistical procedure of ratio estimation to develop estimates of evaluation verified gross and net impacts. There are two basic steps in the process. The first step is to verify energy savings in a sample of measures. DNV GL accomplished this first step via CATI surveys and on-site visits. The second step is to expand the sample results to the population of measures. This is accomplished by calculating the ratios of verified-to-reported and attributable-to-verified for the sample. The ratios are also referred to in this discussion as adjustment factors or rates.

Expansion of sample results to the population via ratio analysis

The calculation of the adjustment factors for tracking system gross and net savings uses appropriate case weights corresponding to the sampling rate. The four directly calculated adjustment factors are the installation rate, the calculation adjustment, the operating hour's adjustment, and the net-to-gross factor.

Each of these is calculated as a ratio estimator over the sample of interest (Cochran, 1977, p.165). The formulas for these factors are given below.

Notation: The following terms are used in calculating the adjustment factors:

G_{TRKj}= tracking estimate of gross savings for measure *j*
G_{DOCj}= tracking estimate of gross savings for measure *j*, adjusted for documentation discrepancies
G_{TECHj}= tracking estimate of gross savings for measure *j*, adjusted for documentation and technology discrepancies
G_{QTYj}= tracking estimate of gross savings for measure *j*, adjusted for documentation and technology discrepancies and non-installation (quantity discrepancies)
G_{HRSj}= tracking estimate of gross savings for measure *j*, adjusted for documentation and technology discrepancies, non-installation (quantity discrepancies) and evaluation logged operating hours

G_{VERj}= verified gross savings for measure *j*, including HVAC interactive effects
N_{Vj}= net savings determined from the CATI surveys.
W_{Oj}= weighting factor for measure j used to expand the on-site sample to the full population
W_{Cj}= weighting factor for measure j used to expand the CATI sample to the full population

Documentation Adjustment

The documentation adjustment R_{DOC} is calculated using the on-site sample as

$$R_{DOC} = \frac{\sum_{j \in O} G_{DOCj} w_{Oj}}{\sum_{j \in O} G_{TRKj} w_{Oj}}$$

Technology Adjustment

The documentation adjustment R_{TECH} is calculated using the on-site sample as

$$R_{TECH} = \frac{\sum_{j \in O} G_{TECHj} w_{Oj}}{\sum_{j \in O} G_{DOCj} w_{Oj}}$$

Quantity Adjustment

The documentation adjustment R_{QTY} is calculated using the on-site sample as

$$R_{QTY} = \frac{\sum_{j \in O} G_{QTYj} w_{Oj}}{\sum_{j \in O} G_{DOCj} w_{Oj}}$$

Operational Adjustment

The documentation adjustment R_{HRS} is calculated using the on-site sample as

$$R_{HRS} = \frac{\sum_{j \in O} G_{HRSj} w_{Oj}}{\sum_{j \in O} G_{QTYj} w_{Oj}}$$

HVAC Interaction Adjustment

The interactive adjustment R_{HVAC} is calculated using the on-site sample as

$$R_{HVAC} = \frac{\sum_{j \in O} G_{VERj} w_{Oj}}{\sum_{j \in O} G_{HRSj} w_{Oj}}$$

Net-to-Gross Ratio

The net-to-gross ratio R_A is calculated from the CATI sample as[24]

$$R_A = \frac{\sum_{j \in C} N_{Vj} w_{Cj}}{\sum_{j \in C} G_{VERj} w_{Cj}}$$

Standard errors

---

[24] For the net-to-gross ratio, the verified gross savings for measures in the CATI survey (G_{VERj}) were estimated based on the gross savings adjustments found for measures of the same measure type in the on-site sample. For measure types in the CATI that were not represented in the on-site sample, verified gross savings were assumed to be equal to tracked savings in calculating the ratio.

The ratio estimator is calculated using a SAS® macro provided by SAS for ratio estimation by domains. The procedure also returns the standard error of the estimate.

The standard error is calculated recognizing the sample as drawn from a finite population: the measures completed within the analysis period with associated energy impacts in the program-tracking database. This calculation uses the Finite Population Correction (FPC) factor. This factor is a reduction to the calculated variance that accounts for the fact that a relatively large fraction of the population of interest has been observed directly and is not subject to uncertainty. It is appropriate to apply precision statistics, such as confidence intervals, based on the standard error calculated in this manner when quantifying the results of the program during the study period only.

Combined Ratio Estimators

The Gross Realization Rate, $R_{E,}$ is calculated by chaining together the installation rate and the calculation adjustment:

$$G_T$$

This is an example of a chained ratio estimator using a nested sample. The standard error for the chained ratio is approximated by the formula:

$$SE(ABCDE) \approx ABCDE\sqrt{\left[\left(\frac{SE(A)}{A}\right)^2+\left(\frac{SE(B)}{B}\right)^2+\left(\frac{SE(C)}{C}\right)^2+\left(\frac{SE(D)}{D}\right)^2+\left(\frac{SE(E)}{E}\right)^2\right]}.$$

(This formula overstates the standard error, because it ignores the correlation between the numerator of each ratio and the denominator of the ratio to its right, which reduces the variance of the product.)

Likewise, the net savings realization rate, R_N, is calculated by chaining together the gross savings realization rate and the net-to-gross ratio:

$$R_N = R_{VER} R_A$$

The same standard error approximation formula allows (an over-estimate of) the standard errors of each of the realization rates to be calculated from the two separate standard errors. The standard error for the chained ratio is approximated by the formula:

$$SE(AB) \approx AB\sqrt{\left[\left(\frac{SE(A)}{A}\right)^2+\left(\frac{SE(B)}{B}\right)^2\right]}.$$

D. Net to Gross Battery and Algorithm

This appendix excerpts the key free ridership and participant like spillover questions posed in the survey and reviews how those responses were utilized in the estimation of net to gross by measure.

Free-ridership questions

After confirming the proper respondent was on the phone, the instrument inquired about what influence, if any, the program had on the decision to install equipment through the program. As there are several dimensions to the decision to purchase and install new equipment, the battery discusses the timing of the installation and the quantity and the efficiency level of the equipment installed. These questions reference both the overall effect of the program (including audit recommendations and technical assistance) and the specific effect of the financial incentive. The questions are listed below. Please note that these questions are measure-specific and are repeated for up to two measure categories.

FR5 I'd like to go over all the assistance you received from National Grid. According to our records, National Grid paid up to 70 percent of the total cost of the energy efficient <MEASURE CATEGORY> project implemented through the program.

If (R6<>3): In addition, as I previously mentioned, National Grid paid for the energy audit you received.

[IF R9 = YES: National Grid also provided interest-free financing for up to 24 months for your portion of the project costs.]

If National Grid had not paid up to 70 percent of the implementation cost [IF (r6<>3): OR provided the energy audit] [IF R9 = YES: or provided interest-free financing], would your business have implemented <u>any type</u> of < MEASURE CATEGORY> project <u>at the same time</u>?

FR6A Would you have implemented the <MEASURE CATEGORY> project earlier than you did, at a later date, or never?

FR6B How much [EARLIER/LATER] would you have implemented the <MEASURE CATEGORY> project?

FR7A Without the program incentive would your business have implemented the <u>exact same quantity</u> of < MEASURE CATEGORY> equipment…

IF FR5=YES or DK (OR FR5=2 AND FR6A=3): at that same time?

IF FR5=2 AND FR6A=3: within (TIMEFRAME IN FR6b)] of when you implemented it through the program?

FR7B Compared to the amount of <MEASURE CATEGORY> that you implemented through the program, what percent of the project do you think your business would have purchased on its own during that timeframe?

FR8 You said your business would have installed [IF FR7A=YES: all; IF FR7A= NO: (FILL WITH FR7B %), IF FR8 = DK/R, FILL IN WITH "some"] of the equipment on its own if the program had not been available.

Thinking about the <MEASURE CATEGORY> equipment you would have installed on your own, what percent of this equipment would have been:

a. of the same high efficiency as what was installed through the National Grid program?
b. lower efficiency than what was purchased but higher than standard efficiency or code?
c. standard efficiency or code

Consistency check questions

The instrument also included questions that would identify and correct inconsistent responses. For example, if participants reported that they were likely to install the equipment without the program but also reported that they would not have installed the energy efficient equipment within two years, the interviewer asked them to confirm which statement was more accurate. These questions are listed below.

FR1 On a scale of 0 to 10, with 0 being not at all likely and 10 being very likely, how likely is it that your business would have implemented the same [IF QUANTITY > 1: quantity] and efficiency of <MEASURE CATEGORY> at that same time if National Grid had not provided <ALL ASSISTANCE>?

C3 On a scale of 0 to 10, with 0 being no influence and 10 being a great deal of influence, how much influence did the incentive you received from National Grid have on your decision to implement the high efficiency <MEASURE CATEGORY> project?

C4A Now I want to focus on what it would have cost your business to install this equipment on its own without the program. On a scale of 0 to 10, with 0 being not at all likely and 10 being very likely, how likely is it that your business would have paid the additional cost of the equipment on top of the amount you already paid, to implement the same quantity and efficiency of <MEASURE CATEGORY> equipment at that same time?

C8A (ASK IF AT LEAST SOMEWHAT LIKELY TO HAVE INSTALLED THE MEASURE WITHOUT THE PROGRAM BUT LATER STATES WOULD HAVE WAITED AT LEAST TWO YEARS (FR1 > 3 AND FR6b > 24 MONTHS OR NEVER) and FR5<>1) Earlier in the interview, you said there was a (FR1 SCORE) in 10 likelihood that you would have implemented the same quantity and efficiency of < MEASURE CATEGORY> equipment at that same time in the absence of the National

Grid program assistance. But you also said you would not have implemented the < MEASURE CATEGORY> project within 2 years of when you did. Which of these is more accurate?

1. The likelihood of installing this without the program assistance was (FR1 SCORE)
2. Would not have installed anything within 2 years
3. Something else (SPECIFY)

C9C I'd like to better understand your purchase decision. In your own words, please describe what impact, if any, all the assistance you received through the program had on your decision to install the amount of energy efficient <MEASURE CATEGORY> equipment at the time you did?

As inputs into the algorithm, Tetra Tech constructed a scoring system based on the influence and consistency check questions above. The scoring calculates two scores: a quantity score and an efficiency score. The quantity score represents the percentage of the incentivized equipment that would have been installed in absence of the program. The efficiency score is the percentage of savings *per unit installed* that would have occurred without the program. For equipment that is reported to be more efficient than standard but less efficient than what was installed through the program, we assume 50 percent of the savings for those measures. Multiplying these two scores together gives the percent of the incentivized savings that would have occurred without the program. This percentage is the raw free-ridership estimate. Table 39 details these calculations.

Table 39: Quantity and Efficiency Scores

Score	Responses	Result
Quantity Score (FR_QTY)	If would have installed same quantity without program (FR7A = YES)	FR_QTY = 1
	If would have installed fewer quantity without program (FR7A = NO)	FR_QTY = FR7B
	If never would have installed (FR6A = never)	FR_QTY = 0
Efficiency Score (FR_EFF)	If would have installed at least some equipment on their own	FR_EFF = FR8A + (FR8B*.50)
	If never would have installed (FR6A = never)	FR_EFF = 0
Initial Free-ridership Score	The percent of the rebated savings that would have occurred without the program.	FR_EFF * FR_QTY

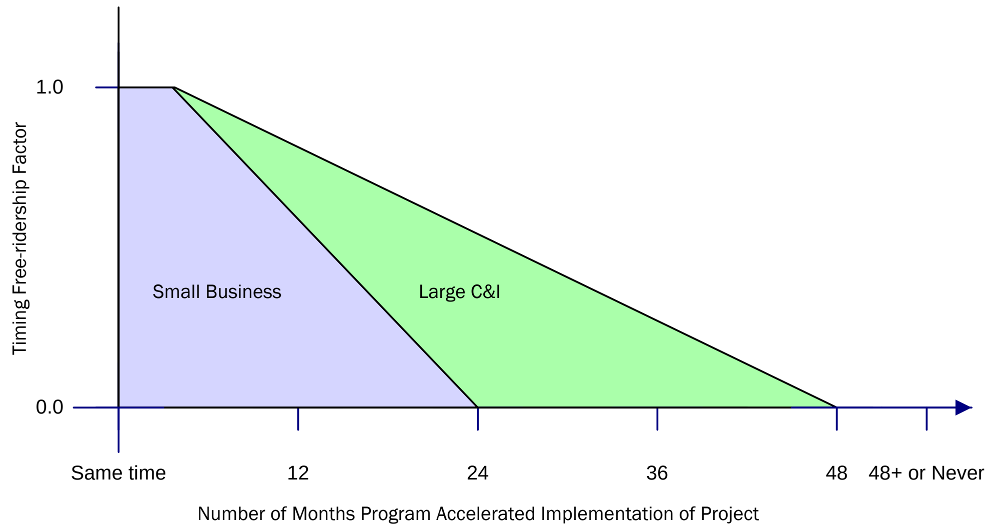
The product of these two scores is then adjusted by a timing factor. The timing factor adjusts the raw free-ridership estimate downward for all or part of the savings that would have occurred without the program, but not until much later. By doing so, the program is given credit for accelerating the installation of energy efficient equipment. For example, if the participant states that he or she would have installed equipment at the same time regardless of the program, the quantity-efficiency factor is not adjusted. However, if the participant states that, without the program, they would have completed the project more than six months later than they actually did; any free-ridership identified in the quantity-efficiency factor is adjusted downward[25]. The degree of the adjustment depends on the program. As the equipment planning schedule for small businesses is likely shorter than the planning schedule for large businesses, small business programs receive a greater acceleration benefit. This reduced adjustment for small businesses reflects the increased effect the program has on the planning schedule. This adjustment is detailed in Table 40 and visualized in Figure 7. In the NY Method, Timing free-ridership is averaged with the initial free ridership score rather than multiplied. This calculation is provided in the last row of this table.

Table 40: Timing Factor Adjustment

Score	Responses	*Result*
Timing Factor—Small Business Programs (FR_TIMING)	Would have installed at the same time without the program (FR5 = Yes)	*FR_TIMING = 1*
	Would have installed within six months of when participant actually did without the program (FR6b <= 6 months)	*FR_TIMING = 1*
	Would have installed sometime between 7 and 24 months of when participant actually did without the program (FR6b > 6 months &< 24 months)	*FR_TIMING = 1-((FR6B-6) * .056)*
	Would have installed sometime after 24 months of when participant actually did without the program (FR6b > 24 months)	*FR_TIMING = 0*
	Would have never installed without the program (FR6A = Never)	*FR_TIMING = 0*
Core Method: Adjusted Free-ridership Score	*The raw free-ridership estimate adjusted for all or part of the savings that would have occurred without the program, but not until much later*	*FR_TIMING * Initial Free-ridership Score*
NY Method: Adjusted Free-ridership Score	*The raw free-ridership estimate adjusted for all or part of the savings that would have occurred without the program, but not until much later*	*(FR_TIMING + Initial Free-ridership Score) / 2*

[25] Projects that were accelerated by fewer than 6 months are not adjusted. As installation timelines are subject to shifting, we assume these projects are just as likely to have been installed at the same time.

Figure 7: Timing Free-ridership Factor by Number of Months the Program Accelerated Implementation



1. Influence of technical assessment

The initial free-ridership score is further adjusted by the influence of any program-sponsored technical assistance or audit and by the influence of previous program participation. If a participant rates the influence of the technical assistance as high (7 or greater on a scale of 0-10), the free-ridership score is reduced by half. This reduction is necessary because the previous factors focus on the specific effect of the program incentive and the overall effect of the program. Without this adjustment, the influence of the technical assessment is under-represented.

C2 On a scale of 0 to 10, with 0 being no influence and 10 being a great deal of influence, how much influence did the information provided by the energy audit have on your decision to implement the high efficiency <MEASURE CATEGORY> project?

2. Influence of past program participation

Likewise, if a participant has previously participated in the program, they are asked about the influence of that past participation on their perceptions and behaviors. Participants are asked to state whether they agree or disagree with four statements about the effect past participation has had on their decision-making. Based on the number of statements with which they agree, their free-ridership is reduced by 75 percent, 37.5 percent, or not reduced at all. This reduction is done to account for the influence positive program experiences have had on participants' purchasing decision – with the program administrators, implementers, or the equipment incented.

PP3 I'm going to read you several statements. For each statement, please tell me whether you agree or disagree that this statement applies to your business. There are no right or wrong answers; we just want your honest opinion. Our previous experience implementing energy efficient projects through a National Grid program. . . .

a. Has made our firm more likely to consider energy efficient equipment
b. Has made our firm more likely to install energy efficient equipment

c. Has given us more confidence in the financial benefits of energy efficient equipment
d. Has given us more confidence in the nonfinancial benefits of energy efficient equipment

As mentioned previously, the previous program participation adjustment is made to account for the market effects associated with implementing energy efficiency programs over time. These market effects will result in net savings estimates that do not capture the full cumulative effect of the program. This methodology attempted to capture some of these market effects by making this adjustment for previous program participation. While it could be argued that the influence of previous participation should count as spillover rather than reduced free-ridership, the traditional definition of spillover does not count measures installed through a program as spillover. Table 41 details these adjustments.

Table 41: Adjustments for the Influence of Technical Assessments and Previous Participation

Adjustment	Responses	Result
Technical Assessment Adjustment	No technical assessment, audit, or study conducted	No adjustment
	Participant would have performed assessment, audit, or study without program assistance or it was not influential (C2 ≤ 6)	No adjustment
	Participant **would not** have performed assessment, audit, or study without program assistance and it was influential (C2 > 6)	Adjusted Free-ridership Score * .5
Previous Participation Adjustment	No previous participation in program	No adjustment
	Agrees with four statements regarding the positive influence of past participation (PP3)	Adjusted Free-ridership Score * .25
	Agrees with three statements regarding the positive influence of past participation (PP3)	Adjusted Free-ridership Score * .625
	Agrees with two or fewer statements regarding the positive influence of past participation (PP3)	No adjustment

<u>Spillover Questions and Algorithm</u>

The "like" spillover estimates are computed based on how much more of the same energy efficient equipment the participant installed outside the program that were, in fact, influenced by the program. This is a conservative approach because it assumes the exact same equipment, including efficiency level and size. The following questions, in conjunction with the savings assigned to that same equipment by the program, are used to estimate possible spillover savings:

S1A Now I'd like you to think of the time since you participated in the Small Business Services program in <YEAR>. Has your company implemented any <MEASURE CATEGORY> projects for this or other facilities in New York **on your own,** which is without a rebate from National Grid?

S1B Was this equipment of **the same or higher level of efficiency** as the equipment you installed through the program, or was it of a **lower level of efficiency** as the equipment you installed through the program?

S1C Was this equipment more energy efficient than standard efficiency or code equipment?

S2A Compared to the amount of <MEASURE CATEGORY>that you installed through the program at <ADDRESS>, how much <MEASURE CATEGORY>equipment did you install on your own?

For respondents that answer “Yes” to S1A and “Same or higher” to S1B, spillover savings are calculated as the measure-specific savings identified by the program multiplied by the quantity identified in S2A. For respondents that answer “Yes” to S1A and S1C, spillover savings are calculated as 50 percent the measure-specific savings identified by the program multiplied by the quantity identified in S2A. If the respondent answers “No” to S1A or S1C, there are no identifiable “like” spillover savings.

For those measures, a program-attributable spillover rate is then calculated based on the following questions:

S3A Did a recommendation by the contractor, engineer, or designer who you worked with under the Small Business Services program influence your decision to implement some or all of this efficient <MEASURE CATEGORY> equipment on your own?

S3B Did your experience with the energy efficient projects implemented through the Small Business Service program influence your decision to implement some or all of this efficient <MEASURE CATEGORY> equipment on your own?

S3C Did your participation in any past program offered by National Grid influence your decision to implement some or all of this efficient <MEASURE CATEGORY> equipment on your own?

S3D On a scale of 0 to 10, where 0 is “no influence at all” and 10 is “a great deal of influence”, how much influence did your participation in the National Grid program have on your decision to install this equipment without an incentive?

S4a Why didn’t you implement this <MEASURE CATEGORY> project through a National Grid program?

S4b [IF THE EQUIPMENT WOULD NOT QUALITY] Why wouldn’t the equipment qualify?

If the respondent reports that the contractor influenced their decision to install the like equipment on their own, we attribute the program with 50 percent of those savings based on the influence the program has on the trade allies. If the respondent reports that either their experience with the program-sponsored project or past programs influenced their decision to implement the like equipment, we attribute the program with 100 percent of the spillover savings. Equipment that would not qualify for a National Grid program or was installed outside National Grid’s service territory is not attributable and receives a spillover rate of 0 percent.

To summarize:

If (S3A=yes AND (S3B = no AND S3C = no)), spillover rate = 50%.

If (S3B=yes OR S3C = yes), spillover rate = 100%.

That rate, applied to the estimated spillover savings, results in the program-attributable spillover savings for that participant for that measure category.

E. Billing Analysis Methods

This appendix provides the specifications of the models used for the two different billing analysis approaches pursued in this report.

<u>Site-Level Approach</u>

The site-level approach to billing analysis follows the PriSM approach[26]. Consumption data for the pre- and post-installation periods for each site are modeled separately. The model is flexible to the presence of both heating and cooling load and changing set points for both of those processes. The models facilitate the estimation of a normalized annual consumption (NAC) for the pre-and post-installation periods. The NAC transforms time- and weather-dependent billing data into a uniform annual measure of consumption that reflects consumption under typical weather conditions in both pre- and post-installation periods. The difference between the pre- and post-installation estimates represents the pre-/post- change associated with the program participation. For comparison group members, the same process is completed around a "non-installation" date. This difference captures the pre-/post- difference of a site where there was no program participation.

The site-level pre-/post- differences, for both participants and the comparison group, are combined in second-stage regression that estimates participation effects where comparison group change is controlled for. The second-stage models can include measure-level program gross savings to provide a richer characterization of program effects.

Stage One

The degree-day regression for each premise and year (pre or post) is modeled as:

$$E_m = \mu + \beta_H H_m + \beta_C C_m + \varepsilon_m$$

where,

E_m = Average consumption per day during interval *m*

H_m = Specifically, $H_m(\tau_H)$, average daily heating degree days at the base temperature(τ_H) during meter read interval m, based on daily average temperatures over those dates

[26] PRisM (Advance Version 1.0) Users' Guide. Fels, M.F., and K. Kissock, M.A. Marean and C. Reynolds. Center for Energy and Environment Studies, Princeton New Jersey. January 1995.

C_m	=	Specifically, $C_m(\tau_C)$, average daily cooling degree days at the base temperature(τ_C) during meter read interval m, based on daily average temperatures over those dates
μ	=	Average daily base load consumption estimated by the regression
$\beta_{H,}\beta_C$	=	Heating and cooling coefficients estimated by the regression
ε_m	=	Regression residual

This model is estimated across a range of degree-day bases and without degree-day terms. We then select the model, from across all of these different specifications, that best fits the data. The two optimal models for each site, for pre- and post-installation periods, are used to calculate NAC. To calculate NAC for the pre- and post-installation periods for each site and timeframe, combine the estimated coefficients μ, $\beta_{H,}$ and β_C with the annual normal-year or typical meteorological year (TMY) degree days H_0 and C0 that has been calculated at the site-specific degree-day base(s), τ_H and τ_H. Thus, for each pre- and post- period at each individual site, use the coefficients for that site and period to calculate NAC.

$$NAC = \mu * 365 \quad + \beta_H H_0 + \beta_C C_0$$

The site-level change in consumption is calculated as

$$\Delta NAC = NAC_{post} - NAC_{pre}$$

Stage Two

The second-stage regression can take many forms. The simplest form simply captures the difference of the site-level ΔNAC between participants and comparison groups in the parameter estimate on a single dummy variable. A more informed model includes information on which sites installed various measure groupings and what the program tracking data recorded as the projected measure grouping savings for that site. This approach is sometimes referred to as a statistically adjusted engineering model (SAE) because it provides adjustments (essentially realization rates) on the engineering savings projections in the tracking data. The SAE model has the following structure:

$$\Delta NAC_j = \Sigma_k \, \gamma_k I_{kj} + \Sigma_k \, \rho_k T_{kj} + \varepsilon_j$$

where,

ΔNAC_j	=	Change in NAC for customer j
I_{kj}	=	0/1 dummy variable, equal to 1 if customer j received measure group k in the current year, 0 if customer j is in the comparison group and/or did not receive

measure group k

T_{kj} = Program gross savings from the tracking data for measure group k for current-year participating customer j, 0 for customer j in the comparison group

γ_k, ρ_k = Coefficients estimated by the regression

ε_j = Regression residual

Two useful results can be derived from the regression estimates: overall program savings given the mix of measures installed, and average measure-specific savings if only that measure had been installed.

Overall program savings is calculated using the following equation:

$$\text{Program savings} = \Sigma_k \, \gamma_k I^*_k + \Sigma_k \, \rho_k T^*_k$$

where,

I^*_k ,T^*_k = The site-level averages of I_{kj} and T_{kj} in the post- period across all participants (including zeros for those who did not install that particular measure group)

Measure-specific savings are calculated using the following equation.

$$\text{Measure savings} = \gamma_k + \rho_k T^{**}_k$$

where,

T^{**}_k = The average measure group program gross savings across those sites with measure group k. The dummy variable, I, is treated similarly so I^{**}_k reduces to one.

Pooled Approach

The evaluation team originally proposed to use a pooled, fixed-effects model for the SBS billing analysis. The primary advantage of this approach over a site-level, PriSM-type billing analysis is the potential avoidance of a control group. The focus of the analysis on installations from the period from October 2010 to June 2011 made it possible to construct a robust comparison group. This allowed the more flexible PriSM approach to become the primary analysis approach that we pursued. In the interest of completeness, we still estimated the pooled specification of the billing analysis.

The following model is similar in structure to the second-stage model estimated above except for a few details. The pooled approach weather-normalizes (goes to an NAC basis) and estimates savings all in one step. This explains why the specification appears to be so much more complicated – weather terms need to be interacted throughout. The added features in the pooled specification are the fixed effects which capture site-specific characteristics across time (μ_i) and month-specific characteristics across sites (ϕ_m). This improves the precision of pooled specification estimates while maintaining some of site-level specificity that is a feature of the PriSM approach. The one drawback of the pooled specification is that all sites are constrained to the same heating and cooling degree-day bases. This removes the site-specific weather flexibility that is a feature of the PriSM approach.

The pooled model equation that is consistent with the second-stage specification above is:

$$E_{jm} = \mu_j + \phi_m + \beta_H H_{jm} + \beta_C C_{jm} + \beta_{Hp} H_{jm} P_{jm} + \beta_{Cp} C_{jm} P_{jm} +$$
$$\Sigma_k \gamma_k I_{kjm} + \Sigma_k \rho_k T_{kjm} +$$
$$\Sigma_k \gamma_{kH} I_{kjm} H_{jm} + \Sigma_k \rho_{kH} T_{kjm} H_{jm} +$$
$$\Sigma_k \gamma_{kC} I_{kjm} C_{jm} + \Sigma_k \rho_{kC} T_{kjm} C_{jm} + \varepsilon_{jm}$$

where,

E_{jm} = Average consumption per day for customer j during interval m

H_{jm} = Specifically, $Hm(\tau_H)$, average daily heating degree days at the base temperature (τ_H) during meter read interval m, based on daily average temperatures over those dates

C_{jm} = Specifically, $Cm(\tau_C)$, average daily cooling degree days at the base temperature (τ_C) during meter read interval m, based on daily average temperatures over those dates

P_{jm} = 0/1 dummy variable, equal to 1 in the post-installation period for all customers j; 0 otherwise

I_{kjm} = 0/1 dummy variable, equal to 1 in the post-installation period if customer j received measure group k; 0 otherwise

T_{kjm} = Equal to program gross savings per day in the post-installation period if customer j received measure group k; 0 otherwise

μ_i = Site-specific fixed effect

ϕ_m = 0/1 indicator for each time interval m, time series component that tracks systematic change over time

β_H, β_C = Pre-installation period heating and cooling coefficients estimated by the regression

$\beta_{H\pi,}\beta_{Cp}$ = Post-installation change heating and cooling coefficients estimated by the regression

$\gamma_{k,}\ \gamma_{kH,}\ \gamma_{kC}$ = Coefficients capturing baseload, heating, and cooling effects correlated with the installation of measure group k

$\rho_{k,}\ \rho_{kH,}\ \rho_{kC}$ = Coefficients capturing baseload, heating, and cooling effects correlated with the program gross savings of measure group k

ε_{im} = Regression residual

As with the second-stage model for the site-level approach both overall and measure-specific results can be derived from the estimated parameters. The process is the same for the pooled specification. For example, overall program savings are calculated with the following equation.

$$\text{Program savings} = 365*(\Sigma_k \gamma_k I^*_{kjm} + \Sigma_k \rho_k T^*_{kjm} + \Sigma_k \gamma_{kH} I^*_{kjm} H^*_{jm} + \Sigma_k \rho_{kH} T^*_{kjm} H^*_{jm} + \Sigma_k \gamma_{kC} I^*_{kjm} C^*_{jm} + \Sigma_k \rho_{kC} T^*_{kjm} C^*_{jm})$$

Where I* and T* are means of site-level post-installation period variables and H* and C* are typical year daily average degree days.

Model Diagnostics

As part of our standard diagnostics, after estimating the models, we look at plots of residuals and check for influential outliers that may be driving the results disproportionally. We employed two main strategies. The first was to remove the top 1% extreme values and rerun the models. If the differences between the results are statistically significant, we investigate further.

The second strategy is to use DFFITS. DFFITS is a common diagnostics statistics for assessing the influence of individual observations to the proposed model. It works by calculating the studentized difference in the predicted value when removing each observation in the dataset one at a time. Following common practice, we flag cases in which DFFITS is greater than 2.

The results from running the models with and without the identified outliers indicated that there was not a significant difference in the ensuing results.