<u>Summary Report</u> –
A Sampling Methodology for Custom C&I Programs

Prepared for:
Sub-Committee of the

Technical Evaluation Committee

September 21, 2012

Prepared by:
Dan Violette, Ph.D. & Brad Rogers, M.S., MBA

Navigant Consulting, Inc.
1375 Walnut Street, Suite 200
Boulder, CO 80302
303.728.2500
www.navigant.com

NAVIGANT CONSULTING

Table of Contents

1. Introduction

This Summary report presents the basic sample design recommendations and approaches. The Main report discusses some of the more technical and nuanced issues that have come up in recent audits of the savings verification studies. Past savings verification studies of Union Gas Limited (Union) and Enbridge Gas Distribution (Enbridge) custom programs have undergone 3rd party audits where the sample design and realization rate calculations are examined in detail. The processes used have been reviewed in-depth, often delving into the judgments made to ensure that the analyses are transparent and accurate. It is hoped that the greater detail in the Main report will streamline these auditor discussions in future reviews. Also, the Main report also contains three technical appendices discussing key issues and presenting the calculations with equations for developing the program realization rate with the achieved confidence and precision of the program savings estimates.

1.1 Project Background

Union Gas Limited (Union) and Enbridge Gas Distribution Incorporated (Enbridge) have delivered Demand Side Management (DSM) initiatives since 1997 and 1995 respectively. Union and Enbridge operate DSM programs including programs that involve custom projects in the industrial, commercial, multi-residential, and new construction sectors. The DSM portfolio for both utilities includes several hundred custom projects annually.

Union and Enbridge DSM activities are regulated by the Ontario Energy Board (OEB) and adhere to the requirements as laid out in DSM Guidelines for Natural Gas Utilities.[1] For custom projects, the resource savings are determined through engineering calculations that are determined at the design stage of each project. There is a need to verify the resource savings through a third-party commercial and industrial (C&I) engineering review.

In 2012, both utilities entered into a new regulatory framework in Ontario that established a new intervener process with the creation of a common Technical Evaluation Committee (TEC) for both utilities. The goal of the TEC is to establish DSM technical and evaluation standards for natural gas utilities in Ontario. The TEC will make recommendations to the OEB on annual Technical Reference Manual (TRM) updates, establish evaluation priorities, and reach consensus on the design and implementation of evaluation studies.

1.2 OEB Requirements for Evaluating Custom Projects

The Ontario Energy Board developed s DSM Guidelines for Natural Gas Utilities prescribes an evaluation approach for custom projects which includes:

[1] "Demand Side Management Guidelines for Natural Gas Utilities." EB-2008-0346. Ontario Energy Board. June 30, 2011.

A special assessment program should be implemented for custom projects. The assessment should be conducted on a random sample consisting of 10% of the large custom projects; and the projects should represent at least 10% of the total volume savings of all custom projects.

The recommended sample methodology conforms to the Guidelines for custom projects.

1.3 Assignment Objective

The objective of this assignment is to develop a methodology for designing a sample and for calculating achieved realization rates and sample confidence and precision. The recommended methodology must meet OEB requirements as well as address the technical and programmatic needs of Union and Enbridge custom programs. Actions taken to achieve this objective are:

- Understand the composition of Union and Enbridge custom programs;
- Review and analyze sample methodologies in selected jurisdictions;
- Recommend a methodology for designing and selecting samples;
- Recommend a methodology for calculating the achieved program realization rates and sample confidence and precision.

The recommended sample methodology is intended to provide sufficient flexibility to allow Union and Enbridge to efficiently meet sample precision needs while the composition, participation, and impacts of their custom programs resemble the current 2011/2012 programs. If the nature of the custom programs changes, adjustments to the recommended methodology may be warranted.

2. Overview of Union Gas and Enbridge Custom Programs

Union's T1/R100 and Commercial/Industrial (C/I) custom programs are aligned under one brand platform, the *EnerSmart* program. Table 1 outlines the rate class divisions of Union's Custom projects. Table 1 shows that the T1/R100 rate class projects represent approximately 30 percent of all projects, but these projects accounted for nearly 70 percent of overall custom program savings. The relative size of projects in different sectors is an important factor in developing sampling plans that are efficient, i.e., produces lower required sample sizes to achieve target levels of confidence and precision. This is important due to the high cost associated with visiting each sample site and conducting a verification study.

Table 1. Union 2011 Custom Projects Overview

Union Custom Sector	# of Custom Projects	Gas Savings	% of Custom Portfolio
T1/R100	200	98,702,955	68.3%
Commercial/Industrial	459	45,472,108	31.5%
Low Income*	13	348,525	0.2%
Total	**672**	**144,523,588**	**100%**

**Low Income numbers are forecast 2012 numbers as this is a new segment for Union in 2012.*

Custom projects are typically tied directly to unique processes or technology requirements resulting in a varied mix of projects. The EnerSmart program was designed to achieve savings in process-specific energy applications, as well as space heating, water heating, and the building envelope. As part of program implementation, each project is validated on a stand-alone basis by a comprehensive professional engineering review and the overall programs are required to pass a TRC screening process. Given the customized nature by which tracking database savings estimates are generated, Union conducts a third party on-site engineering study to verify the results of a representative project sample.

Enbridge offers custom programs for the Commercial and Industrial (C&I) sectors. A variety of incentive based initiatives are offered to C&I sector customers. These initiatives include custom project incentives and a suite of prescriptive offerings aimed at promoting specific measures. Given the myriad of building types and end uses, ownership structures and leasing arrangements, the C&I sector is a complex marketing which to deliver energy efficiency.

Table 2 below outlines the Commercial and Industrial sector divisions of Enbridge custom projects in 2011. For Enbridge, the industrial customer segment only accounts for 14 percent of the overall projects, but these projects account for about 30 percent of the overall savings.

Table 2. Enbridge 2011 Custom Projects Overview

Enbridge Custom Sector	# of Custom Projects	Gas Savings	% of Custom Portfolio
Commercial	780	37,470,116	68.2%
Industrial	127	17,482,847	31.8%
Total	**907**	**54,952,963**	**100%**

There are important differences in the Union and Enbridge custom programs. One difference is the average size of project. The average Enbridge commercial project is about 48K therms compared to about 99K therms for the Union C/I market projects. The average Enbridge industrial project is bout 138K therms compared to the Union T1/R100 industrial projects which average about 493K therms. In general terms, Enbridge's programs serve a market more dominated by commercial customers with smaller average project sizes while Union's programs generally serve a market with more industrial customers which results in larger projects in terms of savings. These factors need to be taken into account in an efficient sample design.

3. Review of Sampling Methodologies in Selected Jurisdictions

This section presents the findings from a review of sampling methodologies used in the evaluation of custom project programs in North America. The reviewed methodologies are all contained within publically available documents.

3.1 *Methodology Review Findings*

C&I programs across North America range in type and size, and they frequently use inconsistent nomenclature. It is common to see custom C&I programs separated from prescriptive programs, but some utilities do combine custom and prescriptive measures into a single program. Stratification approaches and confidence and precision targets are determined differently depending on each utility's regulatory requirements and program organization.

Figure 1 presents an overview of the review. More detail can be found in the Main report and in Appendix C to the Main report which addresses specific studies that were reviewed. A number of specific topics that were addressed in the review are summarized below:

Precision Targets -- Confidence and precision requirements vary widely across the reviewed methodologies. Both one-sided and two-sided confidence intervals are common. Confidence level requirements range from 80% to 90%, and precision requirements ranged from 8% to 20%. These confidence and precision requirements frequently differ in the level at which they are applied. Targets could be set for the overall program, a portfolio of programs, certain customer segments, or for a geographic region (i.e., a transmission zone).

Verification Methods -- On-site verification and evaluation is common industry practice for evaluating savings from custom program targeted at larger C/I customers. There are cases where phone and engineering algorithm verifications have been used for custom programs in some years with more in-depth evaluation work performed in other years, e.g., detailed in-field verifications performed every two years with engineering reviews conducted in the intervening years.

Effect Estimation -- Ratio estimation is used in nearly all of the reviewed studies and has now become a standard practice in the industry. Ratio estimation is a statistical technique whereby prior information from a tracking database—"reported savings"—is verified by independent evaluators. A realization rate is estimated that is the ratio of the verified savings to the reported savings in the tracking system. This ratio is then used to estimate verified program savings. This realization rate approach allows for information on the population of projects to be leveraged with the information gained from the selected sample of participants to increase overall precision of the estimated program savings. This leverage reduces the overall sample size requirements. If stratification is used, realization rate measured for each stratum is used to estimate overall program savings.

NAVIGANT CONSULTING

Figure 1. Summary Comparison of Sample Methodologies in Selected Jurisdictions

N°	Service Territory or Jurisdiction	Organizations Reviewed	Year	Service Type	Timing	Precison Target	Stratify by Size	Stratify by Segment	Ratio Estimation
1	Illinois (Chicago)	Commonwealth Edison Company	2011	Electric	2-stage	90/08 (3yr utility program)	✓		✓
2	Michigan (Detroit)	DTE Energy	2010	Gas & Electric	1-stage	90/10 (utility program)		✓	✓
3	Massachusetts	Massachusetts Energy Efficiency Advisory Council (NSTAR, National Grid, Western Massachusetts Electric Company)	2009	Gas & Electric	1-stage	90/10 (statewide custom C&I)			✓
4	New Mexico	El Paso Electric Company, New Mexico Gas Company, Public Service Company of New Mexico	2011	Gas & Electric	1-stage	90/10 (utility total portfolio)	✓		✓
5	Pennsylvania (Philadelphia)	PECO Energy Company	2011	Gas & Electric	3-stage	85/15 (utility C&I total)	✓	✓	✓
6	Ohio	AEP Ohio	2011	Electric	2-stage	90/10 (utility program, RTO zone)	✓	✓	✓
7	Maryland	EmPower Maryland (Baltimore Gas & Electric, Potomac Electric Power Company, Delmarva Power, Southern Maryland Electric Cooperative, and Potomac Edison)	2011	Gas & Electric	1-stage	80/20 one-sided (utility program)	✓		✓
8	California	California Public Utilities Commission (Pacific Gas & Electric Company, San Diego Gas & Electric, Southern California Edison, Southern California Gas Company)	2009	Gas & Electric	flexible	90/10 (utility program)	✓	✓	✓
9	Vermont	Vermont Department of Public Service (Efficiency Vermont and Burlington Electric Department)	2010	Electric	2-stage	80/10 (utility portfolio)	✓	✓	✓
10	PJM Interconnection (Midwest & Eastern US)	PJM Interconnection	2010	Electric	flexible	90/10 one-sided (utility program, RTO zone)	✓	✓	✓
11	US Federal Facilities	US Department of Energy	2008	not applicable	flexible	not applicable		✓	
12	General International	Efficiency Valuation Organization (IPMVP)	2012	not applicable	flexible	not applicable		✓	

Uses of Stratification -- Stratification approaches varied across the reviewed studies and appear to be customized to fit each utility's program structure, number of projects, sizes of projects, regulatory requirements and stakeholder concerns. The review yielded two common approaches for stratifying each based on size of project. The first approach defines the large strata based on very large projects in the population. Sometimes a census is sought when the very large strata contains only a few projects. The second approach divides the population into strata of roughly equal contribution to total savings. In addition to stratifying by magnitude of project savings, some studies stratified by segment. Segments used for stratification included market sector (e.g., education, multi-family, manufacturing and other customer-type based segments), geography and project-types (space heating, water heating, or industrial process).

Sample Staging -- Schedule requirements for reporting often necessitated a staged approach to sampling in order to begin the evaluation efforts early enough to meet timely reporting requirements in that jurisdiction. About half of the reviewed studies implemented staged sampling. Typically, the two-stage approach is comprised of a stage one sample that is drawn based on either the first two or first three quarters of the year. A second stage draws the remainder of the sample from those projects completed in the fourth quarter. This allows for the evaluation effort to begin on the stage one sample earlier in the year. Verification of C/I custom projects requires time to set up and complete. If the verification assessment was started only after information were available on all the projects completed in a year, the results may not be available to meet program reporting requirements. This staged or rolling sample is becoming quite common across the C/I custom EE programs with some jurisdictions using a rolling sample comprised of three-stages.

Differences between Gas and Electric Programs – Major differences in methods used in evaluating savings between electric and gas utilities were not found. Differences are more likely to be based on program size and number of years evaluating and reporting program savings. Most jurisdictions count both electric and gas savings for custom C&I measures regardless of whether the administrating utility supplies both fuel types.

Bias in Results-- Industry best practices prescribe a demonstration of effort to control for common sources of bias. The principle bias of concern is the over or under representation of certain elements of the population in the sample. Stratification is useful in reducing this potential bias as it tends to make the sample more representative of the population by requiring, at a minimum, that the sample be spread across the different strata. Bias also can be introduced into the analysis by anomalous observations in the sample that for some reason are unique and not representative of other members of the population. In general, little discussion was found addressing the composition of the sample, treatment of outliers, sample replacements, missing data points, or other sample adjustments. This discussion could be addressed in project memos rather than expanding what is often a lengthy final evaluation report. However, this is an area where standard industry practice may not be on par with evaluation practices in other fields. It is not clear whether this deficiency is related only to reporting or if it reflects limitations on current evaluation practice.

4. Recommended Sample Design Methodology

Ratio estimation has become standard practice for the evaluation of large C&I programs as it leverages information available on the population of projects with the sample. The sample design approaches discussed in this section are constructed to make full use of the ability to leverage sample data in combination with information on the population available from the project tracking database. This is important given the relatively high cost of rigorously evaluating custom C&I projects.

The level of specification for sampling protocols observed in jurisdictions across North America varies widely. An overly specified methodology may lead to incompatibilities in future evaluation efforts as the composition, participation and distribution of impacts can change from year to year. However, an overly general methodology may lead to sample designs that are not efficient in meeting the confidence and precision requirements. The recommended sample design methodology is intended to strike a balance between flexibility and specificity to allow Union and Enbridge to best meet their evaluation needs now and in future program years.

4.1 *Stratification*

Stratification in designing samples for evaluating custom C&I programs is recommended both for its benefits in enhancing confidence and precision, but also for its ability to increase the representativeness of the sample. Stratification is the practice of disaggregating the population into sub-groups based on selected criteria.

The specific stratification approach will depend on an evaluation of the population data. If the distribution of project savings for a program is relatively tight[2] and there is not an easily delineated group of large projects. In this case, then stratification by project size alone may not produce substantive sampling efficiencies. However, if the distribution of project savings is wide or there is clear group of large projects, then stratifying by project size will likely produce sampling efficiencies. Stratification by project type (e.g., space heating or industrial process applications) or market segment (manufacturing, education, or multi-family) may also be appropriate.

4.2 *Ratio Estimation*

A ratio estimation approach is recommended. Ratio estimation is often used to increase the precision of estimated means and totals. It is motivated by the desire to use information about a known auxiliary quantity (i.e., reported savings) to obtain a more accurate estimator of the population total or mean (i.e., verified savings). When applying ratio estimation within a stratified population, the separate ratio estimator approach should be used where strata are

---

[2] A "tight" project savings distribution is generally considered to be less than a single order of magnitude. Size based stratification should be considered when the distribution of savings spans multiple orders of magnitude.

defined and realization rates calculated for each stratum and sample weights are used to estimate overall program savings.

4.3 Sample Staging

A rolling sampling approach using a two-stage sample approach is recommended to ensure that reporting requirements can be met. Reporting schedules often do not provide sufficient time to design and evaluate a sample after all the projects in a specific year are finalized. This type of schedule constraint frequently occurred in the jurisdictions reviewed in Section 4. A rolling sample implemented through sample staging as described below can allow evaluation efforts to begin earlier using sampled projects completed early in the program year. By starting the verification effort earlier, staging can help meet reporting requirements by reducing the evaluation workload required between the end of the program year and the targeted reporting deadlines.

The first sampling stage evaluates a sample based on the first three quarters of the program year, and the second stage accounts for the full population at the end of the year. This approach is not intended to create separate strata or separate estimates for each stage. Rather, the first stage is simply part of the population of projects being from which sample projects are selected for verification.

Union's and Enbridge's projects tend to come online more heavily in the fourth quarter with roughly half to three-quarters (depending on which program) of projects completing in the last quarter. This would imply that a 50-50 split between sample stages would be reasonable, given constraints related to the calendar time needed to set up and conduct the verification studies and the reporting requirements. However, if the timing allows, Union and Enbridge might consider placing more of the sample into the fourth quarter when savings from projects completed in the fourth quarter are expected to contribute more than half of program savings. This recommendation is a compromise between the time and resources needed to perform the number of required site verifications, and the need to meet program reporting deadlines.

4.4 Recommended Sample Design Process – Seven Steps

The recommended sample design methodology contains the following seven steps:

Step 1: Review project tracking database for accuracy and quality.
Prior to any stratification or sampling, gains in the precision of the final program estimates can be made in the resulting analysis and precision by reviewing the estimates in the tracking database and making sure that the best possible initial project-based engineering estimates are contained in the tracking database. Many utilities do a second check of the tracking database prior to the sample design and sample selection. Identifying unique projects in the tracking database can help avoid outlier problems later in the analysis.

Step 2: Evaluate the population and define strata.
Examine the population for ways to leverage the sample design to improve efficiencies in meeting target confidence and precision levels. This includes three activities:

1. *Assess the exclusion of extremely small projects* – Ratio estimation weights project realization rates according to project savings. Very small projects typically exert only negligible influence on estimates of the total realization rate, the total savings, and the total achieved precision. For many very small projects, a 100% difference in realized savings versus reported savings would produce a negligible impact on the total estimates. The cost of evaluating the impacts of these small projects exceeds the value of the information obtained from them. It is therefore considered reasonable to exclude the very small projects. Historically, a large number of these small projects typically cumulatively only represent 5% or less of the total program savings based. As a result, excluding small projects that cumulatively account for up to 5% of program savings is recommended. Some judgment is needed to see if there are natural delineations in the small project category. In some years, projects that overall have contributed 3% to program savings have been dropped from the sampling frame. In recent years, this number has been 5%.
2. *Identification of project size strata bounds* – Efficiencies can be gained by stratifying by project size when the distribution of project savings is wide or there is a clear group of large projects. Sorting the projects by savings size can allow for identification of discontinuities in the project size distribution. Set strata bounds first based on natural breaks in the distribution that result in easily delineated groupings. If natural groupings do not exist, other approaches may be used such as stratifying into strata of roughly equal total savings. The number of size-based strata typically ranges from two to four, with three most commonly applied for C&I program evaluations.
3. *Identification of categorical characteristic strata bounds* – Efficiencies can be gained by defining strata along market sector or project-type categories when these categories contain projects that are similar within that segment. For commercial projects, strata could be defined by building type (e.g., schools, office building, multi-family, etc.). Similar buildings could be expected to have a lower variance in the estimated realization rate across sites (i.e., within the strata) than when combined with other building types.

There are tradeoffs that need to be assessed in deciding the strata to be used. This assessment requires an analysis of the data, knowledge of project types, and knowledge about program administration and participation. It should also balance concerns about the efficiency of the sample with helping ensure that the sample is representative of the population to which the results are being extrapolated.

Step 3: Estimate an appropriate variance for each stratum.
In ratio estimation, the variance considered is that of the residuals on the stratum average realization rate rather than the variance of the verified savings. Accordingly, the development

of a sample plan requires that a coefficient of variation (CV) or error ratio should be based on the assumed distribution of individual realization rates for the population of projects in each stratum.

The CVs should be based on the un-weighted realization rates for historic sample data from prior program years, when such data are available. Any changes in program composition, administration, or participation from the previous year will decrease the validity of applying prior year CVs, and the assumed CVs should be adjusted upward by 0.1-0.2 to prevent under-sampling. Navigant does not recommend using a coefficient of variation less than 0.30 so to ensure sample sizes sufficient for robust results and to allow for increasing variances that may result from evolving measurement approaches and program participation. A CV of 0.5 may be assumed when historic data are not available. This is a standard industry assumption and is generally conservative in ratio estimation if the population reported savings in the tracking database are reasonably accurate.

Step 4: Allocate observations to each stratum.

The overall sample should be designed to achieve a targeted 10% precision (two-tailed) at a 90% confidence level (i.e., 90/10).[3] This confidence and precision target is meant to be achieved for each custom program in each program year. This target is at the high end of accepted industry practice. If changes are made to this target, these changes can be addressed in the sample size calculations and do not necessarily warrant changes in the recommended methodology.

Allocating the sample across strata to achieve target confidence and precision often requires an iterative approach where different sample plans are tested. Proportional sampling was found to be widely used in the review of methods, but this should not be the default approach. Proportional sample allocation allocates sample to strata in proportion to some characteristic such as savings, but this approach may not result in the most efficient sample plan. For example, variances can differ across samples with, for example, larger variances expected for different sizes and types of projects. As a result, alternative allocations of sample across strata should be tested using sensitivity analysis around the strata CVs to produce a final sample design. Given the judgment needed to develop a sample design, it is important to test the robustness of the design though simulating different scenarios. This is illustrated in the examples contained in the Main report.

Step 5: Determine criteria for assessing sample representativeness. (optional)

There are often categorical characteristics of the population that are not used in defining strata but may help ensure that a selected sample is a reasonably representative of the population. For example, a market segment may not have been used in defining strata, but a random sample that fails to include at least some observation from certain major market segments might not be viewed as a representative sample by reviewers of the verification assessment.

[3] "Terms of Reference: Sampling Methodology of Custom Projects." Union Gas Ltd. & Enbridge Gas Distribution Inc., June 27, 2012. (see page 4)

You could set up new strata for these factors, but it is expected that a random draw will be representative across these factors and there is a benefit for a simple stratification design.

To address this, some criteria can be defined prior to randomly selecting a sample that can be used to assess the representativeness of the sample. Criteria should be established only for the most important characteristics, and they should only be set at general levels and levels that, if not met, would represent an extreme sample in terms of the sample representing the population (See example is Section 5.5 of the main report).

Step 6: Select a random sample.
The sample for each stratum should be selected at random. This provides an equal opportunity for each project within a stratum to be selected.[4] This can be accomplished in Microsoft Excel using the random number function to assign a random number between 0 and 1 to each project in a stratum.

Recruitment of the full selected sample is often not achievable since some program participants may not respond or refuse to participate in the site verification study. Even when agreement to participate in evaluation activities is required to participate in the program, full recruitment of the selected sample can often not be achieved. Therefore, a set of potential replacement projects may be provided to recruiters to fill in for non-recruited participants.

Potential replacements should be selected from the same random number list of the population from which the original sample was selected. Replacements should be selected in priority of assigned random number until full recruitment is achieved. The full population of a stratum should not be provided to recruiters, whose incentives are not usually aligned to follow the random prioritization of the sample, unless the full sample size is not expected to be achieved.

Step 7: Recruit the sample.
Recruitment of each stratum sample can begin once the sample has been selected and assessed. Recruitment typically occurs over the phone, and may or may not involve scheduling of the on-site evaluation visit. Ensuring the accuracy and completeness of contact information in the tracking database can streamline the recruitment task.

The list of potential replacements may be initially withheld from recruiters to ensure that the originally selected sample projects are pursued fully before being replaced by alternate projects. This can help reduce the possibility for non-response bias in the sample.

4.5 Summary of Sample Design Methodology

The sample design methodology described in this section is meant to apply industry practices to create a cost-efficient sample by leveraging pre-existing project and program information to

[4] Sampling from a savings weighted distribution can also be valid, but it is not recommended here since size based strata are already employed.

the greatest extent possible. The methodology can be described as a "stratified ratio-estimation" approach. The sample is administered in two-stages to make the best use of early observations that can be collected prior to completion of the program year. The methodology provides a step-by-step description of sample design tasks, but leaves flexibility to accommodate program changes in future years and cycles. The main report presents illustrative examples of this sample design process, with samples drawn for each utility for each C&I custom program.

5. Recommended Realization Rate Methodology

This section describes the recommended methodology for determining realization rates and achieved confidence and precision based on data from the sample selected. This section also discusses several potential adjustments that may be needed to ensure that the results appropriately characterize the population and provide the information needed by the utilities and stakeholders.

5.1 Determining Verified Realization Rates

Realization rates should be calculated for each stratum sample and applied to each respective stratum population when estimating total savings. Applying realization rates to population strata is more complicated than assessing the results in a simple random sample without strata, but it is necessary when efficiencies are sought through stratification. Again, sample size and design efficiencies are important due to the high cost of gathering the verification data at each sample site. The calculations needed to develop a verified realization rate from stratified sample data are shown in Appendix B to the Main report.

5.2 Determining Achieved Confidence & Precision

It is important to recognize that a precision levels and confidence levels are inter-related. The calculation for both confidence and precision come from the same basic equation. A given level of confidence or precision is first established; then, the other is solved for. For example, a precision of +/- 10% implies that the stated confidence level should be that level of confidence that the data show results in +/- 10% precision from the mean estimate. This confidence level may turn out to be 90%, 82% or another value. Typically, the confidence level is established first and the precision is solved for. For example, the level of precision achieved at a 90% level of confidence can be calculated and may turn out to be 10%, 12%, 15% or some other number. Regardless, the calculation of confidence and precision are part of the same calculation and one cannot be estimated one without establishing the other. Additional discussion on confidence and precision can be found in Appendix A to the Main report. These confidence and precision calculations take into account the fact that a stratified random sample has been used and use the strata sample weights. The equations for calculating confidence and precision from a stratified sample design are shown in Appendix B to the Main report.

5.3 Sample Adjustments and Related Issues

This section discusses several adjustments that may be needed to provide the information needed by the utilities and other stakeholder on program savings estimates, and the confidence with which these estimates are calculated. Three types of adjustments are discussed in the Main report:

1. Treatment of outliers and influential observations,
2. Replacing sample projects,
3. Post stratification.

Each is discussed below.

Treatment of Outliers and Influential Observations
It is recommended that all verification assessments and estimates of program savings examine the sample data for outliers and influential observations.

> *The rationale for looking for outliers is that they may have a strong influence on the estimates…an influence that may not be desired.* [5]

The reason for looking for evaluating outliers is that there may be a sample case drawn that is well outside the expected bounds of the distribution and that this observation may exert undue influence on the estimates of the analysis (i.e., be an influential observation).

The issue is whether it is appropriate for a single observation to swing the overall results in a substantial manner.[6] If such an observation is found, then further study is needed to determine the most appropriate course of action. In general, a sample of ten from a population of 100 projects implies that each sample point represents ten projects. However, if a selected sample point is truly a unique case and does not represent other projects in the population, then an adjustment may be warranted.[7]

The sample analysis should seek to determine whether or not outliers and influential observations can be viewed as representative members of the main population upon which population estimates may be inferred.

Projects that implement new technologies—whose savings estimates have had less validation—or certain technology classes that are complex and difficult to estimate for the tracking database may be at an increased likelihood to result in outlier realization rates. Identifying such projects in the program tracking database could help isolate them and reduce their chance of skewing program estimates. These projects could be placed into a separate category with different confidence and precision targets for new technologies. Any projects that are truly unique should be identified and addressed during sample design. These steps would not eliminate these projects in terms of their contribution to overall program savings, but would allow for appropriate methods to more accurately estimate program savings. If

---

[5] Kennedy, P. "A Guide to Econometrics." Third Edition. MIT Press 1992, *p.279.*

[6] A simple intuitive example of the impacts an outlier can have on a statistical analysis can be found in a Wikipedia contribution (8/20/2012): *Naive interpretation of statistics derived from data sets that include outliers may be misleading. For example, if one is calculating the average temperature of 10 objects in a room, and nine of them are between 20 and 25 degrees Celsius, but an oven is at 175 °C, the median of the data could be between 20 and 25 °C but the mean temperature will be between 35.5 and 40 °C. In this case, the median better reflects the temperature of a randomly sampled object than the mean; however, naively interpreting the mean as "a typical sample", equivalent to the median, is incorrect. As illustrated in this case, outliers may be indicative of data points that belong to a different population than the rest of the sample set.*

[7] Osborne and Overbay state that: "*Where outliers are illegitimately included in the data, it is only common sense that those data points should be removed… Few should disagree with that statement*" From Osborne, J and, A. Overbay, ***"The Power of Outliers and Why Researchers Should Always Check for Them."***2004 Practical Assessment, Research & Evaluation, volume 9, section 6. Link: http://pareonline.net/getvn.asp?v=9&n=6

sampled, these unique projects should not be considered representative of other projects in the custom program. As a result, addressing this issue in advance can improve the sample design and help avoid resulting program estimates.

The Main report presents ways to identify potential outliers, and assess the influence of those observations. A discussion of how outliers were treated in a recent custom program verification is also presented in the Main report.

Replacing Sample Projects

The final recruited sample should be analyzed and summarized, especially when replacement projects are substituted into the originally selected sample. Recruiters should document the reasons for unsuccessful recruitment of original sample members. Replacement samples should always be selected in priority based on the assigned random number, and full effort should be made to recruit selected replacements before substituting other replacements. If recruitment rates are very poor, this may introduce a significant non-response bias. Low recruitment rates should be investigated and documented, and recommendations may be made to improve recruitment in sub-sequent evaluation years.

Post-Stratification

If a sample did not achieve the desired confidence and precision and the stratification basis is thought to be sub-optimal, post-stratification may be used to retrospectively re-stratify a sample along more appropriate dimensions to demonstrate an improved precision achieved by the sample. Often post-stratification will not improve achieved precision, especially at relatively small sample sizes; but, under certain circumstances this technique may be useful. The Ontario Power Authority notes that:

> *A technique known as post-stratification may be used to develop estimates about sub-populations after the study is complete and can be used if characteristics about the sub-populations are unknown at the time the study in conducted.*
>
> *This advanced technique should be reserved for special situations and utilized only after careful consideration of other options and well documented in the experimental approach of the Draft Evaluation Plan.* [8]

Post-stratification should not be used on a normal basis, and if necessary should inform subsequent program evaluation cycles to improve the sample frame and prevent the need for post-stratification in future years.

5.4 Summary of Realization Rate Methodology

This section presents the method for calculating verified ex-post realization rates as well as for appropriately calculating the confidence and precision levels for the estimated realization rate

[8] "EM&V Protocols and Requirements: 2011-2014." Ontario Power Authority. March 2011, p. 130.

and overall program savings. It also discussed three issues that can lead to adjustments to the sample and recalculation of the realization rate along with confidence and precision levels. There are several important concepts presented in this section:

- The program realization rate is inferred from the sample observations based on the separate realization rates for each stratum.
- The realization rate calculations should apply the strata weights to accurately interpret sample observations. This adds a bit of complexity, but no alternate application of the observed data would be appropriate. This is considered standard practice in the application of a stratification approach in statistics.
- There are some important and legitimate considerations that should be examined when inferring estimates for a population from an observed sample. Three factors considered to be important are discussed in this section:
    1. Outliers and influential observations;
    2. Replacement projects when data cannot be gathered from the originally sampled project;
    3. Post stratification to provide higher precision and greater confidence in the results.

The equations needed to calculate the realization rates and achieved confidence and precision from the sample data are contained in Appendix B to the Main report.