FINAL REPORT

Custom Savings Verification Evaluation for Ontario's Natural Gas Demand-Side Management Program

Ontario Energy Board

Date: November 25, 2024

Table of contents

List of figures

List of tables

EXECUTIVE SUMMARY

To verify the impacts of the Enbridge Gas Distribution, Inc. (Enbridge) demand side management (DSM) programs, the Ontario Energy Board (OEB) undertakes various annual evaluation studies. The Gross Savings Verification Evaluation of the 2023 Natural Gas Demand Side Management Custom Programs is summarized in this document.

In 2023, Enbridge delivered ratepayer-funded DSM programs to customers, including custom programs delivered to large volume, commercial, and industrial customers that encouraged them to reduce their energy consumption by providing customer-specific energy efficiency and conservation solutions. The custom programs offered provide financial incentives, technical expertise, and guidance with respect to energy-related decision-making and business justification to help customers prioritize energy efficiency projects against their own internal competing factors. Multi-residential buildings – other than low-income buildings, which are dealt with separately – are eligible to participate in Enbridge's custom commercial programs.

The OEB evaluates the custom commercial and industrial program results annually as the programs have significant OEB-approved savings targets. Based on the results of the utilities' programs, the utilities may be eligible for performance incentives. The portion of shareholder incentives that come from the custom commercial and industrial programs is based on the amount of verified net natural gas savings achieved by each utility relative to the OEB-approved targets.

- **Verified savings** are utility draft program savings that are audited and confirmed by an independent third party. The process and results of the verification are described in a separate report. The result of the analysis is a ratio that represents the percentage of utility-draft energy savings that are verified by the auditor.
- **Net savings** are those that are caused, or influenced, by the utility. The process and results of the net savings assessment are described in the 2023 Natural Gas Demand Side Management Free-Ridership Based Attribution Evaluation report. The result of the analysis is a ratio that represents the percentage of verified savings that were caused by the utility.

The two ratios are applied to the utility draft savings to produce final verified net natural gas savings according to the equation in the following figure.

This summary reports the verification ratio, which along with claimed savings and the net savings ratio serves as an input used to calculate verified net savings. The customer program results are combined with the results from other utility programs in a "scorecard." The utilities' scorecard results determine overall performance and if the utility is eligible for a shareholder incentive. The following tables shows the gross savings verification ratio for each program and segment.

Table ES-1. Commercial program

Segment	Gross Realization Rate	+/- at 90% Confidence	Sampled Measures	Population Measures	Population M3 Savings	Percent population M3 Savings
Commercial	91.13%	9%	14	405	17,165,686	57%
Low Income and Multi-Residential Multi-Family	97.61%	8%	18	681	12,961,169	43%
Commercial Program	93.85%	7%	32	1,086	30,126,855	100%

Table ES-2. Industrial program

Segment	Gross Realization Rate	+/- at 90% Confidence	Sampled Measures	Population Measures	Population M3 Savings	Percent population M3 Savings
Agricultural	99.12%	1%	29	190	30,754,095	48%
Industrial	96.52%	5%	18	163	32,772,861	52%
Industrial Program	97.69%	3%	47	353	63,526,956	100%

Table ES-3. Large Volume

Segment	Gross Realization Rate	+/- at 90% Confidence	Sampled Measures	Population Measures	Population M3 Savings	Percent population M3 Savings
Large Volume	98.07%	11%	12	31	80,549,726	100%

Findings and recommendations

#	Energy savings and program performance: Finding	Energy savings and program performance: Recommendation	Applies to: Enbridge	Applies to: Evaluation	Primary beneficial outcome: Reduce costs	Primary beneficial outcome: Increase savings	Primary beneficial outcome: Customer satisfaction	Primary beneficial outcome: Decrease risk
1	Enbridge continues to exhibit a strong commitment to accurate energy savings estimates.	Enbridge should continue its cultural commitment to accuracy.	✓			✓	✓	✓
2	The CPSV effort this year found realization rates between 90% and 100%	Continue performing custom savings verification on a regular basis.		✓				✓
		Consider approaches to sampling that can reduce sample sizes and costs.		✓	✓		✓	

#	Energy savings and program performance		Applies to		Primary beneficial outcome			
	Finding	Recommendation	Enbridge	Evaluation	Reduce costs	Increase savings	Customer satisfaction	Decrease risk
3	Some measures in each utility program are routine maintenance, periodic repairs or like for like replacements that are considered standard care in other jurisdictions.	Establish a clear policy regarding eligibility of maintenance, repair and like for like replacement measures for the programs.	✓		✓			✓
4	The close relationships between Enbridge Energy Savings Consultants (ESCs) and customers provide advantages and challenges for evaluation.	Clarify the role of evaluation engineers, customers, and ESCs in the evaluation. Set and communicate clear expectations for each of the three roles so all parties are aligned.	✓	✓			✓	✓
5	Project documentation continues to improve.	Continue to improve data quality.	✓			✓		✓
6	Some Large Volume measures appear as two separate measure rows in the database due to having two sources of incentive funding.	Add a field to the tracking database to link two rows that are a single measure implementation.	✓		✓			

1 INTRODUCTION

On behalf of the Ontario Energy Board (OEB), DNV carried out the Custom Program Savings Verification (CPSV) of Enbridge Gas Inc.'s (Enbridge) natural gas demand-side management (DSM) programs delivered in 2023. The study produced verified savings ratios and verified gross savings totals for the custom projects in the Enbridge programs examined, shown in Table 1-1.

Table 1-1. CPSV by program

Program	2023 Gross Verification
Large Volume	✓
Commercial*	✓
Industrial	✓

*Custom Market-Rate Multi-Residential (Multifamily) and Low Income Multi-Residential projects are expected to be included as a part of this program.

1.1 Evaluation objectives and approach

The study objectives were to:

- Develop accurate verified gross savings for each of Enbridge's custom commercial, industrial, multi-residential (including low-income), and large volume programs carried out in 2023, with disaggregated rates for each of the major program components within these groupings (for example, differentiated by segment/technology type and determined in consultation with the EC, OEB staff, and EAC at the start of the study).
- Establish and maintain transparency throughout the project.
- Follow industry best practices.
- Achieve 90/10 precision[1] at the requested stratification segment levels.

The methodology selected for the CPSV study consisted of engineer reviews of gross savings. Reviews of complex projects included on-site verification and data collection, while less complex projects were verified with Telephone Supported Engineering Reviews (TSERs).

1.2 Study background

To encourage Enbridge to implement public benefits programs designed to reduce overall energy use, called conservation demand-side management (DSM) programs, the OEB reimburses them for the cost of program implementation and provides an incentive, called the shareholder incentive, that reflects the utilities' performance against pre-determined targets. The OEB also compensates the utilities for the revenue lost as a result of the lower natural gas sales.

In the 2023 calendar year, programs delivered by Enbridge targeted all natural gas ratepayers, including residential, multifamily, low-income, commercial, and industrial customers. This study is part of step 4 of an overall conservation program cycle, as shown in Figure 1-1.

Figure 1-1. Conservation program cycle

[1] 90/10 precision refers to 10% relative precision with 90% confidence.

To verify the impacts of the Enbridge DSM programs, the OEB sponsors studies to verify the energy savings achieved. Specifically, this study verifies the engineering calculations, inputs, and assumptions that produce the utilities' claimed gas savings. The results of this study are combined with the results of the 2023 Net-to-Gross study to produce verified net cumulative gas savings for Custom measures in Enbridge's 2023 Commercial, Industrial and Custom Large Volume programs.

2 COMMERCIAL CUSTOM PROGRAM

Enbridge's custom DSM programs for commercial customers encourage customers to reduce their natural gas consumption by recommending and incentivizing energy saving projects and actions.

These custom programs differ from the prescriptive programs by providing additional technical support for projects. They also provide financial incentives based on overall natural gas savings realized by the customer rather than a per-unit incentive.[2]

2.1 Gross savings realization rate

The gross realization rate (GRR) represents the ratio of the savings verified by the evaluation to the savings claimed (or reported) by the utility, as shown in the following equation. A 90% GRR means the verified gross savings for the project or program were 90% of the claimed savings. Differences between claimed and verified savings for each project can arise for a number of reasons, usually related to differences in forecast assumptions, differences in underlying facts, or differences in calculation approaches or parameters.

$$gross\ realization\ rate = \frac{evaluation\ verified\ savings}{utility\ reported\ savings}$$

Table 2-1 shows the first-year gross savings realization rate by customer segment for the Commercial Custom Program. It shows the gross realization rate, statistical precision at the 90% confidence interval, the program-claimed population cumulative cubic meters of natural gas (CCM) savings, and percent of program savings for each customer segment. The percent of program savings represents the relative contribution that each customer segment makes to the overall result.

Enbridge's Commercial custom program overall achieved a 94% gross realization rate. The customer segment gross realization rates varied from 91% to 98%. The largest segment was Commercial with 57% of the population energy savings. Relative precision for the program overall was 7% at 90% confidence.

Table 2-1. First year gross savings realization rate for the Commercial Custom program

Segment	Gross Realization Rate	+/- at 90% Confidence	Sampled Measures	Population Measures	Population M3 Savings	Percent population M3 Savings
Commercial	91.13%	9%	14	405	17,165,686	57%
Low Income and Multi-Residential Multi-Family	97.61%	8%	18	681	12,961,169	43%
Commercial Program	93.85%	7%	32	1,086	30,126,855	100%

2.2 Discrepancy summary

This section presents detailed results of the various project-level discrepancies between program claimed and evaluation verified savings. The verification found discrepancies in 53% of the projects reviewed.

Figure 2-1 shows that 16 of the 32 measures had no adjustment from program claimed to evaluation verified savings, while 16 measures were adjusted based on verification findings. For custom savings verification, we consider verified savings that differ more than 20% from utility tracking savings to be a "large" discrepancy. Moderate adjustments within 20% of utility tracking savings are expected given the level of uncertainty in forecasting energy savings. Nine of the 16 adjusted measures

[2] A more detailed description of the program can be found in Enbridge's 2023 Demand Side Management Annual Report

had verified savings within 20% of utility tracked savings. Of the 7 measures with adjustments greater than 20%, one (1) had an adjustment increasing savings (measure level realization rate greater than 120%) and six (4) had adjustments decreasing savings (adjusted measure level realization rate less than 80%).

Figure 2-1. Adjusted realization rate (ARR) summary – Commercial Custom program

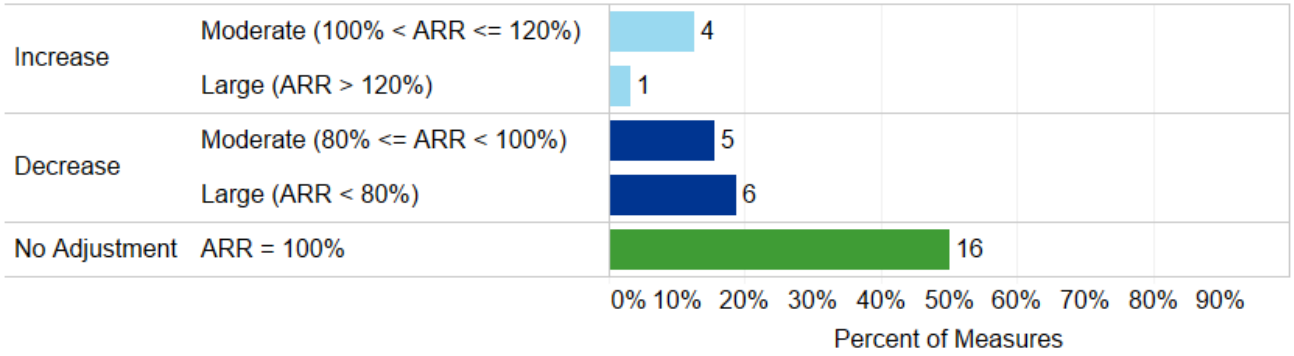


Figure 2-2 plots the claimed first year savings and the realization rate for each measure in the sample. The plot is sorted with the smallest measure on the bottom and largest on the top. The left plot shows the relative size of each measure. The right plot shows the gross realization rate for each measure. In both plots, measures with light blue bars have a realization rate greater than 100% (verified savings greater than utility claimed savings). Measures with dark blue bars represent a gross realization rate less than 100% (verified savings lower than utility claimed savings). Measures with green bars represent a gross realization rate of 100%.

Figure 2-2. Sample measure realization rates sorted by size – Commercial Custom program

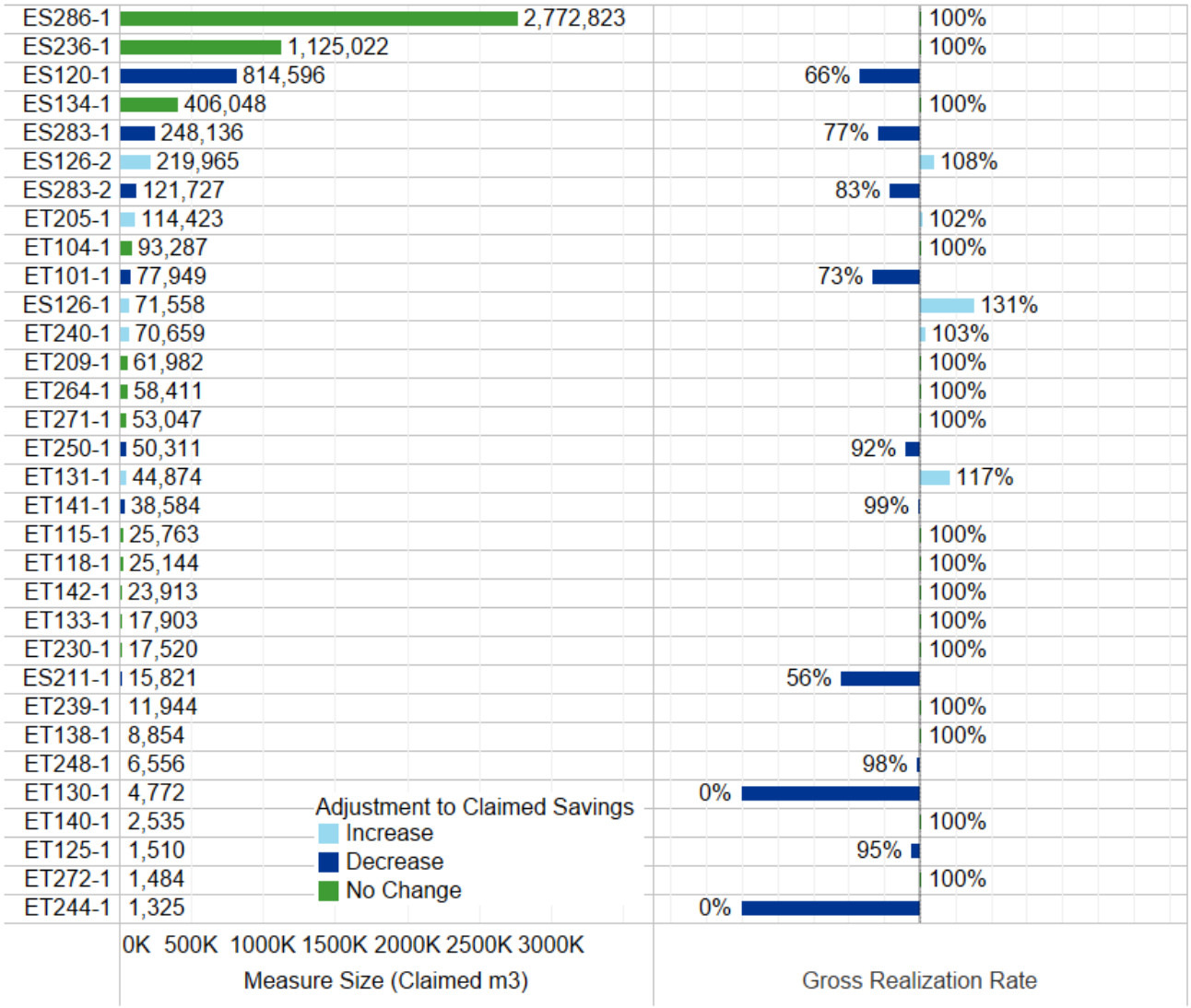
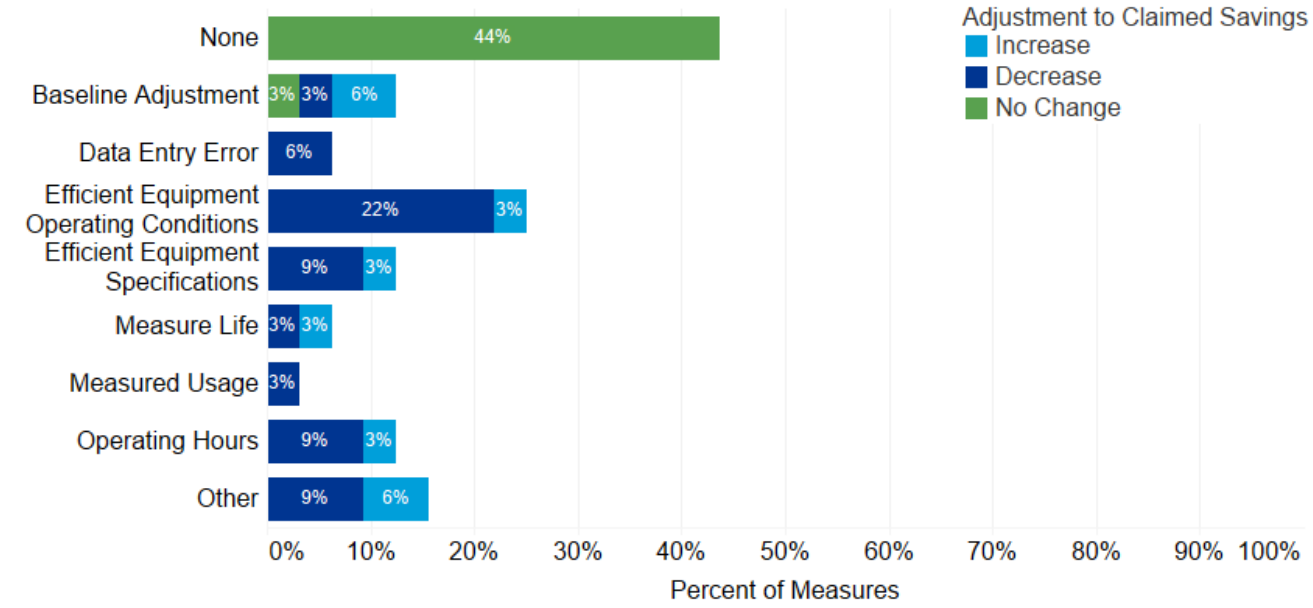


Figure 2-3 shows the types of discrepancies found by the verification. Each measure may have more than one discrepancy. The verification found no discrepancies for 44% of sampled measures. Operating conditions were the only type of discrepancy found for more than 20% of measures. The utility can reduce this type of discrepancy by documenting projects more thoroughly with sources for the assumptions used and more complete documentation of conditions found at the time of installation (see recommendations in section 5); however, this type of discrepancy is partially outside of utility control. One measure had a baseline adjustment that resulted in no change to savings.

Figure 2-3. Savings discrepancies – Commercial Custom program



3 INDUSTRIAL CUSTOM PROGRAM

Enbridge's custom DSM programs for industrial customers encourage customers to reduce their natural gas consumption by recommending and incentivizing energy saving projects and actions.

These custom programs differ from the prescriptive programs by providing additional technical support for projects. They also provide financial incentives based on overall natural gas savings realized by the customer rather than a per-unit incentive.[3]

3.1 Gross savings realization rate

The gross realization rate (GRR) represents the ratio of the savings verified by the evaluation to the savings claimed (or reported) by the utility, as shown in the following equation. A 90% GRR means the verified gross savings for the project or program were 90% of the claimed savings. Differences between claimed and verified savings for each project can arise for a number of reasons, usually related to differences in forecast assumptions, differences in underlying facts, or differences in calculation approaches or parameters.

$$gross\ realization\ rate = \frac{evaluation\ verified\ savings}{utility\ reported\ savings}$$

Table 3-1 shows the first-year gross savings realization rate by customer segment for the Industrial Custom Program. The table shows the gross realization rate, statistical precision at the 90% confidence interval, the program-claimed population cumulative cubic meters of natural gas (CCM) savings, and percent of program savings for each customer segment. The percent of program savings represents the relative contribution that each customer segment makes to the overall result.

Enbridge's custom program overall achieved a 98% gross realization rate. The customer segment gross realization rates varied from 97% to 99%. The largest segment was Industrial with 52% of the population energy savings. Relative precision for the program overall was 3% at 90% confidence.

Table 3-1. First year gross savings realization rate for the Industrial Custom program

Segment	Gross Realization Rate	+/- at 90% Confidence	Sampled Measures	Population Measures	Population M3 Savings	Percent population M3 Savings
Agricultural	99.12%	1%	29	190	30,754,095	48%
Industrial	96.52%	5%	18	163	32,772,861	52%
Industrial Program	97.69%	3%	47	353	63,526,956	100%

3.2 Discrepancy summary

This section presents detailed results of the various project-level discrepancies between program claimed and evaluation verified savings. The verification found discrepancies in 28% of the projects reviewed.

Figure 3-1 shows that 34 of the 47 measures had no adjustment from program claimed to evaluation verified savings, while 13 measures were adjusted based on verification findings. For custom savings verification, we consider verified savings that differ more than 20% from utility tracking savings to be a "large" discrepancy. Moderate adjustments within 20% of utility

[3] A more detailed description of the program can be found in Enbridge's 2023 Demand Side Management Annual Report

tracking savings are expected given the level of uncertainty in forecasting energy savings. Eight of the 13 adjusted measures had verified savings within 20% of utility tracked savings. Of the four (4) measures with adjustments greater than 20%, 1 had adjustments increasing savings (adjustment greater than 120%) and three (3) had adjustments decreasing savings (adjustment less than 80%).

Figure 3-1. Adjusted realization rate (ARR) summary – Industrial Custom program

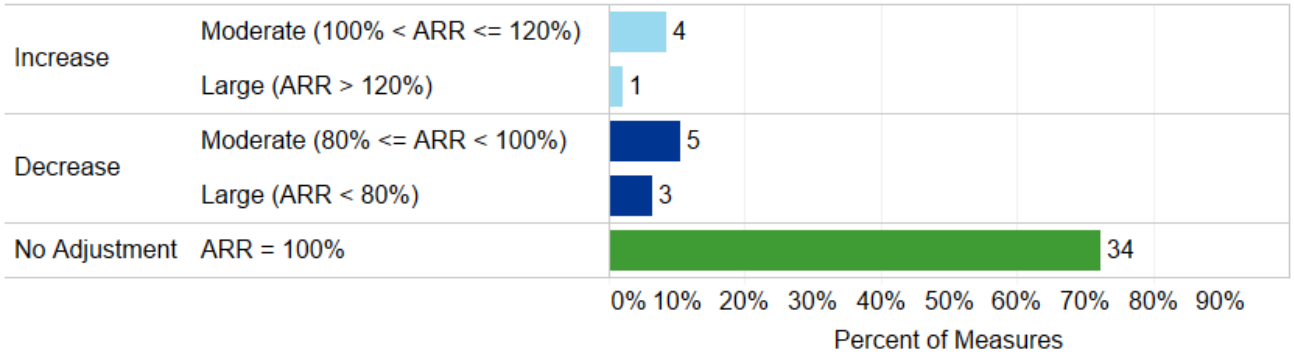


Figure 3-2 plots the claimed first year savings and the realization rate for each measure in the sample. The plot is sorted with the smallest measure on the bottom and largest on the top. The left plot shows the relative size of each measure. The right plot shows the gross realization rate for each measure. In both plots, measures with light blue bars have a realization rate greater than 100% (verified savings greater than utility claimed savings). Measures with dark blue bars represent a gross realization rate less than 100% (verified savings lower than utility claimed savings). Measures with green bars represent a gross realization rate of 100%.

Figure 3-2. Sample measure realization rates sorted by size – Industrial Custom program

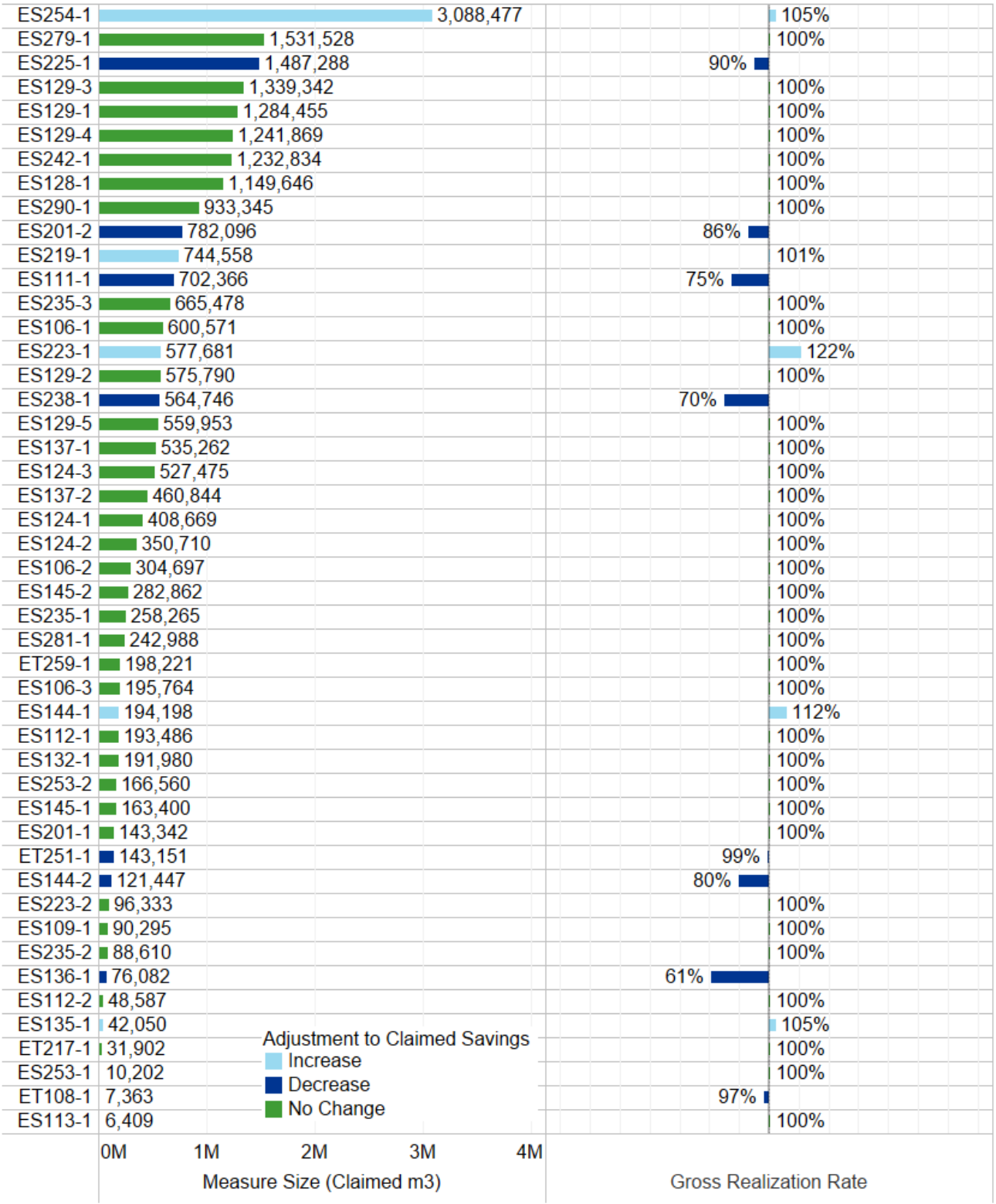
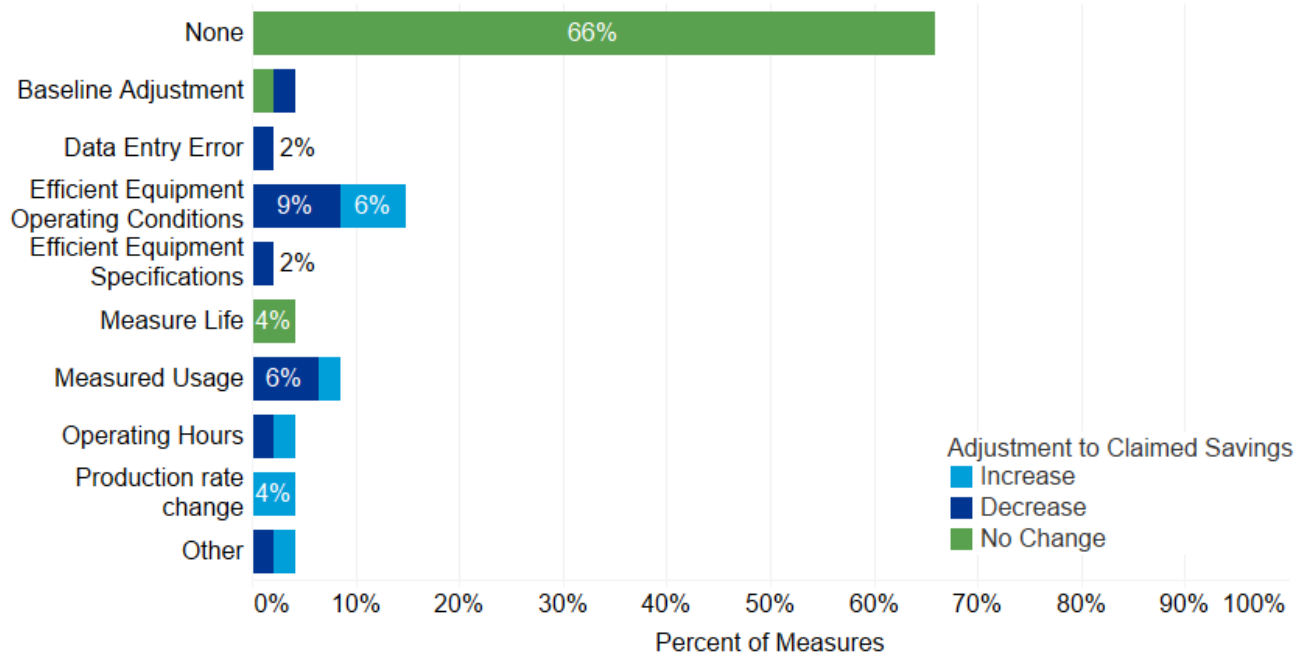


Figure 3-3 shows the types of discrepancies found by the verification. Each measure may have more than one discrepancy. The verification found no discrepancies for 66% of sampled measures. Operating conditions were the only type of discrepancy found for more than 10% of measures. The utility can reduce this type of discrepancy by documenting projects more thoroughly with sources for the assumptions used and more complete descriptions of conditions found at the time of installation (see recommendations in section 5); however, this type of discrepancy is partially outside of utility control. One measure had a baseline adjustment that did not impact savings, while two measures had measure life adjustments, but no discrepancy that affected first year savings.

Figure 3-3. Savings discrepancies – Industrial Custom program



4 LARGE VOLUME

Enbridge encourages the adoption of energy efficient equipment, technologies, and actions via its Large Volume program. In 2023, the Large Volume program was applicable to customers in Rate T2 or Rate 100.

The program uses a direct access budget mechanism for the customer incentive budget process. This mechanism collects funds from each customer through rates. Customers must use these funds to identify and implement energy efficiency projects, or the funds become available for use by other customers in the same rate class. This “use it or lose it” approach ensures each customer has first access to the amount of incentive budget funded by their rates. The Large Volume program is the only “direct access” program offered in Ontario.

Custom projects implemented as part of this program and claimed in 2023 were included in the CPSV study.

4.1 Gross savings realization rate

The GRR represents the ratio of the savings verified by the evaluation to the savings claimed (or reported) by the utility, as shown in the following equation. A 90% GRR means the verified gross savings for the project or program were 90% of the claimed savings. Differences between claimed and verified savings for each project can arise for a number of reasons, usually related to differences in forecast assumptions, differences in underlying facts, or differences in calculation approaches or parameters.

$$gross\ realization\ rate = \frac{evaluation\ verified\ savings}{utility\ reported\ savings}$$

Table 4-1 shows the cumulative gross savings realization rate for the Large Volume program. The table shows the gross realization rate, statistical precision at the 90% confidence interval, the program-claimed population CCM savings, and percent of program savings.

The Large Volume program overall had a 98% annual gross realization rate. The absolute precision (+/-) for the program was 11% at 90% confidence.

Table 4-1. First year gross savings realization rate for Large Volume program

Segment	Gross Realization Rate	+/- at 90% Confidence	Sampled Measures	Population Measures	Population M3 Savings	Percent population M3 Savings
Large Volume	98.07%	11%	12	31	80,549,726	100%

4.2 Discrepancy summary

This section presents detailed results of the various project-level discrepancies between program claimed and evaluation verified savings. The final realization rate for the program was 98% and the verification found discrepancies for 67% of the projects reviewed.

Figure 4-1 shows that four (4) out of 12 measures had no adjustment from program claimed to evaluation verified savings, while eight (8) measures were adjusted based on verification findings. For custom savings verification, we consider verified savings that differ more than 20% from utility tracking savings to be a “large” discrepancy. Moderate adjustments within 20% of utility tracking savings are expected given the level of uncertainty in forecasting energy savings. Three (3) of the eight (8) adjustments had verified savings within 20% of utility tracked savings. Of the five (5) measures with adjustments greater

than 20%, three (3) had adjustments increasing savings (adjustments greater than 120%) and two (2) had adjustments decreasing savings (adjustment less than 80%).

Figure 4-1. Adjusted realization rate (ARR) summary – Large Volume program

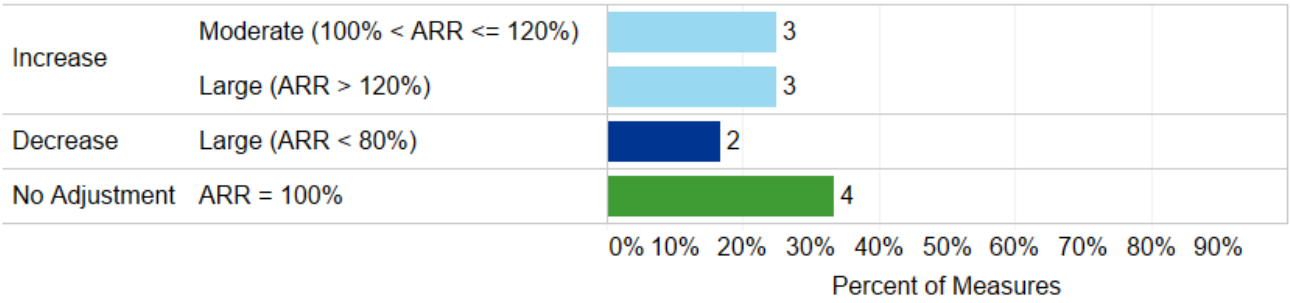


Figure 4-2 plots the claimed cumulative savings and the realization rate for each measure in the sample. The plot is sorted with the smallest measure on the bottom and largest on the top. The left plot shows the relative size of each measure. The right plot shows the gross realization rate for each measure. In both plots, measures with light blue bars have a realization rate greater than 100% (verified savings greater than utility claimed savings). Measures with dark blue bars represent a gross realization rate less than 100% (verified savings lower than utility claimed savings). Measures with green bars represent a gross realization rate of 100%.

Figure 4-2. Sample measure realization rates sorted by size – Large Volume program

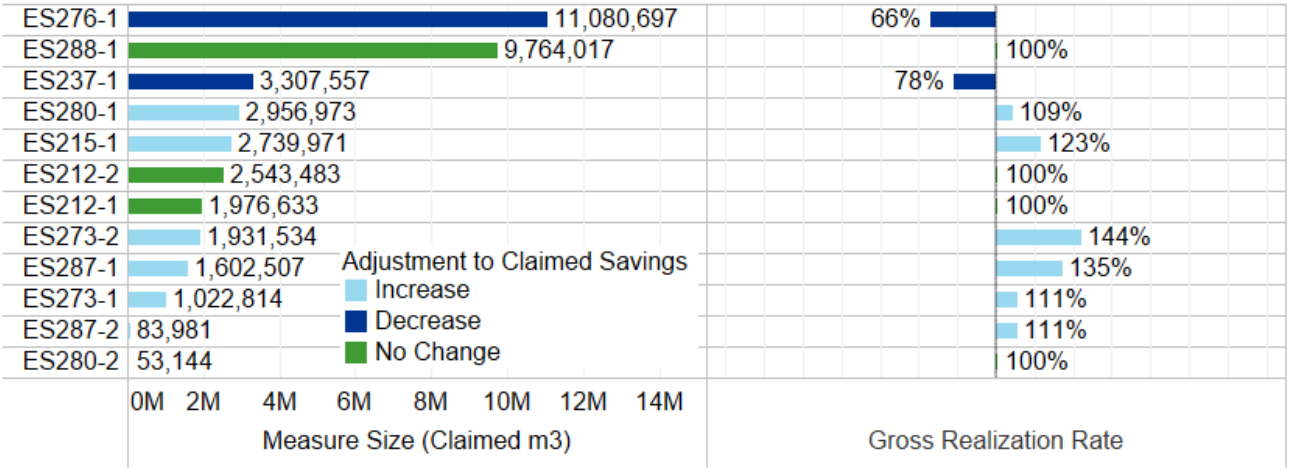
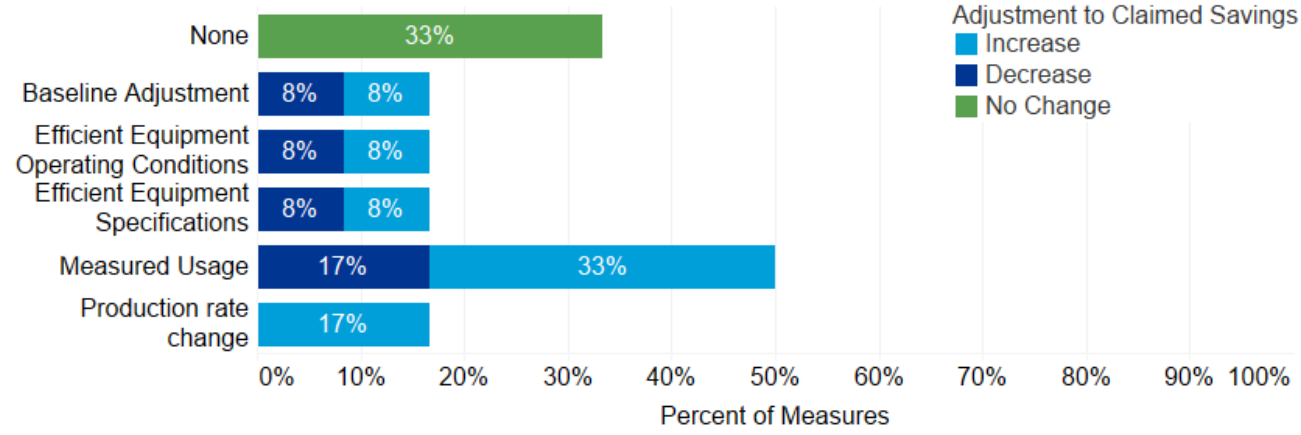


Figure 4-3 shows the types of discrepancies found by the verification. Each measure may have more than one discrepancy. The verification found no discrepancies for 33% of sampled measures. The most common discrepancy between claimed savings and verified savings (50% of measures) was updates to measured energy usage data provided by customers to the verification team. Savings based on measured energy usage are expected to result in some discrepancy during verification because the verification has access to a longer time period of post-installation data than the implementation team. In several cases the implementation team was working with very limited post-installation period data to model savings, which increases the risk of a large adjustment in verification.

Each of four other discrepancy types (baseline adjustment, EE equipment operating conditions, EE specifications and production rate changes) were found for 17% of measures. The utility can reduce this type of discrepancy by documenting projects more thoroughly with sources for the assumptions used and more complete descriptions of conditions found at the

time of installation (see recommendations in section 5); however, these types of discrepancies are partially outside of utility control.

Figure 4-3. Savings discrepancies – Large Volume program



5 FINDINGS AND RECOMMENDATIONS

Table 5-1 presents the key findings and recommendations from the study. It shows the party to whom the recommendation applies and the primary beneficial outcome of the recommendation. We classified outcomes into four categories: reduce costs, increase savings, increase (or maintain) customer satisfaction and decrease risk (multiple types of risk are in this category including risk of adjusted savings, risk to budgets or project schedules, and others). Details of the findings, recommendations and outcomes follow the table.

Table 5-1. Energy savings and program performance recommendations

			Applies to		Primary beneficial outcome			
#	Finding	Recommendation	Enbridge	Evaluation	Reduce costs	Increase savings	Customer satisfaction	Decrease risk
1	Enbridge continues to exhibit a strong commitment to accurate energy savings estimates.	Enbridge should continue its cultural commitment to accuracy.	✓			✓	✓	✓
2	The CPSV effort this year found realization rates between 90% and 100%	Continue performing custom savings verification on a regular basis.		✓				✓
		Consider approaches to sampling that can reduce sample sizes and costs.		✓	✓		✓	
3	Some measures in each utility program are routine maintenance, periodic repairs or like for like replacements that are considered standard care in other jurisdictions.	Establish a clear policy regarding eligibility of maintenance, repair and like for like replacement measures for the programs.	✓		✓			✓
4	The close relationships between Enbridge Energy Savings Consultants (ESCs) and customers provide advantages and challenges for evaluation.	Clarify the role of evaluation engineers, customers, and ESCs in the evaluation. Set and communicate clear expectations for each of the three roles so all parties are aligned.	✓	✓			✓	✓
5	Project documentation continues to improve.	Continue to improve data quality.	✓			✓		✓
6	Some Large Volume measures appear as two separate measure rows in the database due to having two sources of incentive funding.	Add a field to the tracking database to link two rows that are a single measure implementation.	✓		✓			

Finding 1: Enbridge continues to exhibit a strong commitment to accurate energy savings estimates. The utility has made significant investments in developing calculation tools that model savings accurately, such as the commercial and industrial Etools calculator, which is very thorough in attempting to model savings for key measures.

Enbridge's engineers have a strong understanding of their customers' building and process systems and show a commitment to finding accurate savings estimates. In this evaluation and in previous rounds of CPSV, the Enbridge engineering team has appropriately questioned evaluation findings that increased savings as well as those that decreased savings.

- **Recommendation 1:** Enbridge should continue its cultural commitment to accuracy.
- **Outcome 1:** Accurate energy savings.

Finding 2: The CPSV effort this year found realization rates between 90% and 100% and identified adjustments for 40 percent of projects. Across the programs, adjustments increased savings for 16 measures and decreased savings for 21 measures. 16 measures had a large adjustment (verified savings more than 20% different from tracked), which was a decrease from the 2017 verification.

- **Recommendation 2a:** Continue performing custom savings verification on a regular basis. Even a study that results in an adjustment of near 100% is still valuable because the programs know that their savings estimates will be reviewed. Knowing a review will be conducted improves the quality of ex ante estimates. The review itself also results in information that improves future program savings estimates.
- **Recommendation 2b:** Consider approaches to sampling that can reduce sample sizes and costs. Consistent realization rates of close to 100% are an indication that frequent smaller sample CPSV may provide the benefits cited in recommendation 2a while allowing for lower cost.
- **Outcome 2:** Accurate energy savings.

Finding 3: Some measures in each utility program are routine maintenance, periodic repairs or like for like replacements that are considered standard care in other jurisdictions.

- **Recommendation 3:** Establish a clear policy regarding eligibility of maintenance, repair and like for like replacement measures for the programs.
- **Outcome 3:** Reduced free-ridership risk.

Finding 4: The close relationships between Enbridge Energy Savings Consultants (ESCs) and customers provide advantages and challenges for evaluation.

A major advantage is that evaluation response rates were higher than they would have been otherwise due to ESC involvement in recruitment and regular attendance at site visits. Another advantage is that at some sites the ESC was able to help ensure both customers and evaluation engineers are talking about the same equipment or parameters, reducing miscommunication risk.

In evaluating some sites, the evaluation faced challenges ensuring that the data collected was coming from the customer rather than the ESC. Customers at times would defer to the ESC for some questions, which risks introducing confirmation bias and less independence for the evaluation.

- **Recommendation 4:** Clarify the role of evaluation engineers, customers, and ESCs in the evaluation. Set and communicate clear expectations for each of the three roles so all parties are aligned.
- **Outcome 4:** Independent and accurate evaluation with a positive customer experience.

Finding 5: Project documentation continues to improve. In this evaluation, some specific areas for improvement were identified:

- Project data or details missing
  - Basecase heating system details (quantities, efficiencies and conditions)
  - Trend data used for ex ante savings estimates
  - Measure loading order in Virtual Grower
- Measure descriptions not matching what was installed
- Use of black box tools
- Hardcoded information in calculation spreadsheets
- Undocumented assumptions and inputs
  - Values (such as CFM, temperature setpoints etc) provided with no documentation
- Insufficient access to customer data (by customers).
- **Recommendation 5:** Continue to improve data quality. Possible steps include:
  - Include explicit sources for all inputs and assumptions in the project documentation, with supporting evidence wherever possible
  - Store background studies and information sources with the project files and make them available to evaluators.
  - Provide evaluators full access to customer data.
  - Provide pre- and post-installation photos, where available.
  - Consider increasing documentation requirements for projects above certain incentive or gas savings amount
  - Institute a checklist as part of project closeout to ensure all relevant project documentation is assembled as ready for verification
- **Outcome 5:** Lower evaluation risk. Properly explaining and sourcing the savings calculation method and assumptions allows the evaluating engineer to more easily identify what needs to be verified. It also makes it easier to determine whether the methods and assumptions are reasonable and use ex ante assumptions rather than seek documented values elsewhere.

Finding 6: Some Large Volume measures appear as two separate measure rows in the database due to having two sources of incentive funding. These were not always easy to identify in the data.

- **Recommendation 6:** Add a field to the tracking database to link two rows that are a single measure implementation.
- **Outcome 6:** Consistent identification of multiple row measures will reduce re-work for sampling and expansion in the evaluation.

APPENDIX A. GLOSSARY OF TERMS AND KEY CONCEPTS

Term	Description
Adjustment factor	The adjustment factors are ratios of savings that allow evaluation findings from a sample of projects to be applied to and "adjust" the population of program savings. Realization rates, and ratios are other common terms.
Attribution	The energy savings or other benefits that are the result of a utility energy program's influence, including free ridership and spillover effects (see definitions in this Glossary).
Baseline, base case	Energy used / equipment in place if the program measure had not been done.
Building envelope	Exterior surfaces (e.g., walls, windows, roof, and floor) of a building that separate the conditioned space from the outdoors.
C&I	Commercial and industrial
Code	An action or standard required by local or federal laws for safety, environmental, or other reasons. For example, a building code that requires a minimum fuel efficiency for furnaces.
Cost effectiveness	Refers to the analysis that determines whether or not the benefits of a project/measure (see Glossary) are greater than the costs. It is based on the net present value of savings over the equipment life of the measure.
Cost effectiveness test - PAC	A test that compares the utility's avoided cost benefits with energy efficiency program expenditures (incentives plus administrative costs).
Cost effectiveness test – TRC-Plus	A test that compares benefits to society as a whole (avoided cost benefits plus non-energy benefits) with the participant's cost of installing the measure plus the cost of incentives and program administration.
Custom project savings verification (CPSV)	Activities related to the collection, analysis, and reporting of data for purposes of measuring gross custom program impacts.
Customer	Unique customers can be identified based on the account number and the contact information provided by Enbridge. A customer may have multiple site addresses, decision makers, and account numbers. Customers can only be identified for records for which we received contact information. i.e., records associated with account numbers that have measures in the sample or backup sample).
Customer Incentive	An incentive is a transfer payment from the utility to participants of a DSM program. Incentives can be paid to customers, vendors or other parties as part of a DSM program.
Demand side management (DSM)	Modification of perceived customer demand for a product through various methods such as financial incentives, education, and other programs
Domain	Grouping of like projects. A domain may be defined as projects within a specific sector or a category of measure types, end uses or other.
Dual baseline	Savings calculation approach which addresses or combines the savings associated with early replacement and the savings after the early replacement period. This concept is relevant to the measurement of lifetime gas savings (CCM) but not first-year annual savings.
Early replacement (ER)	Measure that replaces a piece of equipment that is not past EUL and in good operating condition.
Early replacement period (ER Period)	Years that the existing equipment would have continued to be in use. This is the same as RUL. This concept is relevant to the measurement of lifetime gas savings (CMM) but not first-year annual savings.
Effective useful life (EUL)	The length of time that a measure (see definition in Glossary) is expected to provide its estimated annual gas savings. EUL depends on equipment lifetime and measure persistence (see Glossary definition).
Energy solutions advisors	Energy Solutions Consultants (ESCs) work with customers on a one-to-one basis to address the unique processes and opportunities within each customer facility, identify energy savings opportunities and promote Enbridge's DSM offerings.
Estimated useful life (EUL)	Typically, the median number of years that the measure will remain in service.

Term	Description
Ex ante	Program claimed or reported inputs, assumptions, savings, etc.
Ex post	Program inputs, assumptions, savings, etc. which are verified after the claimed savings are finalized. Does not include assessment of program influence.
Free rider	A customer who would install or perform the same energy-saving measure (see definition in Glossary) without utility influence.
Free ridership	The portion of a program's verified energy savings that would naturally occur without the utility program.
Free ridership-based attribution	The portion of a program's verified energy savings that the utility influenced if one only considers free ridership and not spillover. Free ridership-based attribution is the complement of free ridership. (free ridership-based attribution = 100% - free ridership).
Gross savings	Gross savings are changes in energy consumption and/or demand directly caused by program-related actions by participants, regardless of reasons for participation (savings relative to baseline, defined above).
In situ	Existing measure, conditions, and settings.
In-depth interviews	Structured technical interviews administered by evaluation engineers and market researchers either in person or more frequently, over the phone, IDIs offer more flexibility than CATIs and are best leveraged for complex projects and topics.
Incentive	An incentive is often a payment from the utility to participants of a DSM program. Incentives can be paid to customers, vendors, or other parties.
Incremental cost	The difference in purchase price (and any differences in related installation or implementation costs), at the time of purchase, between the energy-saving measure (see Glossary definition) and the base case measure. In some early retirements and retrofits, the full cost of the efficient technology is the incremental cost.
Industry standard practice (ISP)	Common measure implemented within the industry.
Input assumptions	A common practice used within an industry but not formally defined by code or regulation.
Lifetime cumulative savings	Total natural gas savings (CCM) over the life of a DSM measure. It can be claimed, gross, or net. Sometimes referred to as just "cumulative" or "lifetime."
Maintenance (Maint.)	Repair, maintain, or restore to prior efficiency.
Measure	Equipment, technology, practice, or behavior that, once installed or working, results in a reduction in energy use. Measures are identified in the tracking data as unique line items for which savings within a custom project are quantified. Multiple measures may belong to the same project.
Measure persistence	How long a measure remains installed and performs as originally predicted in relation to its EUL. This considers events like business turnover, early retirement of installed equipment, and other reasons measures might be removed or discontinued.
Measurement and Verification (M&V)	Verification of savings using methods not including attribution/free ridership assessment.
Metric	This is a term used by the OEB to measure a utility's program achievement. Under the DSM framework, programs are grouped into categories, called scorecards. Each program within a scorecard is assigned at least one metric that is used to measure utility performance. The metric for many programs is annual savings, or a reduction in natural gas consumption, while other programs have non-savings metrics such as the number of program participants. Within each scorecard, various metrics are combined to produce an overall scorecard achievement.
MF	Multifamily (multi-residential)
Net-to-gross	The ratio of net energy savings to gross savings. The NTG ratio is applied to gross program savings to convert them into net program savings.
New construction (NC)	New buildings or spaces.
Non-early replacement period (non-ER period)	Years after the ER period up to the EUL.

Term	Description
Non-energy impacts	Sometimes called non-energy benefits, these are the wider socio-economic or environmental outcomes that arise from energy efficiency improvements, aside from energy savings. NEIs can include but are not limited to impacts such as improved safety, improved health, and job creation. For example, offering participants may benefit from increased property value, and improved health and comfort. The TRC-Plus test includes a 15% adder to the benefits calculation to account for NEIs.
Normal replacement (NR)	Measure that replaces a piece of equipment that is past EUL and in good operating condition.
Offering	One or more DSM activities or measures which a utility may use to affect a specifically identified target market in their choices around the amount and timing of energy consumption.
Persistence	The extent to which a DSM measure remains installed and performing as originally predicted in relation to its EUL.
Portfolio	A group of DSM programs which have been selected and combined in order to achieve the objectives of a utility's DSM Plan.
Program	The programs outlined in Enbridge's Multi-Year Plan are comprised of one or more offerings and address the needs of a subset of Enbridge's customer base.
Program evaluation	Activities related to the collection, analysis, and reporting of data for purposes of measuring program impacts from past, existing, or potential program impacts.
Program spending	The amount spent running energy-savings programs, not including the costs of running (called overhead costs) the larger portfolio of programs. This value can be divided into spending for program measures and incentives, as well as program-specific costs.
Project	Projects are identified in the tracking data based on the project code. A project may have multiple measures as indicated by sub-codes in the current data tracking system.
Rate class	The OEB establishes distribution rate classes for Enbridge. Distribution rate classes group customers with similar energy profiles.
Realization rate	A combination of adjustment factors, which represents ratios between two savings values. For example, the final realization rate is the ratio between evaluated savings and program claimed savings.
Remaining useful life (RUL)	The number of years that the existing equipment would have remained in service and in good operating condition had it not been replaced. This is the same as the ER period.
Replace on burnout (ROB)	Measure that replaces a failed or failing piece of equipment.
Retrofit add-on (REA)	Measure that reduces energy use by modifying an existing piece of equipment.
Scorecard	A scorecard allows for multiple different kinds of metrics such as cumulative natural gas savings and/or participants enrolled to be used simultaneously to measure annual utility performance. Each utility has a scorecard identified for each program year, which can be found in the Ontario Energy Board Decision and Order EB-2021-0002.
Scorecard Achievement	The verified value for program-specific metric targets (annual savings, applications, etc.) of each scorecard identified by the Annual Scorecard. This is the value that is verified as the achieved value by the Annual Verification report and used for calculation of the shareholder incentive.
Shareholder Incentive	As part of the current DSM Framework, an annual performance incentive is available to the gas utilities in the event program performance is at or above 75% of the OEB-approved targets up to a maximum of 125%.
Site	Sites are identified based on unique site addresses provided by Enbridge through the contact information data request. A site may have multiple units of analysis, measures, and projects. Sites can be identified by the evaluation only for records for which we receive a site id.

Term	Description
Spillover effects	These are reductions in energy consumption and/or demand that occur as a result of the presence of a utility DSM program, but are beyond program-related savings and are not part of the utility's verified savings. These effects could result from many factors including additional efficiency actions that program participants take outside the program as a result of having participated, changes in store availability of energy-using equipment, and changes in energy use by program non-participants as a result of utility program advertising.
System optimization (OPT)	Improve system or system settings to exceed prior efficiency.
TRM	Technical Resource Manual, which is a document that identifies standard methodologies and inputs for calculating energy savings.
TSER	Telephone-supported engineering review.
Unit of analysis	The level at which the data are analyzed, which in 2023 will likely be a "measure" or sub-project level for Enbridge.
Vendors	Program trade allies, business partners, contractors, and suppliers who work with program participants to implement energy saving measures.

APPENDIX B. FINAL SAMPLE ACHIEVEMENT

Commercial Custom: Summary of participant data collection

Table B-1 summarizes the CPSV data collection efforts for the Commercial Custom Program. The table shows the portion of the program that:

- Completed on-site visits
- Completed telephone supported engineering reviews (TSER)
- Did not respond to an evaluation attempt at contact, or refused verification
- Was not contacted by the evaluation team.

The data collected is represented as the number of sites, the number of measures, and first-year ex ante natural gas savings (ex ante m3). The proportion of the program in each category is also represented in Table B-2. In the table, size categories within segments (e.g. Industrial) are ordered with 1 being the smallest stratum within each segment. The study had a customer response rate of 75%, which is higher than recent comparable studies in central North America.

Table B-1. Summary of CPSV data collection for Commercial Custom program

Data collection category	Targeted # measures	# sites	# measures	Ex ante m^3
Completed On-Site		7	9	5,795,697
Completed TSER		23	23	812,751
Attempted Contact, Not Completed		10	35	4,808,969
Not Attempted		763	1,019	18,709,438
Total	35	803	1,086	30,126,855

Table B-2. CPSV sample achievement for Commercial program

Segment	Stratum	Max m3	Target	Completed Measures	Frame Measures	Completed m3	Population m3
Commercial	1	14,944	4	3	253	22,282	1,763,968
	2	40,413	4	2	93	41,584	2,182,322
	3	107,135	4	4	43	286,769	2,617,012
	4	814,596	3	3	13	1,335,067	2,768,314
	5	3,936,225	3	2	3	3,897,845	7,834,070
Low Income and Multi-Residential Multi-Family	1	13,539	5	5	398	16,698	2,409,060
	2	27,085	4	4	152	84,479	2,978,868
	3	57,673	4	3	89	141,943	3,396,812
	4	219,965	4	5	41	533,643	3,928,293
	5	248,136	1	1	1	248,136	248,136
			Grand Total	32	1,086	6,608,447	30,126,855

Industrial Custom: Summary of participant data collection

Table B-3 summarizes the CPSV data collection efforts for the Industrial Custom Program. The table shows the portion of the program that:

- Completed on-site visits
- Completed telephone supported engineering reviews (TSER)
- Did not respond to an evaluation attempt at contact, or refused verification
- Was not contacted by the evaluation team.

The data collected is represented as the number of sites, the number of measures, and first-year ex ante natural gas savings (ex ante m3). The proportion of the program in each category is also represented in Table B-4. In the table, size categories within segments (e.g., Industrial) are ordered with 1 being the smallest stratum within each segment. The study had a customer response rate of 86%, which is higher than comparable studies in central North America.

Table B-3. Summary of CPSV data collection for Industrial program

Data collection category	Targeted # measures	# sites	# measures	Ex ante m^3
Completed On-Site		26	43	24,262,540
Completed TSER		4	4	380,637
Attempted Contact, Not Completed		5	81	13,712,491
Not Attempted		131	225	25,171,288
Total	44	166	353	63,526,956

Table B-4. CPSV sample achievement for Industrial program

Segment	Stratum	Max m3	Target	Completed Measures	Frame Measures	Completed m3	Population m3
Agricultural	1	78,988	4	4	114	97,100	3,232,999
	2	191,980	4	7	32	965,634	4,231,024
	3	298,378	4	5	20	1,124,575	4,805,489
	4	600,571	4	6	12	3,037,117	5,380,262
	5	1,171,415	4	3	7	2,380,919	6,288,144
	6	1,418,983	3	3	4	3,865,666	5,284,649
	7	1,531,528	1	1	1	1,531,528	1,531,528
Industrial	1	112,851	4	4	98	221,828	4,085,872
	2	267,939	4	3	27	584,360	5,203,176
	3	408,669	4	2	18	759,379	6,084,045
	4	702,366	4	4	11	2,372,268	6,109,767
	5	1,232,834	3	3	7	3,127,038	6,714,235
	6	3,088,477	2	2	2	4,575,765	4,575,765
			Grand Total	47	353	24,643,177	63,526,956

Large Volume: Summary of participant data collection

Table B-5 summarizes the participant data collection efforts for CPSV of the Large Volume program. The table shows the portion of the program that:

- Completed on-site visits
- Did not respond to an evaluation attempt at contact, or refused verification
- Was not contacted by the evaluation team.10

The data collected is represented as the number of sites, the number of measures, and cumulative ex ante natural gas savings (ex ante CCM). The proportion of the program in each category is also represented in Table C-6. In the table, size categories are ordered with 1 being the smallest stratum. The study had a customer response rate of 80%, which is higher than recent comparable studies in central North America.

Table B-5. Summary of CPSV data collection for Large Volume program

Data Collection Category	Targeted # Measures	# Sites	# Measures	Ex Ante m3
Completed On-Site		8	12	39,063,311
Attempted Contact, Not Completed		2	9	33,687,899
Not Attempted		8	9	7,798,516
Total	14	18	30	80,549,726

Table B-6. CPSV sample achievement for Large Volume program

Segment	Stratum	Max m3	Target	Completed Measures	Frame Measures	Completed m3	Population m3
Large Volume	1	1,913,687	3	4	16	2,762,446	8,426,602
	2	2,539,584	3	2	5	3,908,167	11,059,954
	3	2,956,973	3	3	4	8,240,427	10,795,261
	4	15,193,375	5	3	5	24,152,271	50,267,909
			Grand Total	12	30	39,063,311	80,549,726

APPENDIX C. ADDITIONAL RESULTS

Table C-1. Cumulative cubic meter realization rate – Commercial Program

Segment	Gross Realization Rate	+/- at 90% Confidence	Sampled Measures	Population Measures	Population CCM Savings	Percent population CCM Savings
Commercial	83.87%	15%	14	405	337,913,454	56%
Low Income and Multi-Residential Multi-Family	97.10%	6%	18	681	267,391,282	44%
Commercial Program	90.25%	9%	32	1,086	605,304,736	100%

Table C-2. Cumulative cubic meter realization rate – Industrial Program

Segment	Gross Realization Rate	+/- at 90% Confidence	Sampled Measures	Population Measures	Population CCM Savings	Percent population CCM Savings
Agricultural	99.64%	1%	29	190	476,096,535	46%
Industrial	99.07%	8%	18	163	556,156,911	54%
Industrial Program	99.30%	5%	47	353	1,032,253,447	100%

Table C-3. Cumulative cubic meter realization rate – Large Volume Program

Segment	Gross Realization Rate	+/- at 90% Confidence	Sampled Measures	Population Measures	Population CCM Savings	Percent population CCM Savings
Large Volume	89.64%	18%	12	31	741,145,146	100%

Table C-4. First year gross savings realization rate – Commercial Program – non-FPC precision

Segment	Gross Realization Rate	+/- at 90% Confidence	Sampled Measures	Population Measures	Population M3 Savings	Percent population M3 Savings
Commercial	91.13%	11%	14	405	17,165,686	57%
Low Income and Multi-Residential Multi-Family	97.61%	8%	18	681	12,961,169	43%
Commercial Program	93.85%	7%	32	1,086	30,126,855	100%

Table C-5. First year gross savings realization rate – Industrial Program – non-FPC precision

Segment	Gross Realization Rate	+/- at 90% Confidence	Sampled Measures	Population Measures	Population M3 Savings	Percent population M3 Savings
Agricultural	99.12%	2%	29	190	30,754,095	48%
Industrial	96.52%	6%	18	163	32,772,861	52%
Industrial Program	97.69%	3%	47	353	63,526,956	100%

Table C-6. First year gross savings realization rate – Large Volume Program – non-FPC precision

Segment	Gross Realization Rate	+/- at 90% Confidence	Sampled Measures	Population Measures	Population M3 Savings	Percent population M3 Savings
Large Volume	98.07%	17%	12	31	80,549,726	100%

APPENDIX D. SITE-LEVEL SAVINGS RESULTS

Table E-1. Commercial site-level savings results

Measure Number	Tracking m3	Verified First Year m3	First year m3 Realization Rate	Verified Average Annual m3	Tracking CCM	Verified CCM	CCM Realization Rate
1	2,772,823	2,772,823	100%	2,772,823	55,456,460	55,456,460	100%
2	1,125,022	1,125,022	100%	1,125,022	16,875,330	16,875,330	100%
3	814,596	534,240	66%	534,240	20,364,900	10,684,800	52%
4	406,048	406,048	100%	406,048	2,030,240	2,030,240	100%
5	248,136	191,870	77%	191,870	6,203,408	4,180,575	67%
6	219,965	236,846	108%	236,846	3,299,475	3,552,690	108%
7	121,727	100,611	83%	100,611	1,825,905	1,509,165	83%
8	114,423	116,842	102%	116,842	686,538	701,052	102%
9	93,287	93,287	100%	93,287	559,722	559,722	100%
10	77,949	57,153	73%	57,153	1,169,242	857,295	73%
11	71,558	93,504	131%	93,504	1,073,377	1,402,560	131%
12	70,659	72,724	103%	72,724	1,766,475	1,454,480	82%
13	61,982	61,982	100%	61,982	1,549,545	1,549,550	100%
14	58,411	58,411	100%	58,411	876,164	876,165	100%
15	53,047	53,047	100%	53,047	1,326,181	1,326,175	100%
16	50,311	46,536	92%	46,536	1,257,779	1,163,256	92%
17	44,874	52,353	117%	52,353	673,110	785,295	117%
18	38,584	38,066	99%	38,066	964,608	951,650	99%
19	25,763	25,763	100%	25,763	644,087	644,075	100%
20	25,144	25,144	100%	25,144	628,604	628,600	100%
21	23,913	23,913	100%	23,913	597,814	597,825	100%
22	17,903	17,903	100%	17,903	447,566	447,575	100%
23	17,520	17,520	100%	17,520	437,995	438,000	100%
24	15,821	8,866	56%	8,866	158,208	88,660	56%
25	11,944	11,999	100%	11,999	179,154	179,985	100%
26	8,854	8,854	100%	8,854	53,124	53,124	100%
27	6,556	6,401	98%	6,401	163,904	160,025	98%
28	4,772	0	0%	0	71,580	0	0%
29	2,535	2,535	100%	2,535	35,490	35,490	100%
30	1,510	1,427	95%	1,427	22,651	21,405	94%
31	1,484	1,484	100%	1,484	37,105	37,100	100%
32	1,325	0	0%	0	33,131	0	0%

Table E-2. Industrial site-level savings results

Measure Number	Tracking m3	Verified First Year m3	First year m3 Realization Rate	Verified Average Annual m3	Tracking CCM	Verified CCM	CCM Realization Rate
1	3,088,477	3,250,975	105%	3,250,975	61,769,540	65,019,500	105%
2	1,531,528	1,531,528	100%	1,531,528	15,315,278	15,315,280	100%
3	1,487,288	1,343,760	90%	1,343,760	29,745,760	26,875,200	90%
4	1,339,342	1,339,340	100%	1,339,340	20,090,130	20,090,100	100%
5	1,284,455	1,284,454	100%	1,284,454	25,689,100	25,689,080	100%
6	1,241,869	1,241,851	100%	1,241,851	24,837,380	24,837,020	100%
7	1,232,834	1,232,834	100%	1,232,834	24,656,680	24,656,680	100%
8	1,149,646	1,149,646	100%	1,149,646	22,992,920	22,992,920	100%
9	933,345	933,344	100%	933,345	4,666,725	4,666,725	100%
10	782,096	675,529	86%	675,529	3,910,480	3,377,645	86%
11	744,558	754,424	101%	754,424	14,891,160	15,088,480	101%
12	702,366	524,248	75%	713,026	14,047,320	14,260,520	102%
13	665,478	665,478	100%	665,478	6,654,780	6,654,780	100%
14	600,571	600,571	100%	600,571	12,011,420	12,011,420	100%
15	577,681	702,731	122%	702,731	11,553,620	14,054,620	122%
16	575,790	575,790	100%	575,790	11,515,800	11,515,800	100%
17	564,746	394,882	70%	580,900	11,294,920	11,618,000	103%
18	559,953	559,954	100%	559,954	8,399,295	8,399,310	100%
19	535,262	535,262	100%	535,262	8,028,930	8,028,930	100%
20	527,475	527,475	100%	527,475	10,549,500	10,549,500	100%
21	460,844	460,844	100%	460,844	9,216,880	9,216,880	100%
22	408,669	408,669	100%	408,669	8,173,380	8,173,380	100%
23	350,710	350,710	100%	350,710	3,507,100	3,507,100	100%
24	304,697	304,697	100%	304,697	3,046,970	3,046,970	100%
25	282,862	282,863	100%	282,863	5,657,240	5,657,260	100%
26	258,265	258,265	100%	258,265	5,165,300	5,165,300	100%
27	242,988	242,988	100%	242,988	2,429,880	4,373,784	180%
28	198,221	198,221	100%	198,221	3,964,420	2,775,094	70%
29	195,764	195,658	100%	195,658	1,957,640	1,956,580	100%
30	194,198	217,946	112%	217,946	3,883,960	4,358,920	112%
31	193,486	193,486	100%	193,486	2,902,290	2,902,290	100%
32	191,980	191,979	100%	191,979	2,879,700	2,879,685	100%
33	166,560	166,560	100%	166,560	1,665,600	1,665,600	100%
34	163,400	163,400	100%	163,400	2,451,000	2,451,000	100%
35	143,342	143,342	100%	143,342	2,866,840	2,866,840	100%
36	143,151	142,174	99%	142,174	858,906	853,044	99%
37	121,447	97,700	80%	97,700	2,428,940	1,954,000	80%
38	96,333	96,333	100%	96,333	577,998	577,998	100%

Measure Number	Tracking m3	Verified First Year m3	First year m3 Realization Rate	Verified Average Annual m3	Tracking CCM	Verified CCM	CCM Realization Rate
39	90,295	90,297	100%	90,297	1,354,425	1,354,455	100%
40	88,610	88,610	100%	88,610	886,100	886,100	100%
41	76,082	46,238	61%	29,396	1,521,640	587,920	39%
42	48,587	48,587	100%	48,587	485,870	485,870	100%
43	42,050	44,072	105%	44,072	841,000	881,440	105%
44	31,902	31,902	100%	31,902	319,020	319,020	100%
45	10,202	10,202	100%	10,202	204,040	204,040	100%
46	7,363	7,153	97%	7,153	103,082	100,142	97%
47	6,409	6,408	100%	6,408	128,180	128,160	100%

Table E-3. Large volume site-level savings results

Measure Number	Tracking m3	Verified First Year m3	First year m3 Realization Rate	Verified Average Annual m3	Tracking CCM	Verified CCM	CCM Realization Rate
1	11,080,697	7,346,807	66%	7,346,807	221,613,941	146,936,140	66%
2	9,764,017	9,764,017	100%	9,764,017	48,820,085	48,820,085	100%
3	3,307,557	2,571,884	78%	2,571,884	66,151,140	51,437,680	78%
4	2,956,973	3,236,350	109%	3,236,350	2,956,973	3,236,350	109%
5	2,739,971	3,382,338	123%	1,690,863	8,219,913	5,072,589	62%
6	2,543,483	2,543,483	100%	2,543,483	2,543,483	2,543,483	100%
7	1,976,633	1,976,633	100%	1,976,633	11,859,798	11,859,798	100%
8	1,931,534	2,772,774	144%	2,772,774	1,931,534	2,772,774	144%
9	1,602,507	2,158,358	135%	2,158,358	40,062,675	53,958,950	135%
10	1,022,814	1,139,069	111%	1,139,069	15,342,210	17,086,035	111%
11	83,981	93,149	111%	93,149	1,679,620	1,862,980	111%
12	53,144	53,144	100%	53,144	318,864	318,864	100%

APPENDIX E. KEY DOCUMENTS

The Scope of Work embedded below includes the technical background and methodology used in this study. Also provided are the sample design memo and the site report template used for reporting site specific results.

Scope of Work

Scope of Work for OEB CPSV 2023

Sample Design Memo

CPSV Sample Design (2023 wave 2

Site report template

Site report template for OEB CPSV 2023

DNV

CUSTOM PROJECTS SAVINGS VERIFICATION FOR ONTARIO'S NATURAL GAS COMMERCIAL AND INDUSTRIAL DSM PROGRAMS

Scope of Work

Ontario Energy Board

Date: December 17, 2023

Table of contents

List of figures

List of tables

1 OVERVIEW

This document has been prepared for the Ontario Energy Board (OEB) and provides the scope of work for the Custom Program Savings Verification (CPSV) of Enbridge Gas Inc.'s (Enbridge) natural gas demand-side management (DSM) programs delivered in 2023. The study will produce verified savings ratios and verified gross savings totals for the custom projects in the Enbridge programs examined, shown in Table 1-1.

Table 2. CPSV by Program

Program	2023 Gross Verification
Large Volume	✓
Commercial*	✓
Industrial	✓
Low Income Multi-Residential	✓

**Custom Market-Rate Multi-Residential (Multifamily) projects are expected to be included as a part of this program.

1.1 Evaluation Objectives

The overall objectives of this project are to:

Develop accurate verified gross savings for each of Enbridge's custom commercial, industrial, multi-residential (including low-income), and large volume programs carried out in 2023, with disaggregated rates for each of the major program components within these groupings (for example differentiated by segment/technology type and to be determined in consultation with the EC, OEB staff and EAC at the commencement of the study).

- Establish and maintain transparency throughout the project.
- Follow industry best practices.
- Achieve 90/10 precision[4] at the requested stratification segment levels.

1.2 Evaluation Approach

The evaluation approach will differ from the previous evaluations in that the data collection will be divided into two phases. The primary advantage of dividing the evaluation into two separate phases is to reduce the risk of CPSV reporting causing a delay in the 2023 Annual Verification. By evaluating a portion of sites in the first quarter (Q1) of the year, we will have less sites to recruit, visit and report on during Q2 and Q3.

The methodology selected for the CPSV study consists of engineer reviews of gross savings. Reviews of complex projects will include on-site verification and data collection, while less complex projects will be verified with Telephone Supported Engineering Reviews (TSERs).

[4] 90/10 precision refers to 10% relative precision with 90% confidence.

1.3 Deliverables

This study will result in one final deliverable:

- 2023 Custom Gross Savings Verification Summary Report

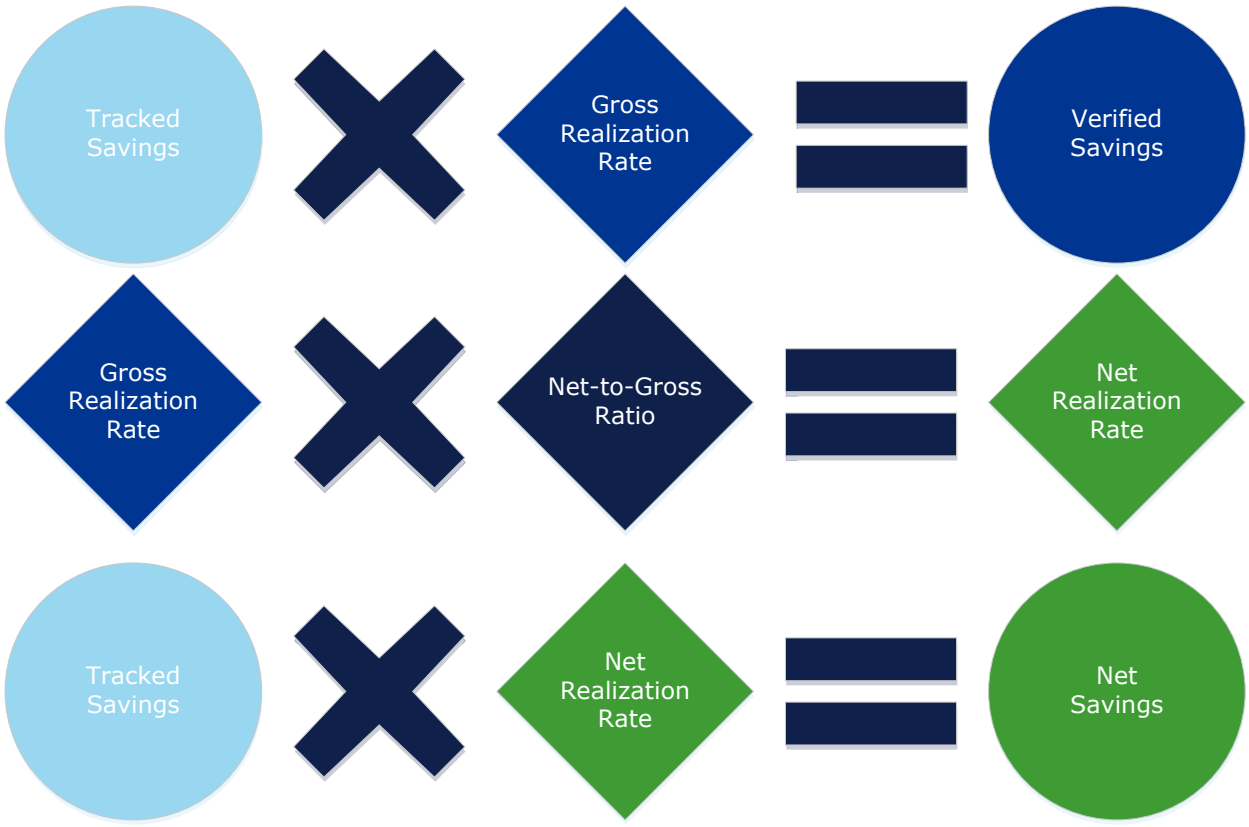
Interim deliverables will include:

- Workplan
- Sample Design Memo (phase 1 and updated for phase 2)
- Presentation of workplan and sample designs
- Documentation and contact information requests for sample and backup
- Advance letter
- Site verification reports, including live calculation worksheets
- Comment matrices for comments received on the workplan and final report

2 METHODOLOGY

The study methodology uses the efforts of the CPSV analysis to produce an adjustment factor, called the gross realization rate (RR), that can be applied to the reported savings data (or tracked savings) to produce the verified gross savings. Figure 4 shows how the gross RR is applied to the tracking savings to produce the verified gross savings. The figure also shows the net-to-gross and net realization rates, which will be applied in conjunction with findings from this evaluation as shown.

Figure 4. Ratios used to estimate verified and net savings



The analysis is built on calculating the ratio of verified gross savings to the tracking estimate of gross savings for implemented measures. The gross realization rate includes corrections to the numbers of units installed, changes in operating hours, changes in operating levels, etc.

The next sections describe the process used to develop the RR from the engineering adjustments in greater detail. They also describe the process for expanding the results of the sample to the population, and the methodology for adjustment factors.

2.1 Realization Rate

The RR is developed through data collected during the CPSV effort, which will verify achieved gross savings for measures at a sample of sites.

For an individual measure the engineering verification factor is derived from the data collected during the participant survey data collection for TSER projects and through the on-site visits for other projects. Differences between the reported measure and the measure installed at the facility are accounted for here. The engineering adjustment factor is the ratio of the evaluator-verified savings to the program-reported savings.

The majority of the CPSV process involves determining the evaluator-verified savings estimate for each measure. The measure-level results are then combined using weights from the sample design to an overall adjustment factor.

To get the evaluation-verified savings for each evaluated measure, the CPSV effort will verify savings based on the applicable baseline(s) and measure life based on the best available information.

DNV will use a dual baseline approach for estimating lifetime and first year energy savings. Figure 5 shows how we will assemble the verified savings for each measure.

Notation:

VGS_S = Verified Gross Savings based on Standard efficiency equipment baseline (annual)

VGS_E = Verified Gross Savings based on pre-existing equipment baseline (annual)

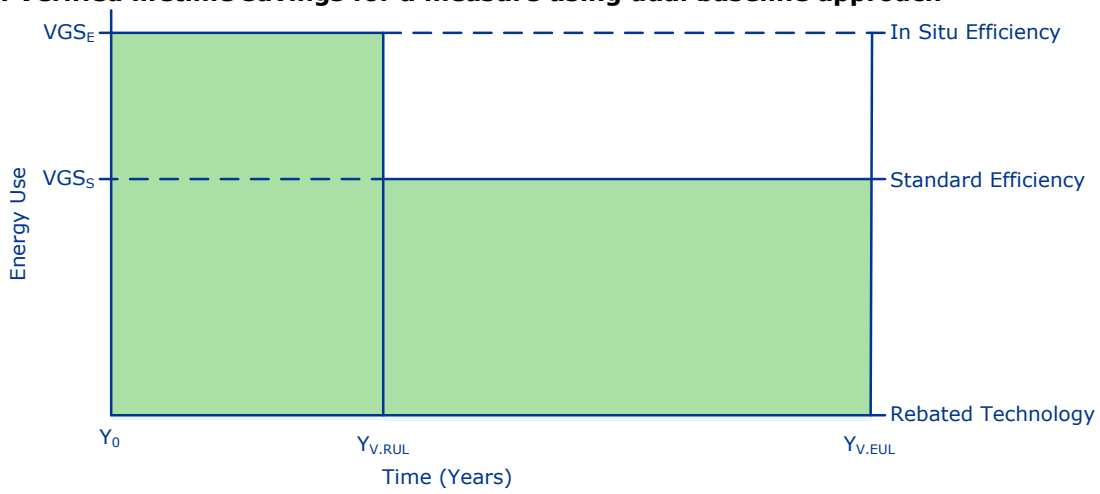
VGS_L = Verified Gross Lifetime Savings

Y_0 = Year of measure implementation

$Y_{V.EUL}$ = Verified Estimated Useful Life (Years) of installed efficient equipment

$Y_{V.RUL}$ = Verified Remaining Useful Life (Years) of replaced equipment[5]

Figure 5. Verified lifetime savings for a measure using dual baseline approach



The verified lifetime savings are calculated as the difference in energy usage of the incentivized measure and the energy use of the in-situ measure for the remaining useful life of the in-situ measure plus the verified savings based on the standard baseline measure for the rest of the (verified) life of the new measure.

Equation 1.

$$\boldsymbol{VGS_L = VGS_E \times Y_{V.RUL} + VGS_S \times (Y_{V.EUL} - Y_{V.RUL})}$$

The verified first year savings could be calculated one of two ways. The first is to set verified first year savings as equal to as the difference in energy usage of the incentivized measure and the energy use of the in-situ measure (equation 2a). For replace on burnout measures the verified savings are based on the standard (historically the standard has been code or minimum viable available alternative) baseline measure (equation 2b).

[5] RUL of existing equipment is also applicable as defining the estimated useful life for some retrofit add-on measures

Equation 2a. $$\boldsymbol{VGS_{FY-ADV} = VGS_E}$$

Equation 3b. $$\boldsymbol{VGS_{FY-ROB} = VGS_S}$$

The second approach is calculated as the verified lifetime savings divided by the (verified) life of the new measure (equation 3). This approach would be used for both advancement and replace on burnout measures. It is an average annual savings rather than a literal first year savings.

Equation 3. $$\boldsymbol{VGS_{FY} = (VGS_E \times Y_{V.RUL} + VGS_S \times Y_{V.EUL})/Y_{V.EUL}}$$

A decision on which approach will be used will follow the kickoff meeting.

2.2 Sample Expansion

Samples are a necessary part of program evaluation. Sampling reduces costs and customer burden. Nonresponse, whether due to a lack of desire to respond, or because the person that should respond cannot, means that evaluating the entire population usually cannot be done. Any time we evaluate a sample of savings from a program, we must expand the sample results to the population. Expanding the results to the population produces results that are representative of the population rather than the sample. Expansion is a key part of calculating important program metrics such as total verified gross savings. More detail on sample expansion is provided in Appendix B .

Expansion is done using weights that are determined based on the sample design. The weight is a numeric quantity associated with each responding unit and conceptually represents the amount of the target population the responding unit represents during the analysis. The sample weight is some function of the total number of units in the sample frame. In the CPSV study, the sample weight will be built from the inverse probability of selection, incorporating additional adjustment factors to account for nonresponse and coverage errors (such as a lack of completes in a specific sampling stratum).

Notation:

N_x = number of units of analysis in stratum X

n_x = number of completed sample units of analysis in stratum X

The weight W_x is calculated as:

$$W_x = N_x/n_x$$

The method used to develop the verified savings will not affect the weight. In the CPSV, each level of rigour is measuring the same thing (verified savings), only varying in their level of detail. In this case, we are looking at energy savings with reliable, valid methods that avoid systematic bias, but with additional magnification on the largest, most variable projects. It is similar to measuring a length using millimetres or eighths of an inch. Both provide accurate measurements of length, but the millimetre measurement is more precise. In terms of expansion, both measurements would get equal weights (once put into comparable units, of course).

DNV uses the ratio estimation method to expand our results to the population. The energy saving estimates (tracking savings, installed savings, or verified savings) of the sampled units (measures, projects, sites) are present in both the numerator and the denominator of the ratios. When combined with the sample weights, the ratio estimation method produces unbiased, savings weighted adjustment factors.

The ratio estimator calculated for this study (the gross realization rate) is a weighted sum of verified savings divided by the weighted sum of tracking savings. The mathematics of ratio estimation and an example calculation can be found in Ratio Estimation.

3 SUMMARY OF ACTIVITIES

The DNV team has divided the project into six distinct functional activities which are presented in Table 3. These activities are discussed in greater detail in the following sections.

Table 3. Key activities

Key Activities
Step 1: Project Kickoff
Convene a project kickoff meeting
Solicit OEB and EAC feedback on draft scope of work
Step 2: Sample Design
Explore the tracking data
Define the unit of analysis
Stratify the data
Design the sample
Select the sample
Step 3: Data Collection
Request and collect project documentation
On-site verification of a sample of projects
Telephone Supported Engineering Review (TSER) of a sample of projects
Step 4: Data Analysis
Analyze TSER and on-site data
Calculate estimates
Step 5: Reporting
Monthly status reports
Bi-monthly updates
Draft deliverables
Final report and presentation
Step 6: Project Management
Complete evaluation on time, on budget and within scope
Keep OEB and EAC informed on progress

3.1 Project Kickoff

DNV will host a project kickoff with OEB and EAC. Discussion at the kickoff and written comments provided by EAC members will inform updates to this workplan.

3.2 Sample Design

At the kickoff meeting, DNV plans to engage the OEB and EAC in an up-front discussion on the options for sample design and reporting categories. Based on this discussion, we plan to complete a draft sample design that will be provided to the EAC for review.

The key guiding principles for the sample design approach that we plan to discuss with the EAC include:

3. Independent free ridership, spillover and CPSV sample designs.

    a. The integrated analysis approach across the three custom studies requires different timing of data collection to adhere to best practices for each study type. In addition, the free ridership and spillover studies will need to evaluate more than double the number of measures as studied in the CPSV, limiting the cost savings of a nested

net and gross approach. Samples for net ratios (free ridership and spillover) will be drawn independently from gross verification samples and independent from one another due to timing issues.

   b. Independent sample designs allow us to provide different stratification options to the EAC for the gross and net samples. This should increase the precision of each study without increasing sample sizes due to a combined sampling stratification. For example, the net sample will not require sampling by rigour level of the CPSV.

   c. A more straightforward sample design for each study will be easier for stakeholders to understand and use, while also reducing complexity for data collection recruitment.

4. A sample based on categories found in Enbridge tracking databases or simple aggregations thereof. We will work with the EAC to define strata and reporting domains that are meaningful to the results, while making mapping of those strata and reporting domains to the utility datasets as seamless as possible.

5. The sample will be divided into two phases. This will reduce the risk of CPSV delaying annual verification in 2023. The second phase sample will integrate the completed and refused sites from phase 1 into the final design.

6. Limit customer burden while collecting data cost-effectively. The most recent CPSV of a single program year limited measures per site to four, while the 2017-18 CPSV limited measures evaluated at a site to three. DNV plans to revert to assessing a maximum of four measures per site in the 2023 CPSV. Based on prior CPSV, we anticipate few sites will require the maximum number of measures, but will re-assess in the final sample design.

A CPSV sample design memo for each wave of data collection will be provided and integrated into the final scope of work.

3.3 Data Collection

Data collection for the program includes the interviews with program managers and staff (completed in September 2023); TSER interviews with program participants; and on-site verification at participating customer sites. Any interviews with program staff are for informational purposes only. CPSV results will be based on data collected directly from participating customers.

Objectives

The objective of the data collection step is to collect:

- Program manager and staff information on program services to inform other data collection efforts
- On-site and telephone data from participants about equipment and operations to inform the CPSV

Activities

Each of the data collection activities support verifying gross energy savings.

1. Program orientations with Enbridge staff focused on gross verification information, including programs, facility types and efficiency measures.

2. Program Participants are the primary source of data for the verification.

   a. On-site visits will collect data to support verification of gross savings estimates (on-site sample).

   b. Telephone Supported Engineering Reviews will be implemented in lieu of an on-site visit for sites where the cost of visiting the site outweighs the benefit of in person observation.

3. Participating vendors are a secondary source for the gross study. Vendors may be contacted to provide technical details where the customer indicates they would be better able to answer.

Follow up with participants and/or vendors via phone or email may be required to acquire additional detail not provided during the initial data collection.

Table 4 is a summary of the targeted completes (customers/sites, rather than measures) by data collection type for each phase.

Table 4. Estimated Target Number of Completed Interviews and Site Visits

Target Group	Estimated Number of Interviews/Visits
Planning Phase	
Program Orientation Interviews	4
Phase 1 Data Collection – 2023 Q1 & Q2	
Participant TSERs	8
Participant Site Visits	12
Phase 2 Data Collection – 2023 Q3 & Q4	
Participant TSERs	22
Participant Site Visits	34

3.3.1 Program Orientation

Technical orientation. In order to better understand the calculation tools Enbridge uses for custom measures, DNV will meet with program staff who use and develop the tools. These meetings will ensure the project team has a full understanding of the primary calculation tools employed.

3.3.2 Participant Data Collection and Review

Participant data collection will be a combination of in-depth-interviews and on-site visits.

The principal activities will consist of the following.

Assign initial rigour level. As part of the sample design process, a preliminary rigour level will be assigned to each measure in the population based on measure type, size, and prevalence in the program. The initial level will be updated throughout the calculation planning process as detailed in the activities below.

Request project documentation. Following the primary and backup sample selection, the DNV team will request project documentation from Enbridge. The documentation should include “live” calculation workbooks (with formulas and links) or input files for specific software programs (such as building models), incentive application forms, invoices and supporting documents, and contact information for technical staff at the participating firm. Project documentation will be requested for all sampled and backup measures as well as non-sampled, non-backup measures at sites that have other measures in the sample/backup. Measures not included in the sample/backup will not be verified unless their verification is required as part of the verification of a sampled measure (i.e., the measures are inter-related). In the case of measures that are verified and not required in the sample, appropriate weights will be developed to account for their non-random selection (consistent with the approach used in previous verifications).

Send advance letters. Prior to data collection, DNV will work with Enbridge to send letters (by traditional mail and email to all customers selected for the primary and backup sample, notifying them of the study and asking for their cooperation.

Emails will be sent from utility email addresses and traditional mail will be sent in utility envelopes and signed by utility representatives.

Assign sites to engineering teams. The DNV data collection lead will assign sites to individual engineers. Some sites (such as eTools or Virtual Grower sites) will be assigned to specialists; others based on the type of measure and expertise of the engineer. The assigned engineer will be responsible for the evaluation of that site from assessing the project documentation through producing the final site report, with support from others in their team. Stantec field staff will work with the DNV teams to collect information from sites where on-sites are performed.

A DNV engineering team member will review the documentation for each measure in the primary and backup sample for completeness. A checklist of items provided for each project will be sent back to the utility with a follow-up request for any identifiable missing or incomplete information. We expect that the need to provide additional documentation after this follow-up will be rare. The engineer will use the information provided to update the preliminary rigour assessment based on the complexity of the calculation method and the evaluation's likely ability to access required information from site contacts.

To reduce potential bias and costs associated with completing and reviewing site reports based on incomplete project documentation files, the utility must provide all supporting project documentation within two business days of a completed site visit or TSER. We will notify Enbridge as on-site visits are scheduled and TSERs are completed. This is the last opportunity to provide supporting documentation for the savings calculations. Additional information provided after this—either in written or verbal form—cannot be included in calculations. Rare exceptions to this rule have occurred in the past that have allowed additional information during EAC site review. As in previous CPSVs, DNV expects to assess these exceptions on a case by case basis with the EAC. The final code of conduct for site visits will be attached in Appendix C following EAC review.

On-sites only: recruit and schedule sites. If the data collection plan dictates that a participant receives an on-site visit, the next step is to recruit the site. Stantec staff will call program participants and ask if they're willing to receive an evaluation visit. If the site agrees, the Stantec recruiter will schedule the on-site visit and identify possible times prior to the visit for a follow-up phone call to gather additional information for the site-specific M&V plan (this call will in most cases be made by the assigned DNV engineer). The Stantec recruiter also will send an email to the utility informing them of the date and time of the visit. Consistent with past CPSV, an Enbridge representative may attend CPSV site visits as an observer.

Develop the site-specific M&V plan. DNV engineering team members will produce site-specific M&V plans for all sites in the gross verification sample. They will review the project documentation in greater depth, identify the key savings inputs to research, and develop a data collection plan specific to that site. The plan will include the data collection approach to be used, the expected savings estimation methodology, and a backup approach for when the requested data is not available. For measures with standard calculation approaches, DNV may first develop a standardized data collection plan. All plans will focus on collecting the information necessary to confidently estimate cumulative energy savings, such as hours of operation, equipment setpoints, equipment schedules, facility usage patterns, and standard O&M activities. Special attention will be paid to whether the remaining useful life of pre-existing equipment limits the EUL for the measure. All plans will be reviewed and approved by DNV's engineering team leads prior to data collection. A summary table from the plans will be provided in the site report.

If necessary: complete a TSER (phone call). Telephone calls will be used as the only primary data collection mode for TSER sites and, if necessary, as a planning tool for on-site visits. There are three general types of calls:

- *TSER sites:* for a TSER-only site, a DNV engineering team member will complete an interview with the technical contact for the participating site. The engineer will verify the team's understanding of the measure and collect data or verify

calculation inputs as required by the M&V plan. If necessary, the engineer will follow up with vendors for additional information.

- *Pre-site plan TSER communication:* If data collection is required prior to the site visit, a DNV engineering team member will complete an interview with the technical contact for the participating site. The engineer will verify the team's understanding of the measure and ask about equipment access, data availability, or other information that will inform the M&V plan. Email exchanges may also be used in lieu of or in addition to phone interviews.
- *Post-site plan TSER communication:* If on-site data collection is improved by a phone call after the M&V plan but prior to the site visit, a Stantec engineer will complete an interview with the technical contact at the participating site. These types of TSERs are likely to be completed with sites that have large numbers of measures or where specific data will require preparation by the site contact.

On-sites only: complete the site visit. Stantec engineers will complete the site visits with program participants. The engineer will attempt to physically verify the measure installation and view the associated systems. The engineer will also collect data as required by the M&V plan. Where direct measurement is required, engineers may be required to return the site to retrieve measurement equipment. The field engineer will transfer site notes and data to DNV no later than the Friday following the site visit.

Estimate verified savings and complete site report. The DNV engineering team member responsible for the site's evaluation will use the data from the on-site or TSER to calculate verified savings and complete the site verification report. They will update the calculations with current operating parameters, where they differ from the values used by the utility. Weather-sensitive measures will receive savings based on government-defined typical weather patterns. Where building simulation models are used, DNV will calibrate the model to monthly consumption data and weather files. As necessary, the DNV engineering team member will work in conjunction with their engineering team lead, site modelling experts, and industrial process experts to ensure accurate results. Applicable results from other custom project evaluation studies such as the eTools boiler study will be incorporated transparently in the verified savings and site report. Live calculation spreadsheets will be provided to the EAC.

We will use the same site report previously approved by the EAC and used in the two prior CPSV evaluations.

Complete technical review. Each site report will undergo a technical review conducted by a senior engineer familiar with the Ontario custom C&I programs. The reviewer will ensure there are no avoidable weaknesses in the technical approach, that descriptions are clear, and that approaches are consistent for similar measures. The review will consider:

- Is the measure correctly described?
- Is the calculation method appropriately identified and described?
- Were inputs adequately verified?
- Was anything overlooked?
- Was the planned rigour threshold met?

Complete final consistency review. After the technical review, each site report will undergo a final consistency review by a senior member of the project team. The reviewer will ensure there are no weaknesses in the technical approach and descriptions, there is consistency in our approach and language across similar measures, and the site form conforms to the OEB style guide.

Deliver the draft site report for review. DNV will deliver the draft site reports to the EAC for review in batches every 2 weeks. The number of site reports in each batch will depend on schedule. We ask the EAC to have comments delivered within two weeks of receipt.

The frequency and timing of the batches will be included in the EC cross-project activities schedule.

Meet to discuss EAC comments. The DNV team will have minimum of two days to review the comments for each batch before hosting a discussion (by phone) with the EAC. On the call, the DNV team will be represented by the project sponsor (Ben Jones). We ask that a representative of the EC team also attend these calls.

Finalize the site report. After the EAC site report call, the DNV engineering reviewers will work with the engineering teams to address the remaining comments and finalize the site reports. The final site reports will be uploaded to the project SharePoint site and included in the draft study report.

Summarize site-level results. DNV will summarize the results in a table of all tracked and verified final savings for sampled measures, including realization rates, high level reasons for discrepancy and documentation of changes made following the EAC meeting to discuss the site report. Summary tables with tracked and verified final savings for sampled measures, realization rates, high level reasons for discrepancy will be included as an appendix in the draft and final reports.

Table 5 shows an example timeline to complete the gross verification for a site. Each on-site measure is expected to take approximately seven weeks to complete, including review and revision.

Table 5. Example timeline to complete gross verification for a site

Step	Week 1					Week 2					Week 3					Week 4					Week 5					Week 6					Week 7				
	M	T	W	Th	F	M	T	W	Th	F	M	T	W	Th	F	M	T	W	Th	F	M	T	W	Th	F	M	T	W	Th	F	M	T	W	Th	F
Schedule site visit or interview																																		
Develop M&V plan																																		
Pre-visit TSER (where needed)																																		
Deliver updated site plan to Stantec																																		
Review M&V plan with Stantec																																		
Onsite visit																																		
Completed site M&V data delivered																																		
Calculate savings																																		
Draft site report																																		
Technical review																																		
Final QC review																																		
Draft delivered to EAC																																		
EAC Review																																		
Review EAC Comments																																		
Discuss at EAC Meeting																																		
Update based on EAC comments																																		
Upload Final Report																																		
Stantec																																		
DNV																																		
EAC																																		

Deliverables

- Draft and final advance letter
- Draft and final site verification reports including calculation worksheets

3.4 Data Analysis

The data analysis step takes the data collected in Step 3 and combines it into adjustment factors that represent the population of implemented measures. Those adjustment factors are then applied to the program-level savings to produce verified gross savings.

Objectives

- Determine the population-weighted adjustment factors related to verified gross savings
- Apply the adjustment factors to the appropriate program-reported savings estimates
- Produce the overall verified gross savings

Activities

3.4.1 Analyze Data

We will use the sampling weights created during the sample design process to expand the customer sample in each stratum to represent the full participant population in that stratum. Targeted strata for which we are unable to obtain any responses will either be treated as not represented by the sample or will be collapsed with other cells for sample expansion.

3.4.2 Calculate Estimates

The gross verification will result in verified gross savings that are calculated for each evaluated measure by evaluation engineers. DNV will use the corresponding sample weights and ratio estimation to expand the sample results to the population in each stratum. Strata without responses will either be treated as not represented by the population or collapsed with other strata for sample expansion.

3.5 Reporting

The reporting step encompasses the formal communication between the DNV CPSV team and the OEB/EAC. Reporting includes status and update reports as well as the draft and final reports, which take the results of the analysis from Step 4 and presents them to the OEB, EAC, and other interested stakeholders. The original plan called for 10 calls with the OEB and EAC to discuss deliverables from the Steps 2-5, with five (5) of these 10 meetings to be focused on verification site reports (roughly 20 measures per meeting). Three (3) of the four (4) meetings were planned for discussion of the evaluation plan/sample design (project kickoff) and final gross savings report. The remaining meeting was a contingency to address specific issues that come up in the process.

In addition to meetings, we have built in review time (2 weeks wherever possible) for the EAC to provide comments on key interim and final deliverables including:

- Workplan
- Sample design memo
- All gross savings verification site reports
- Final report

Matrices of comments received, and responses will be provided for all EAC reviewed draft documents, with the exception of the gross savings verification site reports. EAC comments on site reports will be addressed on EAC calls dedicated to site reports, with changes noted in a final gross savings spreadsheet that will be provided with the draft report.

3.5.1 Monthly Status Reports

Every month the DNV project manager will submit a status report to the OEB, via email, which will summarize the past month's activities, notify of the next month's activities, and report on how closely the evaluation is adhering to the original

schedule. However, if there are methodological questions or delays in responses to data requests that could put the evaluation off schedule, the program manager will notify the OEB of these issues immediately for proposed resolution so that the evaluation schedule is not compromised.

The EC will provide a status report to the EAC at every scheduled EAC meeting.

3.5.2 Weekly Status Updates

The DNV project manager will provide the OEB with study weekly updates via teleconference. We will use our SharePoint communication tools to update dashboard indicators on a weekly basis.

3.5.3 Draft Reports

At the conclusion of the evaluation, DNV will submit to the OEB and EAC one draft report that will present all the information in the research objectives.

The report (2023 Verified Savings Report) will include verified savings for Enbridge Gas Inc.'s 2023 Custom programs.

Also included will be verification rates by market sectors, programs, and domains of interest with associated precision estimates.

Along with these key findings, the report will also show how these estimates were derived and what data from the TSERs and on-sites were used to inform the estimates, including any qualitative findings regarding non-incentive based utility services provided through the custom programs.

3.6 Project Management

The project management step is an ongoing step to ensure proper implementation of the project, including the schedule, budget, and scope.

Objectives

- Ensure timely and on-budget deliverables
- Keep the OEB informed of project progress

This step is ongoing over the course of the project, and includes budget and workflow tracking, communication among DNV GL team members and partner firms, and invoicing. The subsequent sections discuss the project timeline and risks to effective project implementation.

3.6.1 Stakeholder Expectations and EAC Review Approach

Whenever possible we plan to provide two weeks of review time for deliverables with deadlines for draft deliverable delivery and EAC comments clearly communicated via the EC SharePoint site. With the exception of CPSV Site Reports, the final deliverables will be accompanied by a comment matrix that includes our response to each comment received.

For utility data and documentation requests, we will work with Enbridge, the OEB and the EAC to establish reasonable deadlines based on the timing of the request. We will communicate in advance when a request will arrive.

3.6.2 Project Timeline

A consolidated schedule of all projects overseen by the Evaluation Contractor can be found on the OEB-EAC SharePoint site.

3.6.3 Risks and Contingencies

The live risk register can be found on the OEB-EAC SharePoint site.

APPENDIX A. GLOSSARY OF TERMS AND KEY CONCEPTS

Adjustment factor	The adjustment factors are ratios of savings that allow evaluation findings from a sample of projects to be applied to and "adjust" the population of program savings. Realization rates, and ratios are other common terms.
Attribution	The energy savings or other benefits that are the result of a utility energy program's influence, including free ridership and spillover effects (see definitions in this Glossary).
Baseline, base case	Energy used / equipment in place if the program measure had not been done.
Building envelope	Exterior surfaces (e.g., walls, windows, roof, and floor) of a building that separate the conditioned space from the outdoors.
C&I	Commercial and industrial
Code	An action or standard required by local or federal laws for safety, environmental, or other reasons. For example, a building code that requires a minimum fuel efficiency for furnaces.
Cost effectiveness	Refers to the analysis that determines whether or not the benefits of a project/measure (see Glossary) are greater than the costs. It is based on the net present value of savings over the equipment life of the measure.
Cost effectiveness test - PAC	A test that compares the utility's avoided cost benefits with energy efficiency program expenditures (incentives plus administrative costs).
Cost effectiveness test – TRC-Plus	A test that compares benefits to society as a whole (avoided cost benefits plus non-energy benefits) with the participant's cost of installing the measure plus the cost of incentives and program administration.
Custom project savings verification (CPSV)	Activities related to the collection, analysis, and reporting of data for purposes of measuring gross custom program impacts.
Customer	Unique customers can be identified based on the account number and the contact information provided by Enbridge. A customer may have multiple site addresses, decision makers, and account numbers. Customers can only be identified for records for which we received contact information. i.e., records associated with account numbers that have measures in the sample or backup sample).
Customer Incentive	An incentive is a transfer payment from the utility to participants of a DSM program. Incentives can be paid to customers, vendors or other parties as part of a DSM program.
Demand side management (DSM)	Modification of perceived customer demand for a product through various methods such as financial incentives, education, and other programs
Domain	Grouping of like projects. A domain may be defined as projects within a specific sector or a category of measure types, end uses or other.
Dual baseline	Savings calculation approach which addresses or combines the savings associated with early replacement and the savings after the early replacement period. This concept is relevant to the measurement of lifetime gas savings (CCM) but not first-year annual savings.
Early replacement (ER)	Measure that replaces a piece of equipment that is not past EUL and in good operating condition.
Early replacement period (ER Period)	Years that the existing equipment would have continued to be in use. This is the same as RUL. This concept is relevant to the measurement of lifetime gas savings (CMM) but not first-year annual savings.
Effective useful life (EUL)	The length of time that a measure (see definition in Glossary) is expected to provide its estimated annual gas savings. EUL depends on equipment lifetime and measure persistence (see Glossary definition).
Energy solutions advisors	Energy Solutions Consultants (ESA) work with customers on a one-to-one basis to address the unique processes and opportunities within each customer facility, identify energy savings opportunities and promote Enbridge's DSM offerings.
Estimated useful life (EUL)	Typically, the median number of years that the measure will remain in service.
Ex ante	Program claimed or reported inputs, assumptions, savings, etc.
Ex post	Program inputs, assumptions, savings, etc. which are verified after the claimed savings are finalized. Does not include assessment of program influence.

Free rider	A customer who would install or perform the same energy-saving measure (see definition in Glossary) without utility influence.
Free ridership	The portion of a program's verified energy savings that would naturally occur without the utility program.
Free ridership-based attribution	The portion of a program's verified energy savings that the utility influenced if one only considers free ridership and not spillover. Free ridership-based attribution is the complement of free ridership. (free ridership-based attribution = 100% - free ridership).
Gross savings	Gross savings are changes in energy consumption and/or demand directly caused by program-related actions by participants, regardless of reasons for participation (savings relative to baseline, defined above).
In situ	Existing measure, conditions, and settings.
In-depth interviews	Structured technical interviews administered by evaluation engineers and market researchers either in person or more frequently, over the phone, IDIs offer more flexibility than CATIs and are best leveraged for complex projects and topics.
Incentive	An incentive is often a payment from the utility to participants of a DSM program. Incentives can be paid to customers, vendors, or other parties.
Incremental cost	The difference in purchase price (and any differences in related installation or implementation costs), at the time of purchase, between the energy-saving measure (see Glossary definition) and the base case measure. In some early retirements and retrofits, the full cost of the efficient technology is the incremental cost.
Industry standard practice (ISP)	Common measure implemented within the industry.
Input assumptions	A common practice used within an industry but not formally defined by code or regulation.
Lifetime cumulative savings	Total natural gas savings (CCM) over the life of a DSM measure. It can be claimed, gross, or net. Sometimes referred to as just "cumulative" or "lifetime."
Maintenance (Maint.)	Repair, maintain, or restore to prior efficiency.
Measure	Equipment, technology, practice, or behavior that, once installed or working, results in a reduction in energy use. Measures are identified in the tracking data as unique line items for which savings within a custom project are quantified. Multiple measures may belong to the same project.
Measure persistence	How long a measure remains installed and performs as originally predicted in relation to its EUL. This considers events like business turnover, early retirement of installed equipment, and other reasons measures might be removed or discontinued.
Measurement and Verification (M&V)	Verification of savings using methods not including attribution/free ridership assessment.
Metric	This is a term used by the OEB to measure a utility's program achievement. Under the DSM framework, programs are grouped into categories, called scorecards. Each program within a scorecard is assigned at least one metric that is used to measure utility performance. The metric for many programs is annual savings, or a reduction in natural gas consumption, while other programs have non-savings metrics such as the number of program participants. Within each scorecard, various metrics are combined to produce an overall scorecard achievement.
MF	Multifamily (multi-residential)
Net-to-gross	The ratio of net energy savings to gross savings. The NTG ratio is applied to gross program savings to convert them into net program savings.
New construction (NC)	New buildings or spaces.
Non-early replacement period (non-ER period)	Years after the ER period up to the EUL.

Non-energy impacts	Sometimes called non-energy benefits, these are the wider socio-economic or environmental outcomes that arise from energy efficiency improvements, aside from energy savings. NEIs can include but are not limited to impacts such as improved safety, improved health, and job creation. For example, offering participants may benefit from increased property value, and improved health and comfort. The TRC-Plus test includes a 15% adder to the benefits calculation to account for NEIs.
Normal replacement (NR)	Measure that replaces a piece of equipment that is past EUL and in good operating condition.
Offering	One or more DSM activities or measures which a utility may use to affect a specifically identified target market in their choices around the amount and timing of energy consumption.
Persistence	The extent to which a DSM measure remains installed and performing as originally predicted in relation to its EUL.
Portfolio	A group of DSM programs which have been selected and combined in order to achieve the objectives of a utility's DSM Plan.
Program	The programs outlined in Enbridge's Multi-Year Plan are comprised of one or more offerings and address the needs of a subset of Enbridge's customer base.
Program evaluation	Activities related to the collection, analysis, and reporting of data for purposes of measuring program impacts from past, existing, or potential program impacts.
Program spending	The amount spent running energy-savings programs, not including the costs of running (called overhead costs) the larger portfolio of programs. This value can be divided into spending for program measures and incentives, as well as program-specific costs.
Project	Projects are identified in the tracking data based on the project code. A project may have multiple measures as indicated by sub-codes in the current data tracking system.
Rate class	The OEB establishes distribution rate classes for Enbridge. Distribution rate classes group customers with similar energy profiles.
Realization rate	A combination of adjustment factors, which represents ratios between two savings values. For example, the final realization rate is the ratio between evaluated savings and program claimed savings.
Remaining useful life (RUL)	The number of years that the existing equipment would have remained in service and in good operating condition had it not been replaced. This is the same as the ER period.
Replace on burnout (ROB)	Measure that replaces a failed or failing piece of equipment.
Retrofit add-on (REA)	Measure that reduces energy use by modifying an existing piece of equipment.
Scorecard	A scorecard allows for multiple different kinds of metrics such as cumulative natural gas savings and/or participants enrolled to be used simultaneously to measure annual utility performance. Each utility has a scorecard identified for each program year, which can be found in the Ontario Energy Board Decision and Order EB-2021-0002.
Scorecard Achievement	The verified value for program-specific metric targets (annual savings, applications, etc.) of each scorecard identified by the Annual Scorecard. This is the value that is verified as the achieved value by the Annual Verification report and used for calculation of the shareholder incentive.
Shareholder Incentive	As part of the current DSM Framework, an annual performance incentive is available to the gas utilities in the event program performance is at or above 75% of the OEB-approved targets up to a maximum of 125%.
Site	Sites are identified based on unique site addresses provided by Enbridge through the contact information data request. A site may have multiple units of analysis, measures, and projects. Sites can be identified by the evaluation only for records for which we receive a site id.
Spillover effects	These are reductions in energy consumption and/or demand that occur as a result of the presence of a utility DSM program, but are beyond program-related savings and are not part of the utility's verified savings. These effects could result from many factors including additional efficiency actions that program participants take outside the program as a result of having participated, changes in store availability of energy-using equipment, and changes in energy use by program non-participants as a result of utility program advertising.

System optimization (OPT)	Improve system or system settings to exceed prior efficiency.
TRM	Technical Resource Manual, which is a document that identifies standard methodologies and inputs for calculating energy savings.
TSER	Telephone-supported engineering review.
Unit of analysis	The level at which the data are analyzed, which in 2023 will likely be a “measure” or sub-project level for Enbridge.
Vendors	Program trade allies, business partners, contractors, and suppliers who work with program participants to implement energy saving measures.

APPENDIX B. SAMPLE EXPANSION AND RATIO ESTIMATION

B.1 Sample Weights

This appendix describes how we calculate the sample weights for each stratum. In lay terms, the weight is simply the number of units in the sample frame (N) divided by the number of completed units in the sample (n). The interpretation of the weight is that each completed sample unit represents N/n units in the population (sample frame).

Notation:

N_x = number of units of analysis in stratum X

n_x = number of completed sample units of analysis in stratum X

The weight W_x is calculated as:

$$W_x = N_x / n_x$$

We can understand the weight as meaning the response for one sampled unit in stratum X is representative of *Wx* units in the population. Table 6 shows a simple example. In the example, we completed 2 surveys with participants in the "North" and 10 surveys with participants in the "South." The weight for the "Northerners" is greater than that of the "Southerners," but because we completed more surveys with "Southerners" the combined weight of the "South" will be in proportion to its share of the population (both the population and sum of weights is 20).

Table 6. Example Sample Weights

Stratum Definition	Sample Frame (N)	Sample Completes (n)	Weight (W)	Interpretation
North	10	2	5 = 10/2	Each response represents 5 Northern participants
South	20	10	2 = 20/10	Each response represents 2 Southern participants

Without sample weights, the data collected from the "North" would be 17 percent (2/12) of the final result, while with weights, the "North" is 33 percent (10/30). The un-weighted result would be less accurate than the weighted result if the measured value differs along North/South lines. For example, if the "North" is more conservative than the "South" then political surveys without sample weights would end up with inaccurate results. If responding to surveys is negatively correlated with conservatism, then the weights help correct for the systemic bias in response rates.

The sample weight associated with an observation is consistent regardless of the segmentation of the data that we report by (reporting domains). This means that we can segment the data multiple ways in the report, with the final overall results remaining consistent no matter the domain.

Special Cases

There are some special cases where the sample weight for a project needs to be set to one (1) in order to use the data collected without biasing the result. Our sample design targets measures within a site and sample weights are developed at that level as well. When we collect data from a customer, we will collect data on all of a customer's sampled and primary backup measures in a single IDI or site visit. This maximizes the data collected on each customer contact, without overburdening multi-measure customers, but requires special handling to ensure that extra data collected does not bias the sample. To eliminate the potential bias of over representing multiple measure sites, we first identify units that were

completed as an add-on when another measure was selected for a site. With the planned process, there will be limited numbers of “extra” measures collected.

For each stratum in our sample design, the units are randomly ordered for selection in a list. If seven units are targeted for the stratum, then the first seven units on the list are the primary sample and the rest of the list comprises the full backup sample (when we request project documentation, we will restrict the backup sample for the request to reduce burden on utility staff). If a site has two measures in different strata and one is selected in the primary sample, we will request documents on both measures and ask about both, regardless of whether the second measure is in the primary or backup sample in its stratum. After collecting data on both measures, we will assess whether the second measure was selected in its stratum based on how far down the list we had to go to complete our target. If the second measure’s spot on the list was selected, then the measure will be counted as a normal complete and included in the stratum’s N/n weight calculation. If the measure’s spot on the list did not come up, the data collected for the measure will be used, but the measure will not be included in the N/n weight for its strata. Instead, it will be given a weight of 1 so that it represents itself and no other measures. For variance estimates, the measure will remain in its sampled stratum.

Table 7 provides an example. Both site A and site B had measures in Stratum X selected in the sample. Each responded to our interview. Both sites also had a measure in Stratum Y. The evaluation completed data collection for both measures for each site. Due to where each of the sites’ second measures were on the original priority list in Stratum Y, the second measure for each site received different weights despite being in the same stratum.

Table 7. Determining non-randomly selected measures

Strata	Priority	Site	Measure	Survey Disposition	Selection Type	Weight
X	1	**A**	**A1**	**Complete**	**Random**	**3/2**
X	2	**B**	**B1**	**Complete**	**Random**	**3/2**
X	3	C	C1	Live		
Y	1	D	D1	Complete	Random	8/3
Y	2	E	E1	Refused		
Y	3	**A**	**A2**	**Complete**	**Random**	**8/3**
Y	4	F	F1	Complete	Random	8/3
Y	5	G	G1	Live		
Y	6	**B**	**B2**	**Complete**	**Not Random**	**1/1**
Y	7	H	H1	Live		
Y	8	I	I1	Live		
Y	9	J	J1	Live		

The measures in Stratum X were each selected randomly. Measure A1 was first on the priority list and measure B1 was second. Because both A1 and B1 were completed and the target was 2 for the strata, site C was not called. Because site C was not called, measure C1 had a final survey disposition of “live.” In the case of Stratum X, there were 3 measures and 2 were completed. This resulted in a sample weight of 3/2 for each of the two completed measures.

In Stratum Y, four measures were completed. In this example the target for the stratum was achieved prior to calling site G. The evaluation attempted data collection for the first 4 measures on the list. Site E refused the survey or otherwise did not respond. Sites D, A, F and G completed the survey, but B did not come up in the priority list until after site G (the first “live” site in the list). In this case measure B2 was not selected randomly and needs to be treated as a special case. Measure B2

is removed from the Stratum Y weight calculation, so the three measures that were completed receive a weight of 8/3 (once measure B3 is removed there are eight measures in the frame, and 3 completed measures). Measure B2 receives a weight of 1.

B.2 Ratio Estimation

The calculation of the adjustment factors for tracking system gross savings uses appropriate case weights corresponding to the sampling rate as discussed above.

This evaluation will produce new values for the gross realization rate shown in this appendix as well as free ridership rates and net-to-gross.

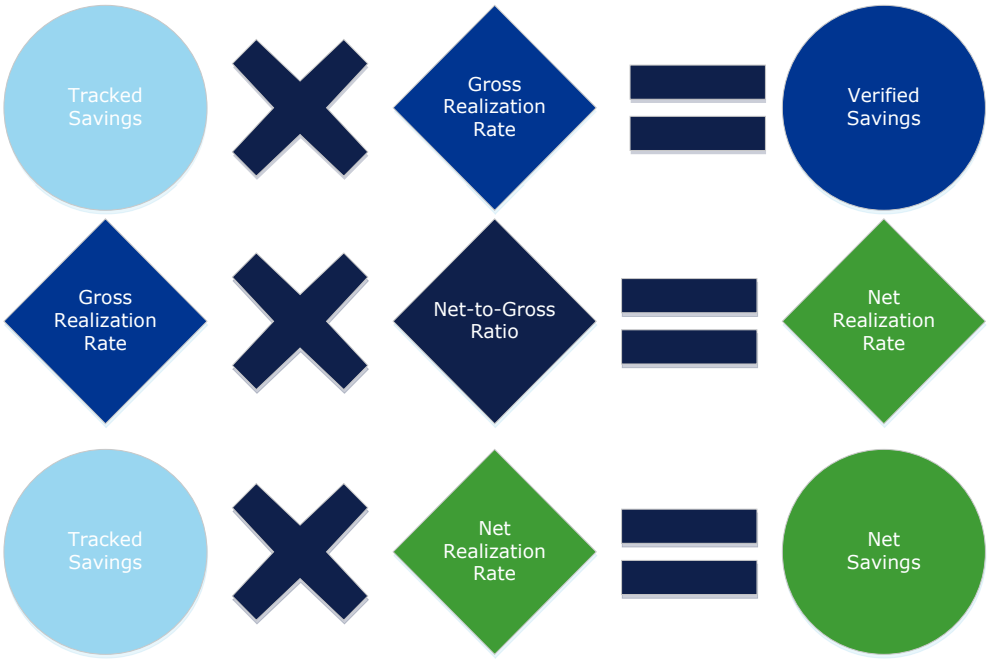
For an individual measure:

- The engineering verification factor is derived from the data collected during the participant survey data collection for TSER projects and through the on-site visits for other projects. Differences between the reported measure and the measure installed at the facility are accounted for here. The engineering adjustment factor is the ratio of the evaluator-verified savings to the program-reported savings.

The majority of the CPSV process involves determining the evaluator-verified savings estimate for each measure. The measure-level results are then combined using weights from the sample design to an overall adjustment factor.

Individual measure results are expanded to the estimate population savings (circles) using ratios (diamonds), as shown in Figure 6. Ratios are applied for each of the primary reporting domains and then summed to calculate the total for the program overall. The gross realization rate is calculated directly from the sample verified and tracked savings (as described below).

Figure 6. Ratios used to estimate verified and net savings



Two general ratio calculation approaches are employed: directly calculated and combined. The description of the process is easiest to understand through an example. The example below has three directly calculated adjustment factors: the installation rate, the engineering 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_{Tj} = tracking estimate of gross savings for measure j

G_{Ej} = engineer verified estimate of gross savings for measure j,

W_{Vj} = weighting factor for measure j used to expand the CPSV sample to the full population

V = number of measures in the CPSV sample

The gross realization rate is calculated directly:

$$R_V = \frac{\sum_{j=1}^{V} G_{Ej} w_{vj}}{\sum_{j=1}^{V} G_{Tj} w_{vj}}$$

Ratio Estimation Example

This section provides an example of the ratio estimation procedure. The results in this section are for explanatory purposes only.

The installed savings, and engineering verified savings, are calculated at the measure level and summed to the Measure Type level for each customer in the sample that completed a survey. Attribution is collected at the measure type level and is a function of the verified measure type savings for the customer. The sample weights are applied to the measure type level savings which is the unit of analysis. Table 8 shows the reported, installed and verified savings and NTG for Example Customer A's four measures reported in the program tracking database.

Table 8. Example Customer A in CPSV and NTG Sample

Measures	Measure Type	Reported m3	Installed m3	Verified m3	NTG
Space Heat Boiler 1	Space Heat	80,000	80,000	100,000	100%
Space Heat Boiler 2	Space Heat	56,000	56,000	55,000	
Process Heat	Process Heat	150,000	150,000	120,000	80%
Steam Trap Repair	Maintenance	12,000	12,000	14,000	20%

DNV engineers confirmed the customer installed all of the measures that were reported by the program; therefore, installed savings are equal to the reported savings. If a measure was initially reported as not installed, a second DNV engineer would contact the customer to verify this result. The engineering review produced adjustments to the installed savings for the first three of Customer A's reported measures, resulting in differences between the verified gross savings and installed savings for those measures.

The attribution rate is calculated for each measure type using the customer survey, and supplier survey if applicable, for Example Customer A using the methods that will be provided with the survey instruments. The measure type level attribution rates are then applied to the aggregated measure type level verified gross savings to estimate measure level net savings. Example Customer A received 100 percent attribution for the two space heat measures, 80 percent attribution for the process heat measure, and 20 percent attribution for the maintenance measure. Table 9 shows the verified gross and net savings for Example Customer A.

Table 9. Example Customer A Net Savings

Measure Type	Verified m3	NTG	Net m3
Space Heat	155,000	100%	155,000
Process Heat	120,000	80%	96,000
Maintenance	14,000	20%	2,800

Similar estimates are created for each customer in the sample. For this example, we assume Example Customers A to F comprise the Industrial Sector sample. Table 10 shows the un-weighted customer and commercial sector savings results.

Table 10. Example Industrial Sector Measure Type Level Sample

Customer	Measure Type	Reported m3	Installed m3	Verified m3	Net m3
A	Space Heat	136,000	136,000	155,000	155,000
A	Process Heat	150,000	150,000	120,000	96,000
A	Maintenance	12,000	12,000	14,000	2,800
B	Process Heat	250,000	250,000	180,000	180,000
B	Maintenance	20,000	20,000	14,000	0
C	Space Heat	150,000	150,000	140,000	35,000
D	Process Heat	80,000	80,000	81,000	81,000
E	Space Heat	70,000	70,000	70,000	0
F	Space Heat	14,000	14,000	13,000	0

Each customer in the sample frame is assigned to a sampling stratum as described in the sampling plan. Each customer in the sample is assigned a sampling weight based on the sample design and the number of completed sample points in each stratum. Assume that Example Customers A and C each have a space heat measure in a stratum that has four measures in the sample frame. The sampling weight for the space heat measures for Customers A and C is equal to the number of customers in the sample frame stratum divided by the number of stratum customers in the sample, or 4/2 = 2. The weighted savings for each customer is equal to the weight times the savings value. Table 11 shows the weights and savings (un-weighted and weighted) for each customer in the Example Industrial Sector if we assume the measure type weights shown.

Table 11. Example Industrial Sector Measure Type Level Weighted Savings

Customer	Measure Type	Weight	Reported m3		Installed m3		Verified m3		Net m3	
			unweighted	weighted	unweighted	weighted	unweighted	weighted	unweighted	weighted
A	Space Heat	2	136,000	272,000	136,000	272,000	155,000	310,000	155,000	310,000
A	Process Heat	3.5	150,000	525,000	150,000	525,000	120,000	420,000	96,000	336,000
A	Maintenance	20	12,000	240,000	12,000	240,000	14,000	280,000	2,800	56,000
B	Process Heat	1	250,000	250,000	250,000	250,000	180,000	180,000	180,000	180,000
B	Maintenance	18	20,000	360,000	20,000	360,000	14,000	252,000	0	0
C	Space Heat	2	150,000	300,000	150,000	300,000	140,000	280,000	35,000	70,000
D	Process Heat	3.5	80,000	280,000	80,000	280,000	81,000	283,500	81,000	283,500
E	Space Heat	15	70,000	1,050,000	70,000	1,050,000	70,000	1,050,000	0	0
F	Space Heat	25	14,000	350,000	14,000	350,000	13,000	325,000	0	0
		TOTALS	**882,000**	**3,627,000**	**882,000**	**3,627,000**	**787,000**	**3,380,500**	**549,800**	**1,235,500**

The next step is to determine program overall adjustment factors. For kWh, the Industrial Sector the installation rate, engineering verification factor, and attribution adjustment factor are:

3,627,000 weighted installed m3 / 3,627,000 weighted reported m3 = 100% installation rate

3,380,500 weighted verified gross m3 / 3,627,000 weighted installed m3= 93.2% eng. verification factor

1,235,500 weighted net m3 / 3,380,500 weighted verified gross m3 = 36.5% attribution adjustment.

The verified gross RR is the product of the installation rate and the engineering verification factor, or 100 percent times 93.2 percent = 93.2 percent for this example. The net RR is the product of the verified gross RR and the attribution adjustment, or 93.2 percent times 36.5 percent = 34 percent for this example.

The same principle can be applied to each Measure Type to get the Measure Type level adjustment factors. With the unit of analysis remaining the same (at the measure type level), the same process can be used to produce adjustment factors for any domain that we are able to define for the whole sample.

Applying Ratios to Domains

Ratio application refers to multiplying the gross RR and net RR times the program tracking savings to produce the total verified and net savings results for a program.

The general formula for total verified gross savings is:

The general formula for total net savings is:

The body of the report discusses how to calculate the population adjustment factors, which are based on a finite, fixed distribution of projects. You can also calculate for subsets, called domains. Viewing domain-level results allows for insights into program performance that can lead to program improvements. Domain-level ratios can also be used to apply ratios and calculate overall program savings totals. The ratio results will be generated for each of the domains of interest (subsets of the population that stakeholders agree are important) and overall, for each of the utility's programs.

The level at which one applies the ratios has an effect on the overall verified and net savings estimate for each program. There are two basic approaches that we take. The first is to apply the overall program ratio. This is appropriate to retrospective evaluation where the population that the applied ratio is the same as the population of study and is static.

The second is to apply the ratio at the domain level. This is appropriate for all uses and recommended for estimating savings for programs or program years that are not the same as the population of study. Another approach is to apply the ratio at the stratum level. This is really a subset of the domain application approach where the domain used is identical to the sample strata.

We recommend applying ratios by domains in most cases in order to improve accuracy. Assuming a sufficient sample size in each domain, domain-level precisions are usually sufficient for the approach. While 90/10 relative precision is typically the

threshold targeted for an overall result, precisions usually have a lower threshold for domain-level application as the resulting precision of the overall result will be better than the component parts.

If one domain has an extreme adjustment, the accuracy of the overall result is improved if domain level ratios are applied to the domain level savings. Table 12 shows an example where we apply the gross RR and net RR directly and by domains. The sample weighted savings in the example closely match the population savings: one domain, process heat, is 3.2 percent different, while the other domains are each within 3 percent and overall, the difference is less than 1 percent. The ratios and resulting savings are also similar, within one percent of one another. Though the results in the example are similar, the final net savings are more accurate when calculated by domains. In the example, both space heat and maintenance measures had very different attributions from process heat, and each were slightly over-represented in the weighted sample savings, which resulted in lower net savings when we applied the overall ratio directly.

Table 12. Example of Ratios Applied Overall vs. by Domains

Measure Type	A Population m3	B Sample Weighted m3	C Gross RR	D Net RR	Verified Gross Savings (A*C)	Net Savings (A*D)
Space Heat	1,950,000	1,972,000	99.6%	19.3%	1,943,078	375,761
Process Heat	1,090,000	1,055,000	83.7%	75.8%	912,810	826,024
Maintenance	585,000	600,000	88.7%	9.3%	518,700	54,600
Overall - Ratios Applied Directly	3,625,000	3,627,000	93.2%	34.1%	3,378,636	1,234,819
Overall - Ratios Applied by Domains and Summed	3,625,000		93.1%	34.7%	3,374,589	1,256,384
Difference			0.1%	-0.6%	4,047	-21,566

Neither applying the overall ratio directly nor by domains has an inherent systemic bias, but when the differences among the domain ratios are significant, applying by domains results in improved accuracy.

The choice between how to apply the ratios does not affect whether or which domains are reported. There is a large inherent value in looking at program results by multiple domains in order to better understand where the program is doing well and what areas have room for improvement.

Criteria for selecting domains for reporting and application

DNV will select the domains that are reported and those that will be applied to estimate gross savings for the programs.

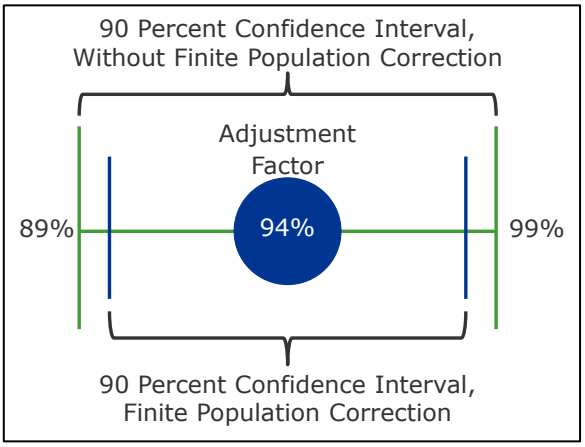
Table 13. Relevant statistics

Term	Definition
Ratio/Adjustment factor	A point estimate of the evaluation findings expressed as a percent.
+/- or Absolute Precision	If the evaluation were repeated several times selecting samples from the same population, 90% of the time the ratio would be within this range of the ratio
Confidence interval	The upper bound is defined by the ratio plus the absolute precision. the lower bound is defined by the ratio minus the absolute precision.
Relative Precision	The relative precision is calculated as the absolute precision divided by the ratio itself. By convention, relative precisions are the statistic that are targeted in sampling (i.e., 90/10 is a relative precision metric)
Finite population correction (FPC)	FPC is a factor that reduces the measured error of samples drawn from small populations (less than 300). FPC applies when the ratio is applied to the same population from which the sample was drawn.

Figure 7 shows an example:

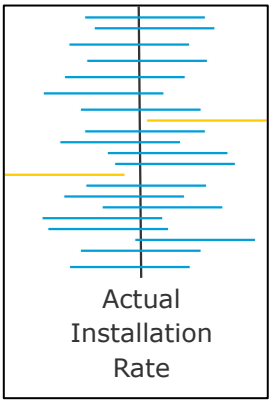
- the adjustment factor (ratio) as a blue point
- the 90 percent confidence interval with finite population correction (blue)
- the 90 percent confidence interval without finite population correction (green)

Figure 7. Ratio Diagram Example



The plus/minus (±) error (%) indicated at the 90 percent confidence interval is the absolute difference between the estimated percentage and the upper or lower confidence bound. For example, in Figure 7, the ratio is 94 percent and the non-FPC 90 percent confidence interval is ± 5 percentage points (i.e., 94 percent ± 5 percent). Another way of saying this is that there is a 90 percent probability that the actual ratio for the next year's program lies between 89 and 99 percent. Figure 8 demonstrates this concept by showing twenty hypothetical confidence intervals calculated from twenty different samples of the same population. Eighteen out of twenty (90 percent) include the true population ratio.

Figure 8. Ninety Percent Confidence Interval



Note: Each horizontal line represents a confidence interval. Yellow confidence intervals do not include the actual ratio.

The relative precision of the ratio is calculated as 5%/94% =5.3%.

For low ratios, relative precisions may be quite high, even when the confidence interval around the ratio is quite narrow. Consider a ratio of 40% with the same 5% absolute precision as in the above example. While the absolute precisions are the same, the latter ratio (40%) has a relative precision of 5%/40% =12.5%.

Because relative precisions can over-represent error for low ratios (and under-represent errors for ratios above 100%), we prefer to set thresholds for reporting and application based on the absolute precision rather than the relative precision. Where prospective application (applying the results of a study to a different program year than the one studied) is used, FPC-off errors are appropriate and the thresholds for reporting and application may be relaxed somewhat depending on context and needs.

For determining which ratios to report and apply we will use the following rules:

- The minimum sample size for a reporting or application domain will be five.
- The absolute precision threshold for reporting ratio for a domain will be +/- 20% at 90% confidence with FPC-on.
- The absolute precision threshold for applying ratio for a domain will be +/- 15% at 90% confidence with FPC-on for retrospective application.
- The absolute precision threshold for applying ratio for a domain will be +/- 20% at 90% confidence with FPC-off for prospective application.

Reporting domains will be defined as combinations of categorizations where sample sizes and precisions allow:

- Stratification segments
- Measure types

APPENDIX C. CPSV RIGOR LEVELS

DNV will use the value of information framework to efficiently apply more evaluation resources (such as labor hours) to the areas with the greatest uncertainty (such as large and complex measures) and fewer resources to the areas with the least uncertainty (such as small simple measures) by defining varying evaluation rigour levels and applying them to each measure. To ensure that the appropriate rigour is communicated to everyone who reviews them, site plans and site reports will use colour-coded table headers according to the assigned rigour level for that measure. Table 14 shows the general descriptions of the evaluation rigour levels and their assigned colours.

Table 14. Rigour level descriptions

Rigour Level	Description	Assigned Colour
Standard	Includes: ▪ Detailed application review ▪ On-site verification and/or telephone interview ▪ Collection of data on key parameters ▪ Revised engineering calculations ▪ Billing data analysis ▪ Possible spot measurements	
High	Includes all approaches described in Standard, plus as applicable: ▪ On-site verification (all) ▪ Billing/interval data analysis ▪ Calibrated standard simulation models ▪ Possible short term post monitoring	
Very High	Includes all approaches described in High, plus as applicable: ▪ Complex calibrated simulation models ▪ Spot measurements ▪ Long-term post monitoring ▪ Supplemental research ▪ Multiple site visits	

Higher rigour sites could involve the addition of elements such as:

- A fully specified regression analysis of consumption information from utility bills with inclusion/adjustment for changes and background variables over the time period of the analysis that could potentially be correlated with the gross energy savings being measured.
- Twelve (12) months post-retrofit consumption data are required.
- Twelve (12) months pre-retrofit consumption data are required, unless program design does not allow pre-retrofit billing data, such as in new construction. In these cases, well-matched control groups and post-retrofit consumption analysis is allowable.
- Sampling must be adequate (in general, a minimum of six data points will be required) for a valid regression-based estimate.
- Building energy simulation models that are calibrated as described in IPMVP Option D requirements. If appropriate, evaluators may alternatively use an engineering model with calibration.
- Retrofit isolation engineering models as described in IPMVP Option B requirements

ONTARIO GAS DSM EVALUATION CONTRACTOR

2023 Natural Gas Demand Side Management - CPSV Sample Design

Ontario Energy Board

Date: April 26, 2024

Table of contents

List of figures

List of tables

1 SAMPLE DESIGN

This section presents the stratification plan using the data provided by Enbridge for 2023 custom projects in the Commercial, Industrial, Low Income, and Large Volume programs.

1.1 CPSV sample design

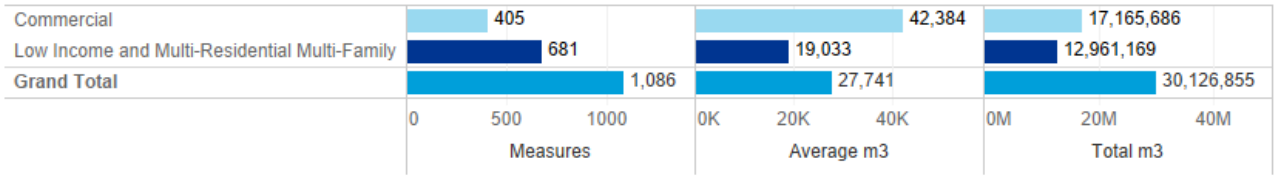
1.1.1 Explore the 2023 tracking data

We describe a row in the tracking data as a “measure”. Enbridge’s tracking data has a clear project identifier that groups rows into projects. For our analysis and the sample design, we use the “measure” row as our unit of analysis.

1.1.1.1 Commercial and low income

For CPSV, the commercial multi-residential multi-family and low income multi-family segments are combined into a single segment. The commercial segment makes up more than half of the custom savings of the combined 2023 Commercial and Low Income programs and less than half of the measures. Figure 9 provides an overview of the number of measures, average measure size in m^3 and total m^3 for both the commercial and multi-family segments.

Figure 9. 2023 commercial and low income programs summary

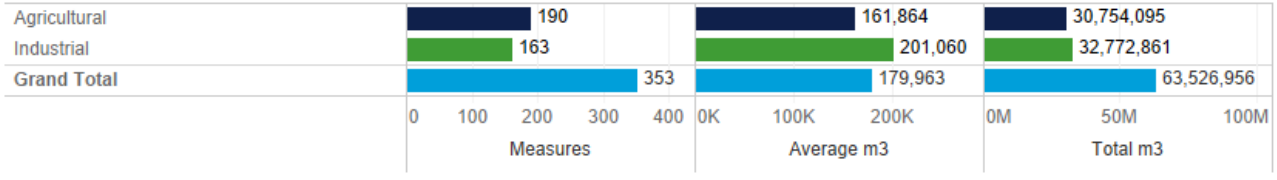


	Measures	Average m3	Total m3	Percent of Program m3
Commercial	405	42,384	17,165,686	56.98%
Low Income and Multi-Residential Multi-Family	681	19,033	12,961,169	43.02%
Grand Total	**1,086**	**27,741**	**30,126,855**	**100.00%**

1.1.1.2 Industrial

The agricultural segment of the 2023 Industrial program makes up less than half of the savings and more than half of the measures in the program. Figure 10 provides an overview of the number of measures, average measure size in m3, and total m3 for both the agricultural and industrial segments.

Figure 10. 2023 industrial program summary

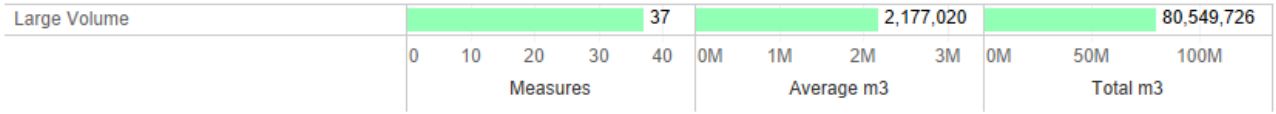


	Measures	Average m3	Total m3	Percent of Program m3
Agricultural	190	161,864	30,754,095	48.41%
Industrial	163	201,060	32,772,861	51.59%
Grand Total	**353**	**179,963**	**63,526,956**	**100.00%**

1.1.1.3 Large Volume

Figure 11 provides an overview of the number of measures, average measure size in m3, and total m3 for the 2023 Large Volume program. The number of projects in Large Volume are low enough that it is unlikely we will be able to disaggregate into reporting categories after the analysis.

Figure 11. 2023 large volume program summary



	Measures	Average m3	Total m3	Percent of Program m3
Large Volume	37	2,177,020	80,549,726	100.00%

1.1.2 Stratification and design

Table 15 shows the estimated error ratio (ER)[6] used in the sample design. The ERs used are based on an average of the 2017-2018 CPSV results and the 2017-2018 CPSV assumptions.[7] We further bounded the ER, that is we would not use an ER less than 0.25 or greater than 0.60 in order to limit the risk of over or under collecting data. Neither bounding rule was required for the 2023 sample designs.

Table 15. Estimated error ratio used in sample designs

Programs	Segment	2017-2018 Program Segment	Error Ratios		
			2017-2018		2023
			Assumed	Actual	Assumed
Combined Commercial and Low Income	Commercial	Enbridge Commercial	0.42	0.23	0.33
	Low Income and Multi-Residential Multi-Family	Enbridge Low Income and Multi-Residential Multi-Family	0.41	0.30	0.35
Industrial	Industrial	Union Industrial	0.39	0.57	0.48
		Enbridge Industrial	0.27	0.30	0.29
		Average			**0.39**
	Agricultural	Union Agricultural	0.27	0.36	0.32
Large Volume	Large Volume	Union Large Volume	0.42	0.24	0.33

The samples were designed to meet two thresholds after wave 2, shown in Table 16.

Table 16. Precision targets

Level	Relative Precision Target Final (after wave 2)
Program	90/10 FPC On
Segment	90/15 FPC On

For the 2023 gross savings verification effort, DNV used a segment-size stratification approach. The segment-size design used two levels of stratification within a program.

- **Segment (Agricultural, Industrial, Commercial, or Multifamily).** Past gross savings verification found that there were some differences in variability for the gross realization rates by segment, which is an indication that stratifying by segment should improve precision (relative to not using segment) for a given sample size. In addition, stratifying by

[6] Another term for error ratio is coefficient of variance (CV)

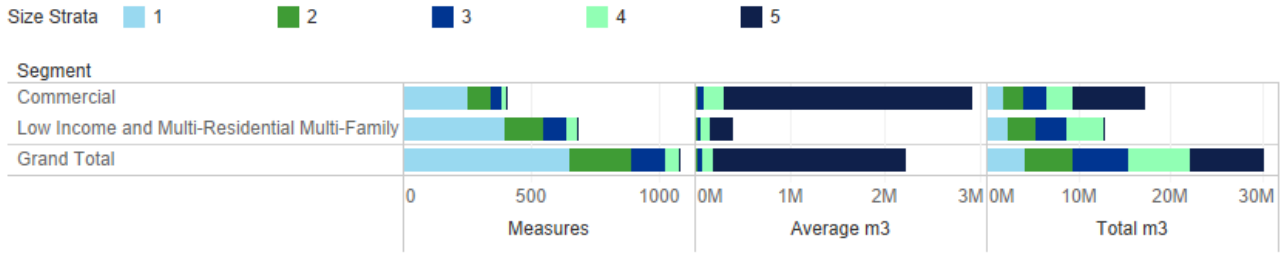
[7] The 2017-18 CPSV assumed ERs were the average of the 2016 CPSV results and 2016 assumption for complex measures (0.4) with the same bounding used in this design. We used the same averaging approach to produce the 2017-18 assumed ER for the programs overall, though theses were not used in the 2017-18 sample design or the final 2023 CPSV sample design.

segment provides value in ensuring coverage of each segment in the sample and ensures sample sizes in each segment support reporting at the segment level. Segments were clearly defined in the tracking data and the evaluation uses these definitions.

- **Measure size (m3).** Within each segment, up to seven strata were assigned. The number of size strata within the categorical groupings were limited to ensure a minimum number of target completes per strata, with the exception of the largest strata which may only have one to three sites in the population for some groupings.

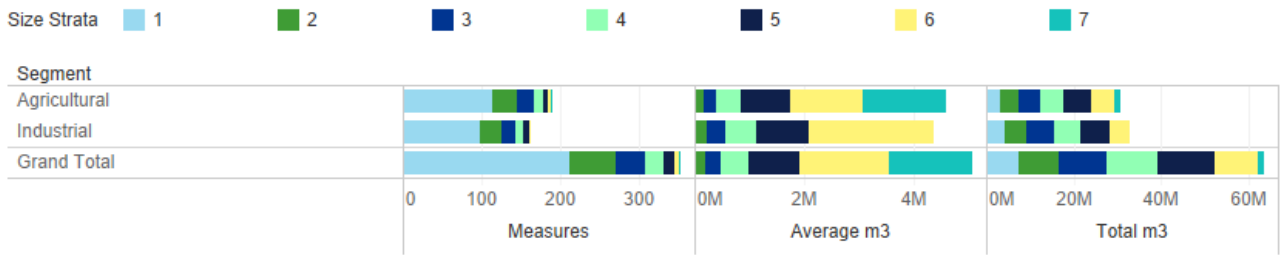
Stratification for each program is shown in Figure 12 thru Figure 14. In each design, strata with the smallest measures are to the left (sky blue) with each stratum further to the right having progressively larger measures. Size is relative within each categorical grouping: for example, the largest measures in stratum 5 in the Commercial segment may be (and in this case, are) larger than those in stratum 5 for the Low Income and Multi-Residential Multi-Family group. Each stratum within a group has similar total savings amounts, except for the largest stratum, which often contains a small number of very large projects whose total savings are greater than the other strata for the segment. At the same time, smaller strata have more measures.

Figure 12. Stratification for commercial and low income



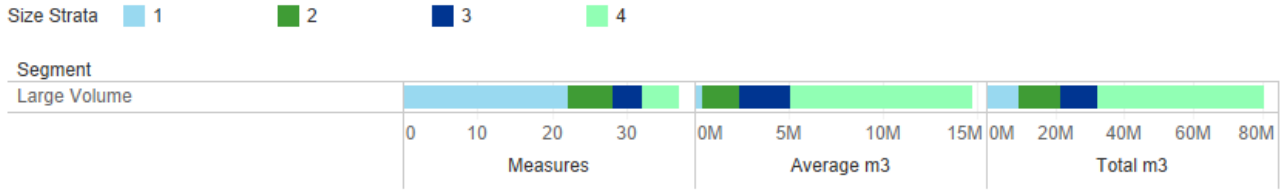
Segment	Stratum	Measures	Average m3	Total m3	Sampled Measures	Backup Measures
Commercial	1	253	6,972	1,763,968	5	2
	2	93	23,466	2,182,322	6	1
	3	43	60,861	2,617,012	7	0
	4	13	212,947	2,768,314	4	1
	5	3	2,611,357	7,834,070	2	0
Low Income and Multi-Residential Multi-Family	1	398	6,053	2,409,060	8	0
	2	152	19,598	2,978,868	7	0
	3	89	38,166	3,396,812	5	2
	4	41	95,812	3,928,293	6	1
	5	1	248,136	248,136	1	0
Grand Total		**1,086**	**27,741**	**30,126,855**	**51**	**7**

Figure 20. Stratification for industrial



Segment	Stratum	Measures	Average m3	Total m3	Sampled Measures	Backup Measures
Agricultural	1	114	28,360	3,232,999	5	2
	2	32	132,220	4,231,024	6	1
	3	20	240,274	4,805,489	7	0
	4	12	448,355	5,380,262	7	0
	5	7	898,306	6,288,144	3	0
	6	4	1,321,162	5,284,649	4	0
	7	1	1,531,528	1,531,528	1	0
Industrial	1	98	41,693	4,085,872	7	0
	2	27	192,710	5,203,176	4	3
	3	18	338,003	6,084,045	5	2
	4	11	555,433	6,109,767	4	3
	5	7	959,176	6,714,235	4	1
	6	2	2,287,883	4,575,765	2	0
Grand Total		**353**	**179,963**	**63,526,956**	**59**	**12**

Figure 14. Stratification for large volume



Segment	Stratum	Measures	Average m3	Total m3	Sampled Measures	Backup Measures
Large Volume	1	22	431,769	9,498,912	3	2
	2	6	1,969,805	11,818,827	3	2
	3	4	2,698,815	10,795,261	3	1
	4	5	9,687,345	48,436,726	5	0
Grand Total		**37**	**2,177,020**	**80,549,726**	**14**	**5**

1.1.3 Selecting a sample design

Table 17. Sample size and anticipated precision by program

Program	Number of Measures		Anticipated Relative Precision @ 90% Confidence	
	Sample Size (n)	Sample Frame (N)	FPC On	FPC Off
Large Volume	14	37	8%	10%
Combined Commercial and Low Income	35	1,086	9%	9%
Industrial	44	353	9%	10%
Total	**93**	**1,476**		

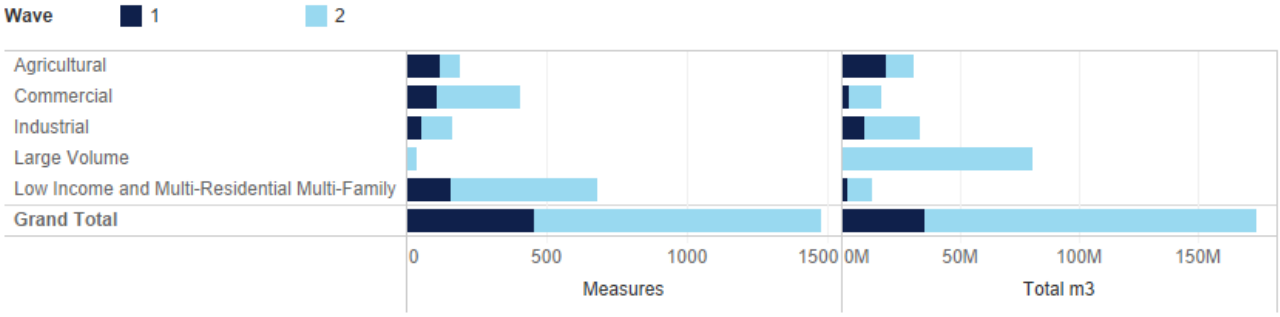
Table 18. Sample size and anticipated precision by program and segment

Program	Segment	Number of Measures		Anticipated Relative Precision @ 90% Confidence	
		Sample Size (n)	Sample Frame (N)	FPC On	FPC Off
Large Volume		**14**	**37**	**8%**	**10%**
Combined Commercial and Low Income	Commercial	17	405	12%	12%
	Low Income and Multi-Residential Multi-Family	18	681	15%	15%
Combined Commercial and Low Income Total		**35**	**1,086**	**9%**	**9%**
Industrial	Industrial	21	163	12%	13%
	Agricultural	23	190	14%	15%
Industrial Total		**44**	**353**	**9%**	**10%**
Total		**93**	**1,476**		

1.1.4 Integration of data collection wave 1 and wave 2 Samples

Data collection for the 2023 CPSV was conducted in two waves. The wave 1 sample was selected from projects in the first three quarters of 2023, while the second wave was selected from projects in the fourth quarter. Figure 22 shows the number of measures and measure savings for the sample frames broken out by wave.

Figure 22. Sample Frame Totals by wave



		Measures	Total m3	Percent of Program m3
Agricultural	1	123	18,918,772	10.86%
	2	67	11,835,323	6.79%
	Total	190	30,754,095	17.65%
Commercial	1	111	3,150,574	1.81%
	2	294	14,015,112	8.05%
	Total	405	17,165,686	9.85%
Industrial	1	56	9,959,289	5.72%
	2	107	22,813,572	13.10%
	Total	163	32,772,861	18.81%
Large Volume	1	3	84,615	0.05%
	2	34	80,465,111	46.19%
	Total	37	80,549,726	46.24%
Low Income and Multi-Residential Multi-Family	1	161	2,640,446	1.52%
	2	520	10,320,723	5.92%
	Total	681	12,961,169	7.44%
Grand Total		**1,476**	**174,203,536**	**100.00%**

Table 19 shows the sample and frame counts for each wave. Measures in the wave 1 sample frame were not eligible for selection in the wave 2 sample. To reduce customer burden, DNV also made wave 2 measures installed by sites who were recruited for or refused data collection in the wave 1 CPSV or FR samples ineligible for selection. This latter decision affected less than 1% of wave 2 sample frame measures. Sample measures in the table for wave 1 reflect the actual completed measures to date.

Table 19. Sample size by program, segment and wave

Program	Segment	Number of Measures Wave 1		Number of Measures Wave 2	
		Sample Size (n)	Frame (N)	Sample Size (n)	Frame (N)
Large Volume		**0**	**3**	**14**	**34**
Combined Commercial and Low Income	Commercial	8	111	16	294
	Low Income and Multi-Residential Multi-Family	9	161	18	520
Combined Commercial and Low Income Total		**17**	**272**	**34**	**814**
Industrial	Industrial	8	56	18	107
	Agricultural	19	123	14	67
Industrial Total		**27**	**179**	**32**	**174**
Total		**44**	**454**	**80**	**1,022**

About DNV

DNV is a global quality assurance and risk management company. Driven by our purpose of safeguarding life, property and the environment, we enable our customers to advance the safety and sustainability of their business. We provide classification, technical assurance, software and independent expert advisory services to the maritime, oil & gas, power and renewables industries. We also provide certification, supply chain and data management services to customers across a wide range of industries. Operating in more than 100 countries, our experts are dedicated to helping customers make the world safer, smarter and greener.

1 SITE REPORT – XX000

Table 1-1. Site Overview

Utility Program	<program>
CPSV ID	
Evaluated (Total) Measures	
Building Type (Verification)	
Data Collection Type	
Data Collection Date	
High Level Description of Project(s) (Verification Description)	.

Table 1-2. Measure Overview(s)

Utility Project ID	<utility measure #>	<utility measure #>
Measure Number	1	2
Rigour Level (Verification)	High Rigour (green headers)	Standard Rigour (blue headers)
Measure Description (Tracking)		
Measure Description (Verification if diff.)		
Program Year		
Installation Date (Tracking)		
Stratum (Verification)		
First Year Savings (Tracking)		
First Year Realization Rate (Verification)		
Key Reasons for Adjustment (Verification)		

Potential Measure Interactions

In 2023, this site had (x) measures, (y) of which were sampled.

1. **ABC-123, Boiler replacement** – (Interactive/Noninteractive) - installed prior (to/after) and on (same/different) system to sampled measure ABC. [If interactive] Ex ante took into account correctly, so no change / Ex ante and ex post differed. Ex post savings reduced by (X) due to the change.

1.1 Verification of Measure 1

1.1.1 Utility Description of Measure

The text below is taken verbatim from the utility documentation except as indicated by brackets [].

Utility Project Description

Area A of the plant is XXXXX

Utility Baseline Description

[Customer] replaced XXX. The plant operates ## shifts per day. Measure is XXXX.

Utility Energy Efficiency Measure Description

Measure is XXXX.

1.1.2 Verifier Interpretation and Additional Information

The following text outlines our understanding of the project prior to data collection.

Verifier Project Description

This is our understanding of the measure.

This is how it saves energy.

Verifier Baseline Description

In the baseline case, XXXXX.

Verifier Energy Efficiency Measure Description

In the efficient case, XXXXXX.

After data collection,

1.1.3 Site Plan Summary

The key sources of uncertainty and how the verification addressed them are provided in Table 1-3.

Table 1-3. Site Plan Summary

<utility measure #>	Primary Data Collection Approach	Backup Data Collection Approach

Top Priority red bold. **Second priority black bold.**

1.1.4 Site Findings

Table 1-4 provides a summary of the findings for parameters in the Site Plan Summary.

Table 1-4. Findings – Measure 1

<utility measure #>	Ex Ante Source	Ex Ante Value	Ex Post Value	Ex Post Source

Items that changed are in red.

1.1.5 Calculation Method

The ex-ante calculation method is based on (*high level method 1 to 2 sentences)*.

Ex post utilized *(state clearly if ex post used ex ante and why or why not. If different method was used, why and what was done instead. METHOD CHANGE ONLY not input or assumption changes)*

1.1.6 Results

Table 1-5 presents the results for the measure. The results below are based on the preceding findings.

Table 1-5. Results – Measure 1

<utility measure #>	Ex Ante Value	Ex Post Finding	% Match	Source or Reason(s) for Difference
Measure Type				
Standard EUL of Measure (Years)				
ER Period (Years)				
Non-ER Period (Years)				
Baseline Type during ER Period				
Baseline Type during Non-ER Period				
Annual m3 Savings in ER Period				
Annual m3 Savings in Non-ER Period				
First Year m3 Savings				
Cumulative m3 Savings				
Measure Incremental Cost				
Annual kWh				
Cumulative kWh				
Annual Water (L)				
Cumulative Water (L)				

1.1.7 Key Findings

1.1.8 Recommendations

1. XXXXX
2. XXXXX
3. XXXXX

About DNV

DNV is an independent assurance and risk management provider, operating in more than 100 countries, with the purpose of safeguarding life, property, and the environment. Whether assessing a new ship design, qualifying technology for a floating wind farm, analyzing sensor data from a gas pipeline, or certifying a food company's supply chain, DNV enables its customers and their stakeholders to manage technological and regulatory complexity with confidence. As a trusted voice for many of the world's most successful organizations, we use our broad experience and deep expertise to advance safety and sustainable performance, set industry standards, and inspire and invent solutions.