# AMPCO INTERROGATORY 1

1.0 Revenue Requirement, Operating Costs and Capital Spending

Issue 1.1

Is the IESO's Fiscal Year 2017 net revenue requirement of $190.8 million appropriate?

## INTERROGATORY

Reference: A2-2-2 Page 21

Preamble: The IESO's Corporate Performance Measure #1 is to execute its strategic plan.

a) Please indicate when the IESO expects to do its annual review of its strategy and the process involved.

b) Please indicate when the IESO expects to update its strategy.

## RESPONSE

a) The IESO's Board of Directors and the Executive Leadership Team conducted an annual review of the strategic plan on March 2, 2017. In reviewing the organization's strategy, the IESO Board of Directors and the Executive Leadership Team assessed the broader environment in which the IESO operates, including stakeholder priorities, to determine whether updates to the strategy are necessary.

b) With a new CEO on board, the IESO Executive Leadership Team intends to review the corporate strategy with any updates expected later in 2017 or early 2018.

Page Intentionally Blank

# AMPCO INTERROGATORY 2

1.0 Revenue Requirement, Operating Costs and Capital Spending

Issue 1.1

Is the IESO's Fiscal Year 2017 net revenue requirement of $190.8 million appropriate?

## INTERROGATORY

Reference: A2-2-2 Page 21

Preamble: The IESO's Corporate Performance Measure #4 is "Operations are well co-ordinated with LDC partners."

a) Please summarize the objective and outcome of this measure and provide the status of the IESO's 2017 target that one major LDC is actively engaged in two-way communication by the end of 2017.

## RESPONSE

a) Please refer to the response to BOMA Interrogatory 1 part f at Exhibit I, Tab 1.0, Schedule 2.01.

Page Intentionally Blank

# AMPCO INTERROGATORY 3

1.0 Revenue Requirement, Operating Costs and Capital Spending

Issue 1.1

Is the IESO's Fiscal Year 2017 net revenue requirement of $190.8 million appropriate?

## INTERROGATORY

Ref: A2-2-2 Page 21

Preamble: Two of the IESO's 2017 target for its Corporate Performance Measure #10: are 80% of priority change initiatives progressing to their approved business case; and 100% of the Operations Readiness Initiatives progressing according to their approved business case.

a) Please provide the business case for the priority change initiative and Operations Readiness Initiative.

b) Please provide the status of the above two initiatives with respect to achievement of the 2017 targets.

## RESPONSE

a) Please refer to the responses to BOMA Interrogatory 8 at Exhibit I, Tab 1.1, Schedule 2.08 and CME Interrogatory 3 at Exhibit I, Tab 1.5, Schedule 3.03.

b) As of Q2-2017, 80% of priority change initiatives and 100% of the Operations Readiness Initiatives are progressing as per the plan.

Page Intentionally Blank

# AMPCO INTERROGATORY 4

1.0 Revenue Requirement, Operating Costs and Capital Spending

Issue 1.1

Is the IESO's Fiscal Year 2017 net revenue requirement of $190.8 million appropriate?

## INTERROGATORY

Reference: B-1-1 Page 11

a) When will the IESO begin supporting the OCCSDC?

b) Please discuss if the IESO's support of the OCCSDC by providing staff to perform work on behalf of the OCCSDC puts the IESO at risk of resource constraints. If not, why not.

## RESPONSE

a) The IESO began support of the OCCSDC in April 2017, when the inter-corporate service agreement was signed between the IESO and the Ministry of Environment and Climate Change.

b) The IESO's support of the OCCSDC will not put the IESO at risk of resource constraints. The IESO manages its resource requirements at an enterprise level, which includes consideration of resources for this initiative. In addition, any required incremental resources will be secured through, and paid for by funding from, the OCCSDC.

Page Intentionally Blank

# AMPCO INTERROGATORY 5

1.0 Revenue Requirement, Operating Costs and Capital Spending

Issue 1.1

Is the IESO's Fiscal Year 2017 net revenue requirement of $190.8 million appropriate?

## INTERROGATORY

Reference: Ex A-2-2 Page 19

Preamble: The evidence indicates the IESO's Internal Audit Group provides independent, objective insight and assurance on governance, risk management, and controls to management and the Board of Directors.

a) Please provide a list of the Internal Audits undertaken in 2015, 2016 and 2017 (actual and planned).

b) Please provide a copy of all 2016 and 2017 audit reports.

## RESPONSE

a) Please refer to the response to SEC Interrogatory 5 at Exhibit 1, Tab 1.1, Schedule 7.05.

b) Please refer to the response to SEC Interrogatory 5 at Exhibit 1, Tab 1.1, Schedule 7.05.

Page Intentionally Blank

# BOMA INTERROGATORY 4

Issues 1.1, 1.3, 5.1

## INTERROGATORY

Reference: Conservation First Framework

(a) The credibility of CDM programs depends, in good part, on the ability to measure the results of such programs. Please provide the steps the IESO is taking beyond what is addressed in evidence to ensure that energy and demand savings are measured to the greatest extent possible. Please confirm that the midterm review is addressing that issue.

(b) Where, in the IESO's view, does the accountability for achieving the Conservation First Framework, 2015-2022 reside? Is it with the IESO, the LDCs, or shared responsibility between the IESO and the seventy-six LDCs, and if shared, how is the accountability for program results determined?

(i) Please confirm that the IESO manages directly at least one aspect of the Conservation First Framework, the industrial accelerator program.

(c) How does the IESO plan to steer the implementation of the Conservation First to ensure that its interim and final targets are met on time? What steps will the IESO take to make its fullest possible contribution to the realization of the program? Please discuss fully.

(d) What is the total FTE, full-time staff, dedicated to guiding the Conservation First Framework?

(e) What impact does the OEB's recent residential rate design change to uniform customer rates for residential customers, unrelated to either demand or consumption, have on the IESO's efforts to implement CDM in the residential sector? How would your response differ if the questions were about the OEB's proposed commercial/industrial rate changes?

## RESPONSE

(a) In order to confirm that the IESO CDM programs achieve their intended outcomes, the IESO Evaluation, Measurement & Verification (EM&V) team, using third-party contractors, conducts rigorous program evaluations for all province-wide, local and pilot programs. In general, program evaluations include market assessments, process evaluations, retrospective outcome/impact assessments (net-verified energy (kWh) and demand (kW) savings) and cost-benefit evaluations. The evaluations reports are produced in accordance with IESO EM&V Protocols and Requirements (please refer to Attachment 1) and present

energy and demand savings as well as recommendations for program improvements. All evaluations reports can be found on the IESO EM&V website[1].

The protocols provide guidance for a robust evaluation, listing guidelines and general instructions. They identify the practice required to evaluate, measure and verify energy savings and demand reductions associated with CDM activities in Ontario. In order to effectively measure the outcomes of each program/pilot, a detailed Evaluation Plan is prepared on the onset of any evaluations. Part One of the EM&V Protocols and Requirements outlines the steps and considerations taken into account when developing an evaluation plan for any program. At a minimum, energy and demand savings are determined at a province-wide level. Where data permits, the IESO conducts more regional or LDC-specific results.

The mid-term review will include evaluation results up to the end of program year 2017. This will include net verified energy and demand savings, and portfolio cost effectiveness.

(b) The accountability for achieving the Conservation First Framework objectives are as described by the March 31, 2014 direction to the OPA and all subsequent directions delivered to the IESO by the Minister of Energy. Please refer to Attachment 2 and Exhibit I, Tab 1.0 Schedule 6, Attachments 1-3.

(i) Confirmed. The IESO directly manages the Industrial Accelerator Program.

(c) The IESO utilizes a number of engagement activities to support the implementation of the Conservation First Framework (CFF). CFF is a regular topic at the IESO's Stakeholder Advisory Committees meetings. The IESO, in collaboration with the LDCs, sit on the Conservation First Implementation Committee (CFIC), including the program related working groups that report into it. The IESO also conducts annual Third Party Verification of all programs and publicly publishes the results. From an overarching viewpoint, the Environmental Commissioner of Ontario annually reports on the progress of CDM activities in the Province.

The IESO continually tracks the performance and spending of the Conservation First Framework to better understand progress to date. Existing programs continue to be refined and adjusted through the IESO-LDC Working Groups based on program performance, market research, customer/stakeholder feedback and recommendations from the annual evaluation, measurement and verification process. In addition, new local, regional and

[1] http://www.ieso.ca/en/sector-participants/conservation-delivery-and-tools/evaluation-measurement-and-verification

province-wide programs are under development based on opportunities identified by stakeholders and channel partners and the recent 2016 Achievable Potential Study.

(d) The 2017 budget FTE that is dedicated to guiding the Conservation First Framework is 58 FTEs.

(e) An assessment of the impacts of a fixed monthly charge for the distribution service portion of a residential customer's total electricity bill on IESO's implementation of the Conservation First Framework would be highly speculative at this time. The fixed monthly rate is being phased in by the OEB gradually, with a fully fixed monthly rate for distribution services for residential customers in place by 2019. The Conservation First Framework ends in 2020, meaning the full impacts of the rate design will only start to be realized in the market near the end of the framework.

The OEB's Board Policy document "A New Distribution Rate Design for Residential Customers"[2] (EB-2012-0410) dated April 2, 2015 includes a discussion on the OEB's assessment of the general potential impacts of a fixed distribution charge on conservation efforts.

The potential commercial/industrial rate design changes are currently under consultation.

[2] https://sites.hks.harvard.edu/hepg/Papers/2015/OEB_Distribution%20Rate%20Design%20Policy_20150402.pdf

Page Intentionally Blank

ONTARIO POWER AUTHORITY

# CONSERVATION FIRST
2015-2020

Verified Savings
Significance Level
EM&V Outcome
Rebound effect
Program Evaluation
Qualitative Data
Impact Evaluation
Market Effects
External Validity
Cost-Benefit
Accuracy

## Evaluation, Measurement and Verification (EM&V) Protocols and Requirements

**EM&V PROTOCOLS**
v. 2.0

# Acknowledgements

The Ontario Power Authority would like to recognize the Efficiency Valuation Organization (EVO) who developed the International Performance Measurement & Verification Protocol (IPMVP). Their work serve as a valuable reference and foundation upon which the "EM&V Protocols and Requirements V2.0" are developed.

Readers wishing more information on program evaluation methods can access the library of materials available from the US Department of Energy and the Office of Energy Efficiency and Renewable Energy at:

**www1.eere.energy.gov/analysis/pe_index.html**

# Table of Contents



ONTARIO POWER AUTHORITY

# Abbreviations

| | |
|---|---|
| **CDM** | Conservation Demand Management |
| **CMVP** | Certified Measurement and Verification Professional |
| **CVRMSE** | Coefficient of Variation of the Root Mean Squared Error |
| **DEP** | Draft Evaluation Plan |
| **ECM** | Energy Conservation Measure |
| **EM&V** | Evaluation, Measurement and Verification |
| **EUL** | Effective Useful Life |
| **FEP** | Final Evaluation Plan |
| **IESO** | Independent Electricity System Operator |
| **IPMVP** | International Performance Measurement and Verification Protocol |
| **LDC** | Local Distribution Company |
| **M&V** | Measurement and Verification |
| **MAL** | Measurement Assumptions List |
| **NEB** | Non-Energy Benefit |
| **NTG** | Net Gross Ratio |
| **NTG** | Net to Gross |
| **O&M** | Operating and Maintenance |
| **OEB** | Ontario Energy Board |
| **OPA** | Ontario Power Authority |
| **RFP** | Request for Proposal |
| **TOU** | Time of Use |
| **TRC** | Total Resource Cost |
| **VORL** | Vendor of Record List |

# Introduction

## Document Introduction

Thank you for your interest in the Conservation First 2015 – 2020 Evaluations, Measurement and Verification (EM&V) Protocols and Requirements V2.0 (the Protocols).

EM&V is critical in establishing Conservation and Demand Management (CDM) as a credible and reliable "first choice" resource in meeting future electricity supply needs of Ontario. EM&V provides information to decision-makers, system planners and program administrators for use in developing long term demand/supply plans, to maximize program performance, and to determine whether energy savings and demand reduction targets are being met.

The EM&V Protocols and Requirements V2.0 helps program and evaluation administrators create and manage objective, high quality, independent, and useful conservation program evaluations. It provides an administrative protocol; governing the "who," "how," "what," and "when" of EM&V. In addition to what has been described above, the "why" is to ensure that the Province and all market players can depend on CDM as a resource. Supporting technical guides, aimed primarily at independent Evaluation Contractors, cover off the remaining "how" elements of completing a high quality evaluation.

## Intended Audience

There are two main audiences for this document:

- **PART 1** is intended primarily for Evaluation Administrators who are charged with managing the program evaluation process
- **PART 2** is intended primarily for Evaluation Contractors, though the information is valuable to Program Administrators as well.

The document is also a resource for program design, as it is important to have a general understanding of evaluation methodologies so that programs are designed in a manner that allows for impacts to be measured and evaluated.

## Background

Across North America, increased attention is being devoted to program evaluations. Today, more than ever, increased scrutiny of government spending and rising energy prices require a prudent review of program investment. As such, linking program resource expenditures with program results has become a necessity.

In general, program evaluations include market assessments, process evaluations, retrospective outcome/impact assessments and cost-benefit evaluations. These types of evaluation studies help

managers determine what adjustments are needed in the program offer to enhance programmatic achievements relative to the committed resources.

Program evaluations are in-depth studies of program performance and customer needs. The benefits of conducting an evaluation are numerous, including:

1. Helping Evaluation Administrators and Program Managers estimate how well the program is achieving its intended objectives;
2. Helping administrators and managers improve their efforts; and
3. Quantifying results and communicating the value of program efforts amidst a multitude of regional, regulatory, and legislative priorities

## Intended Use

The EM&V Protocols and Requirements V2.0 are intended for use by CDM market players in the Province of Ontario who have an interest in CDM Program Design, Delivery and Evaluation. The protocols provide guidance for a robust evaluation, listing guidelines and general instructions. They identify the practice required to evaluate, measure and verify energy savings and demand reductions associated with CDM activities in Ontario. They are not intended for training, nor as an assurance of flawless evaluations. Still, by following these protocols, the appropriate regulatory agencies and administrative agencies can have confidence that each evaluation served is identifiable and comparable to the others using similar processes.

The different types of evaluations require data-collection and analysis methodologies with which some Evaluation Administrators will have little familiarity. It will not be necessary to have in-depth working knowledge of the many methods available. It is highly advisable to have some familiarity with basic evaluation techniques so that selecting and monitoring an Evaluation Contractor is possible, since they will recommend and implement specialized analytical methods.

While the value of program evaluation is well established, the questions of who should do what, how (rigour level and consistency) it should be done, and when (rapid versus after-the-fact feedback as well as recurring studies) are far less well defined. EM&V protocols are intended to address the following key issues:

- The need for separation between the department responsible for program delivery and the department responsible to assess program performance to realize credible and effective evaluation.
- The proper allocation of EM&V costs; typically higher for more project-based evaluations or pilots and typically lower for larger, ongoing programs.
- The proper attribution of savings, when results from multiple evaluations have to be credibly tabulated into a collective total by following common rules and processes.
- The appropriate use of ex ante input assumptions (e.g. the Measures and Assumptions Lists) during program planning, monitoring and evaluation.
- Procedures to identify and prevent duplication of evaluation efforts.
- The realization of "economies of scale" by evaluating similar initiatives and efficiency projects together, such that fewer individual and potentially inconsistent sets of results emerge at the end of a program cycle.

- How the five major streams of evaluation work may be combined or separated in various ways for efficiency and quality:
    - Outcome Evaluation (summative; ex post; conducted to verify cognitive and behavioural changes.
    - Impact Evaluation (summative; ex post; can include M&V engineering conducted for the purpose of developing new or improved ex ante savings estimates).
    - Process Assessment (develop conclusions about program performance; includes audits; can include behavioural research for the purpose of developing new or improved ex ante savings estimates).
    - Market Study (market characterization that can contribute to evaluating the impact of codes and standards, time-of-use rates, market transformation elements of efficiency programs and may also contribute to the development of ex ante savings estimates).
    - Cost Effectiveness (economic analysis that compares the benefits of an investment with the costs ).
- How to incorporate the temporal element of moving from "Resource Acquisition" to "Market Transformation", using "Capability Building".
- To ensure a consistent approach to hiring and managing Evaluation Contractors across the Province, following the lead of other major jurisdictions such as California and New York.

The entire EM&V effort is used to develop a reliable net savings estimate—those savings attributable to or resulting from program-sponsored efforts as distinguished from savings that would have occurred anyway, be that from individual behavioural choice, public acknowledgement, or from naturally occurring market adoption.

### Presentation of Information

This document takes a process-driven approach in presenting the information. The information is presented as a series of steps an Evaluation Administrator would take in managing the evaluation process, from designing evaluations, to hiring Evaluation Contractors, to reporting evaluation results. Of course in the real world, the process is not purely linear – many steps are interrelated and, to some, degree the process is iterative.

### Structure of the Document

The document is divided into two sections:

### PART 1: DEVELOPING, PROCURING AND REPORTING ON EVALUATIONS

Part 1 guides an Evaluation Administrator through the first 12 steps in the overall EM&V process: from documenting a program's market strategy, hiring an evaluation contractor and managing and publishing the evaluation results.

### PART 2: CONDUCTING AN EVALUATION

Part 2 is intended primarily for Evaluation Contractors, but it is also a useful reference for Program Administrators, providing them with a high level understanding of the technical processes required to

carry out the evaluation. Part 2 contains 11 Technical Guides.

Evaluation Administrators need a high-level understanding of the work the Evaluation Contractor is undertaking, therefore it is recommended that Evaluation Administrators also become familiar with the techniques and methods outlined in Part 2.

## EM&V Protocols and Requirements (2011-2014) vs. EM&V Protocols and Requirements V2.0 (2015-2020)

This document replaces the previous version of the EM&V Protocols and Requirements (2011-2014), with an enhanced version that provides additional guidance and clarification on how to undertake an evaluation for energy efficiency and demand management programs.

# Part 1:

# Developing, Procuring and Reporting Evaluations

Audience: Evaluation Administrators

# Introduction to Part 1

The *Conservation First EM&V Protocols and Requirements* helps Program Administrators and Evaluation Administrators create and manage objective, high quality, independent, and useful conservation program evaluations. This Protocol was developed for all staff who plan, commission, and manage program evaluation services across the province.

In the most general sense, Evaluation Administrators are persons or organizations responsible for evaluating energy efficiency, conservation, or demand response initiatives. In the EM&V context, Evaluation Administrators are those who are specifically responsible for designing and implementing the Evaluation Measurement and Verification Plan (EM&V Plan) of energy efficiency, conservation, and demand response initiatives.

Part 1 guides the Evaluation Administrator through the initial steps that lead to conducting the agreed on evaluations by an Evaluation Contractor. The Evaluation Administrator will employ industry best practices for procuring an Evaluation Contractor and working with the selected Contractor to develop and implement the EM&V Plan. Evaluation Administrators are responsible for developing an EM&V plan for a particular program or portfolio. They are also the point-of-contact for EM&V Evaluation Contractors. Evaluation Administrators are sometimes referred to as Evaluation Managers. In general terms, these steps involve the following activities:

- **Hiring an independent, qualified Evaluation Contractor** – this involves inviting qualified vendors to bid on the project and selecting an appropriate contractor from among the bidders.
- **Coordinating Evaluation Contractor's activities** – this involves working with the Evaluation Contractor to determine the detailed research methods that will be used.
- **Managing the evaluation process** – this requires a combination of skills including: balancing resources, overseeing the flow of data and information between persons involved in the evaluation, ensuring quality control with regard to the work being conducted, and ensuring project timelines are satisfied.

The Management Board of Cabinet's Procurement Directive requires that the following principles guide the procurement process:

- **Vendor Access, Transparency, and Fairness:** The procurement process should be conducted in a fair and transparent manner, providing equal treatment to all vendors. Conflicts of interest, both real and perceived, must be avoided. Particular vendors should not be relied on continuously, or routinely be granted contracts, for a particular kind of work.
- **Value for Money:** Goods and services must be procured only after consideration of the business requirements, alternatives, timing, supply strategy, and procurement method.
- **Responsible Management:** The procurement of goods and services must be responsibly and effectively managed through appropriate organizational structures, systems, policies, processes, and procedures.
- **Geographic Neutrality and Reciprocal Non-Discrimination:** Entities subject to Ontario's Trade Agreements must be geographically neutral with respect to vendor access to government business.

- **Documenting the program strategy and offer** – this requires an understanding of the program's logic model.
- **Properly scoping the program evaluation** – this involves selecting elements of the program logic model to be evaluated and drafting the research questions.
- **Identifying analytical approaches to address research questions** – this requires exploration of factors potentially influencing the program and identifying key metrics for each program element to be studied.
- **Specifying evaluation deliverables** – this involves deciding on the frequency and timing of planned evaluations, specifying the primary analytical methods the Evaluation Contractor is expected to use, and creating a detailed timeline of project deliverables.
- **Creating the Draft Evaluation Plan** – the draft evaluation plan forms the basis for the scope of work that is set out in the Request for Proposals (RFP) process which is used to hire an Evaluation Contractor.
- **Assessing the reasonableness of the Evaluation Contractor's findings and conclusions** – this involves linking conclusions to findings and providing context for findings.
- **Publishing the evaluation report** – this includes explaining how net savings figures were arrived at and normalizing them to applicable long-term trends, if appropriate.

The draft EM&V plan defines the Evaluation Contractor's scope of work. When procuring an Evaluation Contractor, the Evaluation Administrator must balance product quality, reliability, and pricing. The following factors will come into play when selecting an Evaluation Contractor:

- Selected areas of study
- Choice of analytical methods
- Availability of staffing
- Timing of evaluation tasks
- Data collection and analysis requirements
- Competitiveness of the offer

Evaluation Administrators and Program Administrators should expect the Evaluation Contractor to propose a variety of approaches for carrying out the work. Given the nature of research, an EM&V plan developed by an Evaluation Administrator is always a draft, with specific research activities developed after work begins and uncertainties managed to achieve the desired levels of precision and accuracy based on the facts revealed.

Though the Evaluation Administrator's job does not end with the hiring of a qualified, independent Evaluation Contractor and drafting an evaluation plan, we have chosen to end Part 1 because the technical aspects of the evaluation, as agreed upon in the evaluation plan, are conducted by the Evaluation Contractor, and Part 2 focuses on those technical elements.

# Step 1: Document Market Strategy and Program Offer

## Key Points / Highlights

**Documenting Market Strategy and a Program's Offer involves the following tasks:**

**1a.** Specify Market Needs

**1b.** Identify Program Strategy

**1c.** Tabulate Impact Forecasts

**1d.** Highlight Program Benefit-Cost Ratios

## Task 1a: Specify Market Needs

To plan a program's evaluation one needs a good understanding of the program. As such, the program description should include discussion of relevant market conditions and the needs of targeted stakeholders.

Given that the purpose of a program is to cause change in the market, the program description should point out key market hurdles and barriers. The descriptions should include a table that identifies and distinguishes between:

**Market Hurdles** – these are temporary obstacles that discourage the adoption of desired behaviours. A well-designed program can, in the short term at least, directly influence market hurdles such that changes to behaviour can occur. For consumers in the business sector, an example of a market hurdle is the payback period or return-on-investment thresholds for investing in energy-efficient equipment. For individual consumers, a market hurdle could be the price of energy efficient appliances. With such hurdles, a financial incentive could help the consumer overcome this one-time investment hurdle.

**Market Barriers** – these are on-going obstacles that prevent adoption of desired behaviours. A well-designed program can also directly influence market barriers, but it typically takes longer for change to occur with market barriers than with market hurdles. For schools, for example, a market barrier might be a lack of trained maintenance staff. If that's the case, a useful program design strategy might be to offer technical training for maintenance staff on energy savings strategies and practices.

The Evaluation Administrator and Program Administrator are both responsible for properly classifying targeted market opportunities as either market hurdles or market barriers.

A program's design reflects an underlying theory about how and why the program activities will achieve the desired results. In particular, the underlying theory illustrates how program activities will help participants overcome one or more market barriers or hurdles, thereby leading to the adoption of energy efficiency or conservation measures.

## Task 1b: Identify Program Strategy

Traditionally, programs were classified as having an underlying strategy that is either:

**Resource acquisition** – these programs address market hurdles and are characterized as involving the direct purchase of GWh or MW. (Programs based on this strategy are referred to as Resource Acquisition Programs). Or,

**Market transformation** – these programs address market barriers and are characterized as involving activities where GWh or MW savings are the logical extension of market-based outcomes. (Programs based on this strategy are referred to as Market Transformation Programs)

The Evaluation Administrator must identify whether the program strategy is resource acquisition or market transformation in nature.

Program strategies have evolved and now some programs are hybrids, meaning they include incentives aimed at overcoming market hurdles and producing short-term energy savings directly and they overcome market barriers, leaving market conditions that are favourable for continued realization of program impacts. Where a program involves a hybrid strategy, the Evaluation Administrator must identify which activities are associated with market transformation and which are intended for resource acquisition.

Regardless of the program type, Program Administrators should forecast the demand impact from the program. This information is required in order to address system reliability. Although the system peak demand savings of all programs offered will be assessed, outcome evaluations may also examine the other benefits. To ensure demand savings can be calculated using a variety of demand definitions, hourly load impacts should be produced to allow for flexibility. More details about calculating demand savings in **Technical Guide 6: Demand Savings Calculation Guidelines.**

**Supportive and technical guidelines on the following are included in Part 2 of this document:**

- **Technical Guide 5:** Gross Energy Savings Guidelines
- **Technical Guide 6:** Demand Savings Calculation Guidelines
- **Technical Guide 7:** Market Effects Guidelines

## Task 1c: Summarize Budget Allocation

The program description should include a summary of the spending on program activities. In short, Program Evaluations focus on the largest program expenditures or on where the largest program impact is forecasted. A simple table showing the budget allocation per class of activity is necessary to address the Program Manager's level of commitment to the program strategies chosen.

**Tabulate Impact Forecasts**

For example, lighting programs make broad assumptions about existing measures: their age, use, and condition. These assumptions change the forecasted energy profile of the lighting measures being removed. Similar assumptions alter the forecasted energy profiles of replacement equipment. The mean difference between these two forecasted energy profiles represents the forecasted energy and demand savings expected for the program.

## Task 1d: Highlight Program Benefit-Cost Ratios

Program cost-effectiveness has broad implications for program planning, design, and implementation. The Program Administrator should develop reasonable forecasts of program costs and savings, making sure that cost-effectiveness screenings fairly represent the anticipated ratio of program costs to benefits. Where verified benefit streams or real costs differ significantly from those forecasted, the Evaluation Contractor should note critical variances and offer conclusions about their impact on program theory. As well, the Evaluation Contractor should make recommendations regarding ways of resolving large differences. Moving forward, the Program Manager will be expected to use this information to narrowing these variances.

When preparing program cost-effectiveness one should consult the cost-effectiveness policy and procedures explained in **Technical Guide 2: Cost-Effectiveness Guidelines**

To help optimize implementation effectiveness, cost-effectiveness studies may be done with regard to specific program activities or with regard to particular measures. These studies can be valuable at the early stages of a program offer, or after program processes have been significantly altered.

### Summary of Actions

- Classify targeted market opportunities as either market hurdles or market barriers
- Identify whether program strategy is resource acquisition or market transformation in nature
- If hybrid strategy involved, identify activities associated with market transformation and resource acquisition
- Include summary of spending on program activities
- Indicate anticipated level of demand and energy savings expected
- Report cost of conserved energy (and cost of demand savings, if demand savings program)

# Step 2: Anticipate Program Causes and Effects

## Key Points / Highlights

**Anticipating Program Causes and Effects involves the following tasks:**

- **2a.** Summarize Resources Available for the Program
- **2b.** Categorize Planned Program Activities
- **2c.** Specify Expected Return on Program Investments
- **2d.** Highlight Potential Outcomes Resulting from the Program Offer
- **2e.** Specify the Desired Impacts from the Program Offer
- **2f.** Illustrate and Annotate Program Logic
- **2g.** Verify Savings Attribution Pathway

## Task 2a: Summarize Resources Available for the Program

Programs allocate resources in an effort to cause energy and demand savings. In the explanation of the theory on which a program is based, program administrators must specify the resources available to achieve the desired effects: namely, the monies and time allocated to the program.

**Capital** is money allocated to fund specific program activities and administrative expenses, including money allocated directly to program service provisioning.

While capital covers the majority of program funding, other contributions, such as in-kind contributions for infrastructure and staff, may add significantly to the program offer without affecting the budget allocation. Where in-kind contributions are relevant to achieving energy and demand savings one should identify them as key resources available to the program.

**Non-capital funding** – Examples of important non-capital sources of program funding:

**Infrastructure (in-kind)** – business and information systems a sponsoring organization may provide to operate the program, such as the organization's procurement services center, its billing information systems, or its training facilities.

**Human (in-kind)** – expertise and support staff offered by an organization to help deliver a program without a direct budget allocation, such as utility account representatives, training center staff, marketing professionals, etc.

**Strategic Relationships** – ties or relationships sponsoring organizations may have with vendors that may provide time or expertise without significant added cost.

## Task 2b: Categorize Planned Program Activities

Program activities are generally categorized based on the nature of their intervention into the marketplace. Here are the main categories into which program activities usually fall:

**Financial Assistance** – this is the payment of cash to encourage customers to engage in desired behaviour. Financial assistance may include direct financial incentives, rebates, or in-store discounts. Other financial assistance in the form of financing, guarantees, or price buy-downs may also be used.

**Technical Assistance** – these are when services are offered to buyers of energy efficiency measures or channel partners. This assistance may be consulting services, training courses, or access to help lines. The goal of technical assistance is to facilitate the introduction, installation, or maintenance of energy efficient technologies within the market.

**Informational and Educational Materials** – this is basically materials focused on communicating technical information, or information about technology options, end-use applications, or emergent practices. The materials can be bill inserts, information brochures, client testimonials, booklets, CDs/DVDs, radio spots, exhibition booths, websites, smartphone apps, etc. The form of media is less important than the message included: namely, technical information rather than promotional material.

Promotional Materials –materials aimed at encouraging program uptake using media to highlight a program's presence within a market. Often promotional materials are not considered part of the planned offer program. However, Evaluation Administrators and Program Managers should insist on including them in order to show how promotional activities contribute to changing market attitudes that then lead to changes in behavior and energy demand.

## Task 2c: Specify Expected Return on Program Investments

Monies paid for goods and services that result in Program Outputs are program expenditures. Program Outputs are the most direct returns that can be measured from program expenditures. Program Managers must highlight on a Program Logic Model the Program Outputs that:

- lead to outcomes along the "Critical Savings Attribution Pathway" and
- involve the expenditure of a significant amount of program resources that have been expended (regardless of their contribution to energy and demand savings). (See **Task 2f: Illustrate and Annotate Program Logic** for examples)

Where possible, Program Managers should specify an average cost per unit of Program Output.

**Program Outputs** are basically the tangible results achieved by a program. Program Outputs are monitored based on metrics the Program Manager establishes, such as: participants served, the number of end-use measures installed, the number of workshops held, pass/fail rates from training programs, etc. It should be noted that though program outputs are critical to a program's success, they are only intermediaries that demonstrate resource allocation and contract compliance.

A **Program Logic Model** is a diagram showing a causal chain with links that go from resource expenditure to long-term outcomes for a program.

### Task 2d: Highlight Potential Outcomes Resulting from the Program Offer

Program Outputs should lead to some anticipated market change. These changes are themselves outcomes that Program Managers must include in the Program Logic Model. The cognitive, structural, and behavioural outcomes necessary to achieve demand and energy savings must be distinguished in the Program Logic Model, along with other market changes that bring about the desired program impacts.

**Types of Outcomes**

**Cognitive Outcomes:** Changes in attitude of people and organizations as a result of a program. Such changes can be reflected in learning, knowledge or understanding, perception, outlook, ambition, desire, etc. They are changes in mental abilities or perceptions that influence people and cause them to change their behaviour in a desired way.

**Structural Outcomes:** Changes in the target market's ability to observe and/or adopt behavioural outcomes as a result of a program. These changes can be reflected in things like enhancement of skills, technological innovation, changes in market structure, increased fiscal support and other market-based changes that support the short-term, intermediate, and long-term abilities of market actors.

**Behavioural Outcomes:** Changes in behaviour as a result of structural or cognitive outcomes achieved by a program. These changes can be reflected in purchasing decisions, stocking practices, technology utilization, energy consumption, load shifting, etc. When assessing whether there have been behavioural changes as a result of a program one must be sure to filter out changes that might have occurred as a result of external influences.

### Task 2e: Specify the Desired Impacts from the Program Offer

For funded conservations programs the desired impact is usually demand and energy savings. However, governmental and sustainability initiatives may be part of a particular program, in which case societal impacts may come into play, such as job creation, emission credits, and so on. The Evaluation Administrator must document the program demand and energy impacts, specifying the hours of demand reduction and the annualized energy savings.

The Evaluation Administrator may also include other societal impacts (job creation, non-energy benefits etc.) in the evaluation, but the Evaluation Administrator must quantify the impacts using standards applicable in the particular industry. For example, an Ontario utility may wish to calculate emission credits associated with electricity demand reduction. To do so, the utility should apply the standards and protocols set out in the International Program Measurement and Verification Protocols (IPMVP) on emissions credits, which may require measurements before and after a retrofit. Furthermore, claiming and selling/assigning any emission credits to other organizations is subject to a complex and changing legal framework. As such, Evaluation Administrators must understand the protocols applicable to all impacts claimed.

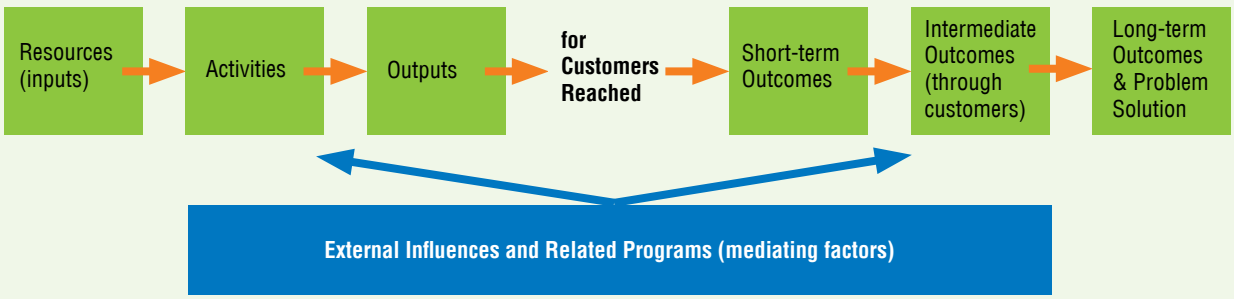
## Task 2f: Illustrate and Annotate Program Logic

As noted, a program logic model is an illustration of program logic as a causal chain from resource expenditure to the long-term impacts of the program.

**Figure 1.0: The Basic Elements of a Logic Model** shows the basic elements of a program logic model. Crafting a good logic model requires that Evaluation Administrators and Program Managers think about what the program is attempting to achieve and what the causal chains are to achieve the desired outcomes.

The arrows linking program activities to outputs, outputs to outcomes, and outcomes to impacts represent the intended cause and effect relationships underlying the program. As such, these linkages must be explained in the EM&V Plans.

**Figure 1.0: The Basic Elements of a Logic Model[2]**



## Task 2g: Verify Savings Attribution Pathway

The Evaluation Administrator should document the intended impacts of the program (reduced energy demand and savings.) and unintended impacts that may occur as a result of the program. For example, a residential demand response/load control initiative could provide a mechanism, for example, *a programmable thermostat*,) that if program participants use at times that are outside those expected (*an unintended impact*) as a result of their increased awareness (a cognitive outcome), changes in heating and air conditioning use (*behavioural outcome*) that result are unintended. The primary reductions in peak demand that result from the thermostat (*intended impact*) are central to the initiative. The Evaluation Administrator should clearly identify at least one (if not more than one) pathway (referred to as an attribution pathway) leading from program resource expenditures directly to energy and demand savings.

**Attribution Pathway:** A relationship from one or more program-sponsored activities to outcomes and impacts being asserted by the Program Administrator or Evaluation Administrator. The pathway is a set of logical connections between resource expenditures and specific impacts so that cause and effect can be attributed to the program offer.

By identifying an attribution pathway, the connection between program intentions and verified program energy and demand savings, including unintended savings impacts, can easily be seen.

By exploring alternative hypotheses about how outcomes evolved one can identify questions about the potential effects of market externals that can be researched. The development of a logic model helps evaluators understand all of the possible ways the program outcomes might ripple through the targeted population. Ripple effects occur, for example, when people mimic desired actions without involvement in the program or as a result of previous participation in the program. Once such additional outcomes are identified, evaluators will know to ask questions about why they occurred. Without such investigation potential outcomes may go unnoticed and both direct and indirect outcomes that could add to program impacts may be missed.

Evaluation Administrators are encouraged to look for, and document, alternate pathways for demand and energy savings. Including these pathways within the logic model provides a means for claiming energy and demand savings that result from unintended, yet highly desirable, market behaviours.

### Summary of Actions

- Specify resources (time and money) available to achieve desired effects
- Highlight on Program Logic Model: Program Outputs that lead to outcomes along Critical Savings Attribution Pathway
- Highlight Program Outputs involving significant expenditure of program resources
- Distinguish types of outcomes resulting from program
- Document program demand and energy impacts
- Document intended impacts of program
- Look for and document any alternative pathways for demand and energy savings

# Step 3: Properly Scope Program Evaluation

## Key Points / Highlights

**Properly Scoping Program Evaluations involves the following tasks:**

**3a.** Select Elements from the Program Logic Model to be Assessed

**3b.** Specify Types of Evaluation to be Completed

**3c.** Clarify Intended Use of Evaluation Findings

**3d.** Draft Research Questions

## Task 3a: Select Elements from the Program Logic Model to be Assessed for the Evaluation

Budgets for evaluations are generally constrained. Therefore, staff will have to make choices regarding the scope of the evaluation.

Evaluation Administrators and Program Managers must choose which elements will be evaluated. The selection of elements should be based on the logic model created under Step 2. Depending on the size and magnitude of the evaluation, all or some elements in the Attribution Pathway (see **Figure 1.0**) can be included in the evaluation.

## Task 3b: Specify Types of Evaluations to be Completed

When all the elements that will be included in the evaluation were selected, the evaluation objectives associated with the elements should be specified. In developing a statement of work for an Evaluation Contractor, the evaluation administrators should determine the types of evaluations that should be requested to ensure that the evaluation objectives can be met.

Types of evaluations include:

- **Outcome Evaluation** – this is conducted to verify cognitive and behavioural changes believed necessary for the realization of program objectives (outcome evaluations are summative and *ex post*).
- **Impact Evaluation** – this is conducted to measure the change in energy consumption or demand caused by the program (Impact Evaluations are summative and *ex post)*. Such evaluations can also include M&V engineering processes used for developing new or improved *ex ante* evaluation estimated savings.
- **Process Assessment Evaluation** – this is conducted to explain the program impact and/or identify lessons learned to inform future program strategies (in other words, to develop conclusions about program performance). Such assessments can include conducting behavioural research for the purpose of developing new or improved ex ante evaluation estimated savings.)

- **Market Study Evaluations** – the study of market characterization is conducted because it can contribute to evaluating the impact of codes and standards, TOU rates, and so on it can act as a benchmark for market transformation elements of efficiency programs and may contribute to the development of *ex ante* savings estimates.
- **Cost Effectiveness Evaluations** – a cost effectiveness evaluation includes "standard" cost effectiveness tests as provided in Technical **Guide 2: Cost-Effectiveness Guidelines**. Where the Evaluation Administrator or Evaluation Contractor deems it appropriate, it may also involve exploring the cost-effectiveness of individual measures, program elements, and/or implementation procedures.

Keep in mind that the analytical methods used in each type of evaluation will depend on the type of program evaluated. For example, program administrators will use a different analytical method for a demand response program impact evaluation and will report different information for such an evaluation than for an evaluation of an energy efficiency program.

When conducting evaluations, one must develop a robust analytical approach that yields statistically significant findings. Part Two of this guide provides guidance on the assessment of conservation programs. The manner in which a program is offered must be considered in the assessment. Therefore, all EM&V plans must provide a strategy that will result in evaluated savings estimates associated with the program.

When applicable, Evaluation Administrators must work with the Evaluation Contractor and apply the methods recommended in the Part Two. Programs must follow the guidance developed in the following sections:

- **Technical Guide 3: Process Evaluation Guidelines** are for all instances where a process assessment is sought or where concerns over operational efficiency have been expressed.
- **Technical Guide 4: Project Level Energy Savings Guidelines** are for single site implementation programs, such as those used for custom industrial process optimization.
- **Technical Guide 5: Gross Energy Savings Guidelines** are for most mass market energy efficiency programs and conservation initiatives.
- **Technical Guide 7: Market Effects Evaluation Guidelines** are for programs thought to change conditions, processes, or practices.
- **Technical Guide 8: Net-to-Gross Adjustment Guidelines** are for all savings claims and primarily used for energy efficiency programs.
- **Technical Guide 9: Demand Response Load Impact Guidelines** are for demand response initiatives.

## Task 3c: Clarify Intended Use of Evaluation Findings

The Evaluation Contractor and Evaluation Administrator must understand how the evaluation will be used beyond the determination of verified savings estimates and must document these intended uses within the EM&V plan. For example, a program design team may commission a research study to assist in designing a program to estimate measure level effectiveness. The intended use of the evaluation findings will influence the evaluation

plan and the manner in which the data is presented.

### Task 3d: Draft Research Questions

Once evaluation objectives are established program administrators must convert them into general and specific research questions that then become the focus of the evaluation effort.

Program administrators should derive the general questions from the evaluation objectives. Each general question implies specific research questions that are capable of being answered through data collection and analysis.

Clear research questions help build consensus among evaluation stakeholders and offer guidance on the areas of investigation, which increases the likelihood of coming up with valuable evaluation findings, insightful conclusions, and useful program recommendations. Properly stated research questions:

(a) flow directly from the evaluation objectives

(b) are specific and solicit significant finding

(c) can yield answers that are actionable and

(d) are answerable within the constraints of the evaluation budget and other resources.

Keep in mind that for each research question there are distinct experimental considerations, such as the sample size and parameters, relevant comparison group, data collection methods, and so on. As a result, few research projects effectively answer more than a handful of research questions. The narrowing of research questions is a fundamental activity within EM&V planning and is necessary for a manageable evaluation. Evaluation Administrators should narrow the inquiry to less than a dozen, well-crafted research questions.

#### Summary of Actions

- Choose the elements to be evaluated
- Ensure evaluated savings estimates are provided rather than deemed savings estimates
- Convert evaluation objectives to general and specific research questions

# Step 4: Identify Analytical Approaches to Address Research Questions

## Key Points / Highlights

**Identifying Analytical Approaches to Address Research Questions involves the following tasks:**

- **4a.** Construct Chain of Logic Connecting Resource Expenditure to Program Impact
- **4b.** Explore Factors that May Influence Program
- **4c.** Document Market Conditions and Research Constraints
- **4d.** Specify the Populations of Interest and Sampling Strategy
- **4e.** Identify Key Metrics for Each Program Element to be Studied

## Task 4a: Construct Chain of Logic Connecting Resource Expenditure to Program Impact

Evaluation administrators must convert the research questions developed in Step 3 into experimental inquiries to estimate demand and energy savings. In general, each research question will require verification of outputs and outcomes and quantification of impacts.

Converting the research question must be done by testing a series of research hypotheses along the "attribution pathway" (see **Step 2**) associated with each research question under investigation.

An example of a hypothesis often used in our industry is that a particular financial incentive caused the participant to adopt the particular energy efficiency measure. Like all hypotheses, that hypothesis may or may not be supported by evidence. Given that it is commonly accepted that some program participants would have adopted the particular measure without the incentive, it is clear that common hypothesis is not always supported. Still, the hypotheses may be supported more often than not. So, the attribution pathway is still valid, but only for a proportion of the participants

Evaluation Administrators and Program Administrators must not stop at an overly simple inquiry; instead they must validate the theory underpinning a program based on a continuous set of hypotheses along the attribution pathway. For the theory to remain valid, the hypotheses must be explicitly stated in the evaluation plan and tested using valid analytical methods.

### Task 4b: Explore Factors that may Influence Program

Considering the unintended impact of external factors helps evaluators isolate and report on program cause and effect. Formalizing the consideration of unintended impacts of a program is necessary to attribute impacts to specific program offers and to allocate savings.

Examining external, non-program factors that might influence an expected outcome can reveal non-program relationships and suggest alternative hypotheses about how outcomes occur. The process of examining the underlying theory, making the logical relationships explicit between the program components, and considering external influences can suggest the need for changes to a program's design or the evaluation plan.

### Task 4c: Document Market Conditions and Research Constraints

Deciding on the resources to dedicate to program evaluation involves simultaneous consideration of:

(1) the importance of the program decisions to which the evaluation will contribute (i.e. achieving CDM targets)

(2) the resources needed to satisfy the evaluation's objectives and,

(3) the resources the program can afford.

Where external influences prohibit the study of critical elements on which the program is based, the constraints prohibiting the analysis should be explicitly stated within the program evaluation plan. The rationale for doing so is not only for simple transparency; rather the reason is grounded in the fact that evaluation staff or contractors will likely see the importance of various elements of program theory the protocols require explicit disclosure of constraints to an area of relevant investigation.

Evaluation Administrator should narrow the areas of investigation before the evaluation contractor begins their work. Doing so after-the-fact can jeopardize the evaluators' autonomy to explore program cause and effect.

### Task 4d: Specify the Populations of Interest and Sampling Strategy

Quantitative research aims to determine the relationship between one or more independent variables (for example, installation of program measures) and a dependent variable (for example, GWh savings) within a target group (for example, low-income households).

Evaluation Administrators may use either a descriptive or experimental study approach to determine the relationship between independent or dependent variables.

A **descriptive study** establishes only the association between variables, such as, the propensity for energy savings among program participants. An **experimental study**, on the other hand, establishes causality between installed energy efficiency measures and observed demand reductions.

In practice, true experiments are difficult to establish for CDM initiatives. So, the industry has adopted quasi-experimental approaches that accept market characterization and measure effectiveness testing that is commonly used to support or confirm findings from other evaluation efforts. Methods such as tabulating descriptive measurements and finding the statistical significance of a relationship between variables are usually not thought of as research designs, but in fact, the process of going from the results of these analytical procedures to answer evaluation questions involves hypothesis testing and, therefore, undergoes a similar process to research design.

If, however, a Program Evaluator needs to determine the proportion of a quantified outcome that can be attributed to the particular program instead of to external influences (that is, the Evaluation Administrator needs to conduct an impact evaluation), then one must use a credible research method. The method should allow to estimate what actions participants would have taken (outcomes) had the program not existed. The difference between what participants would have done and what they actually did, is the amount of the observed outcome that can be attributed to the program.

Evaluation research designs that allow Evaluation Administrators to make claims of effect are called "experimental" or "quasi-experimental" designs.

In the experimental method Evaluation Administrators must fully define the study and comparison populations. They must describe how to determine each sample group, the numbers included in the study, and the resulting precision expected. Unless an exception is granted (and exceptions are typically only granted for market effects), the confidence in the quantitative findings must be at least 90%.

Evaluated savings (as opposed to deemed savings estimates) must be provided, unless unique circumstance prohibit comparison group selection (for example, if evaluating a large industrial energy efficiency program and similar conditions or processes are unlikely to exist for comparison, or where there are no or limited comparison groups such as in new construction). In such cases, refer to the **Technical Guide 4: Project-Level Energy Savings Guidelines** or the **Technical Guide 10: Guideline for Statistical Sampling and Analysis.**

### Task 4e: Identify Key Metrics for Each Program Element to be Studied

The research questions developed earlier will help prioritize the areas of study around essential program elements

Evaluation administrators must identify the sources of data for each question, along with alternative strategies for collecting data where data access or integrity may be suspect. Where there is a lack of data to calculate indicators for each program indicator, one must revisit the research questions.

In a separate table organize the program elements against the program theories. For each program element being studied identify the potential data source and collection method.

### Summary of Actions

- Convert research questions in experimental inquires to estimate demand and energy savings
- Specify irrelevant assumptions to be excluded from investigation
- Ensure evaluated savings are provided
- Create table highlighting key metrics and linking them to relevant theories underlying the program.

# Step 5: Specify Evaluation Deliverables

## Key Points / Highlights

**Specifying Evaluation Deliverables involves the following tasks:**

- **5a.** Draft EM&V Project Gantt Chart(s)
- **5b.** Consider Cross-Cutting Approaches
- **5c.** Identify Study and Comparison Groups
- **5d.** Highlight Analytical Methods Expected
- **5e**. Explore Data Collection Opportunities and Constraints
- **5f.** Change in Hourly (8760s) Load Shapes Explored
- **5g.** Formalize the Draft Evaluation Plan

## Task 5a: Draft EM&V Project Gantt-type Chart(s)

As a result of working through the previous steps, the evaluation requirements have been defined. Using established project management techniques, the Evaluation Administrator must manage the delivery of requirements.

Evaluation deliverables must be depicted in a project chart (for example, a Gantt chart or something similar) showing the timing of each component of the EM&V project and resources related to each component. Evaluation administrators should show the types of evaluations to be completed over the course of the portfolio/program offer. Keep in mind that to show this, the chart may have to include timeframes beyond the program expiration date. For example, for weather-sensitive loads, the chart may have to show timelines that extend to 18 months or more, as utility data may need to be captured for one full year, with an additional six months required to analyze and report the final program year savings.

The Evaluation Administrator must decide on the frequency, duration, and timing of planned evaluations, as well as the types of evaluations that will be completed. The types of evaluations within the scope of the 2011-2014 program offerings are those described on page 29 **(Draft Evaluation Plan Template 2011-2014)** .

Types of studies defined in **Task 3b: Specify Types of Evaluations to be Completed** should be represented as milestones on the project chart. The evaluation administrator must include details regarding each type of evaluation in the project chart, including the start and end dates of major deliverables related to the particular evaluations. Time should be allocated to each major deliverable within the scope of each evaluation including, among other things, the following evaluation activities:

- **Finalizing the Evaluation Plan** – The Evaluation Contractor who will conduct the actual evaluation may need to refine the Draft Evaluation Plan presented to them. When putting together the Final Evaluation Plan, be sure to leverage the Evaluation Contractor's experience and knowledge to ensure that the scope and resources dedicated to the evaluation are optimal and realistic.

- **Developing Data Collection Instruments** – Data collection instruments include surveys, field work, focus groups, etc. The Evaluation Contractor with assistance from the Evaluation Administrator must coordinate data collection from program implementers, utilities and program staff. Keep in mind that, depending on the data available, it may be necessary to allocate significant time and resources for developing data collection instruments throughout the evaluation process.
- **Collecting Field Data** – In-field data collection involves data about the relationship between the LDC and its customers. Note that because such information can be considered sensitive (i.e. use of personal information), the Evaluation Administrator must monitor in-field data collection efforts. Field data can be quantitative (collected from metering studies, mystery shoppers, on-site inspections, etc) and/or qualitative (collected from focus groups, panel studies, process reviews, etc) Whether the data is qualitative or quantitative the collected information must be summarized without bias.
- **Presenting the Findings** – The dates at which the summary of findings will be presented to the Evaluation Administrator must be included on the project chart. These dates are often a couple of weeks after surveys, or at pre-defined periods before the preparation of the draft evaluation report. Evaluation Administrators must ensure the Evaluation Contractor presents a summary of its findings and supporting data in a timely, constructive manner.
- **Delivering the Draft Evaluation Report** – The project chart should specify the date when the first draft of the evaluation report is to be delivered. When setting this deadline it is critical to allow sufficient time for the program administrator and other interested stakeholders to internally review findings and results emerging from the draft evaluation reports.
- **Delivering the Final Evaluation Report** – The delivery date of the final evaluation report must be specified in the project chart.

## Task 5b: Consider Cross-Cutting Approaches

Conducting multiple analyses or evaluations simultaneously is known as cross-cutting. Applying a cross-cutting approach can help optimize evaluations. For example, when one adjusts an end-use measure and that adjustment causes changes to another end-use measure, the resulting change is referred to as a cross effect. A cross-cutting approach can be used to analyze cross effects. Where the Evaluation Administrator thinks using a cross-cutting approach would be useful, the EM&V scope of work should explicitly state that the approach should be used.

Because different scenarios could theoretically result in either overstatement or understatement of program savings, the Evaluation Administrator must disclose how cross-cutting techniques will be used to optimize evaluation cost-effectiveness while adding to the reliability of evaluation findings.

## Task 5c: Identify Study and

**Examples of When Cross-Cutting Is Useful**

A lighting program may involve replacement of incandescent lamps with compact fluorescent lamps that can provide the same lumen output with greater efficiency. Installation of the compact fluorescent lamps also means that less heat would be emitted by the light source, which could have a positive effect on cooling loads (adding to efficiency gains when a space requires cooling) or a negative effect on heating loads (reducing efficiency gains in a residential single family home) if the installations occur in conditioned spaces. To account for these cross effects, cross-cutting analytical approaches must be used where the effects are expected to be substantive.

Energy efficiency initiatives often have some effect on seasonal or peak demand. Therefore, the impact resulting from one or more energy efficiency initiatives affecting the same market should be considered when evaluating demand response initiatives within the same sector. Evaluation contractors will often look only at the direct influence of one program on another where a participant in one program is screened for participation in another. By failing to use a cross-cutting approach in such a case, the Evaluation Administrator risks understating savings.

### Comparison Groups

A brief description of the anticipated study group and comparison groups must be stated for each analytical approach that will be used in the evaluation. The Evaluation Administrator must explicitly state in the evaluation plan the need for a comparison group. Furthermore, the Protocol specifies the methods by which comparison groups are selected – the selection process should be conducted by the Evaluation Contractor. Where possible, the comparison group(s) should be representative of the study group. The EM&V plan must consider comparability between the study and comparison groups in a manner which result in statistical significant findings.

### Task 5d: Highlight Analytical Methods Expected

The Evaluation Administrator develops a list of analytical methods to best achieve the defined objectives in the EM&V Plan. The Evaluation Administrator must specify the primary analytical methods the Evaluation Contractor is expected to use. For example, the Evaluation Administrator may specify that estimated program savings should be based on billing analysis rather than engineering models.

Furthermore, the EM&V plan must include information regarding the savings attribution model. Attribution models are used to define the process an evaluation will follow to determine whether energy and demand savings are due to program influence.

### Task 5e: Explore Data Collection

### Opportunities and Constraints

The Evaluation Administrator must make clear to the prospective Evaluation Contractors what data will be available for analysis and the timing of data acquisition. And the Evaluation Administrator should ask Evaluation Contractors to propose strategies for collecting the desired data and/or options for collecting similar data. If there are any constraints related to the data acquisition, the Evaluation Administrator must highlight these constraints in the scope of work provided to the Evaluation Contractor.

If data acquisition constraints exist, they must not be allowed to affect evaluation practices and the integrity of an evaluation. Most Evaluation Contractors have encountered data constraints and have experience with similar analyses from which they likely can recommend alternatives for data collection.

Where the data constraints are expected to be persistent, the Evaluation Administrator must indicate the steps that are to be taken to ensure EM&V best practices are upheld. Timelines within which data constraints are required to be resolved must be set out in the EM&V plan and time should be built into future evaluation cycles, or at least discussed with the Evaluation Contractor, to ensure the constraints are resolved.

### Task 5f: Explore Changes in Hourly

### (8760s) Load Shapes

With the introduction of smart meters to Ontario's residential sector, some LDCs have usage and demand data that can be analyzed as a part of the evaluation of load shapes. Evaluation contractors that have experience with load shape analysis can provide insight into how interval data can be used for program evaluation.

Given Ontario's electricity reliability standards, using interval data for load shape analysis may be much more illustrative of the achieved impacts than traditional annual estimates of demand and/or energy savings. As a result, when estimating demand and energy impacts, where appropriate, priority may be given to using interval data.

### Evaluation Plan

Evaluation Administrators must create a Draft Evaluation Plan. The Draft Evaluation Plan, must conform to the specifications established in **Step 7: Evaluation Plan Development Guidelines**.

### Task 5g: Formalize the Draft

#### Summary of Actions

- Create project chart showing timing of each component of EM&V project and resources related to each component
- Decide on the frequently, duration, and timing of planned evaluations
- If cross-cutting techniques are used, disclose how they will optimize evaluation cost-effectiveness
- Specify the primary analytical methods Evaluation Contractor is expected to use
- Provide information about savings attribution model used
- If there are constraints related to data collection, highlight them in Evaluation Contractors scope of work

# Step 6: Evaluation Classification Protocols

## Key Points / Highlights

**When undertaking a program evaluation, the following types of evaluation should be taken into consideration**

- **6a.** Impact Evaluations
- **6b.** Process Evaluations
- **6c.** Market Effects Evaluations
- **6d.** Cost-Effectiveness Evaluations
- **6e.** Outcome Evaluations

## Introduction

In the Draft Evaluation Plan, the Evaluation Administrator must specify the types of evaluations to be completed. Impact, Process, Market Effects and Cost-Effectiveness Evaluations are the most discussed evaluations for energy efficiency programs. Another type of evaluation is an Outcome Evaluation. Outcome Evaluations are often useful when there is a need to establish the cause of observed effects. Therefore Outcome Evaluations can be highly relevant to the research

## Task 6a: Impact Evaluations

Impact Evaluations are assessments of both intended and unintended effects that can be attributed to a program, policy, or project. Impact evaluations are the most rigorous of all evaluations since the attribution chain must be established from program outputs through observed outcomes to the realization of tangible impacts. Such evaluations are most appropriately applied to those measures that have a direct causal impact, like the installation of insulation on building heating and cooling efficiency.

For an impact evaluation, the contribution of external factors toward the realization of desired impacts should be limited to factors that are reasonable and can be accounted for within the analysis. In the prior example of building insulation, the external factors are weather and the set point for the interior temperature. For weather effects we generally normalize to some long-term weather trend or establish a reference weather year. For participant behaviours we hypothesize and test whether the program under study substantively influences the behaviours of the target market (participants).

In general, an impact evaluation addresses the following question: *What are the verified quantifiable effects (impacts) attributable to the program?* For CDM initiatives, the primary impacts are energy (GWh) savings and demand (MW) reductions.

**Examples of research questions used in Impact Evaluations:**

- What is the direct impact of the entire program on energy savings and demand reductions?
- What is the direct impact of individual program elements or behaviours on energy savings and demand reductions?
- What is the direct impact of the overall program on non-energy benefits (NEBs)?
- What is the direct impact of individual program activities on non-energy benefits?
- What is the magnitude of observed effects? What proportion of those effects can be attributed to the program?
- What key factors are responsible for the verified savings?
- What could have caused the observed energy saving behaviours, if they were not caused by the program?
- What behaviours were adopted by program participants when compared to those of non-participants?

## Task 6B: Process Evaluations

Process Evaluations are assessments of program policies, procedures and practices, along with a review of organizational controls that contributed to their realization. Unlike management consulting mandates, which tend to be forward-looking, process evaluations are retrospective in nature.

Process Evaluations review practices that were implemented over the period under review, outlining the strengths and weaknesses of program processes and seeking opportunities for improved operational efficiencies.

Process Evaluations verify program expenditures, review the efficacy of the services provided by the program and document the resulting operational outputs to program objectives.

**Examples of research questions used in Process Evaluations:**

- Are program designs and supporting organizational controls adequate?
- Is the program producing the outputs intended?
- Are resources reasonable relative to program objectives?
- How might the program be improved?
- How can the program be modified to improve cost-effectiveness or to enhance the stream of benefits?

The Evaluation Administrator should work with the Program Administrator to re-state specific program concerns into researchable questions to be investigated by Evaluation Contractors. The following general questions are good examples to be reframed for a program Process Evaluation:

- Are program objectives set too high? Too low? What market actors are being served and through what delivery channels?
- Is it easy for customers to join or participate in the program? What motivates them to participate?
- Are the available tools and services supporting program delivery? Are the tools used properly by program delivery agents?
- Are customers participating at expected levels? Are some customer groups participating more than others? Why?
- Which tools and services are being used? By what groups? Are customers satisfied with the program?
- Are the resources assigned to the various program components adequate to achieve the desired objectives?
- Is the program leveraging available funds effectively? How could additional resources be applied? Are detailed program expenditure records maintained?
- How can the program better serve non-participants and hard-to-reach populations? What recommendations do participants and non-participants have for the program?
- Would administrative improvements better support the provisions of program services?

## Task 6c: Market Effects Evaluations

Market effects evaluations assess the changes, due to program, policy, and projects, in both short-term and long-term structural elements of the market place, as well as the cognitive processes and behaviours of key market actors that lead directly to energy savings and demand reductions.

For resource acquisition programs, market effects evaluations serve to measure the net effect of programs by accounting for key major net-to-gross effects: spillover and free ridership. Market effects evaluation also seeks to attribute transformational impacts on the market resulting from application of codes and standards, legislation, innovation, and capability-building initiatives.

Evaluation Administrators should include market effects evaluations when Program Administrators suggest intended changes to target markets, or when they espouse a long-term approach with proposed exit strategies, or suggest that actors' behaviours will persist beyond the scope of the intervention.

**Examples of research questions used in market effects evaluations:**

- Have changes occurred in the willingness or ability to produce, distribute, or service new energy efficient technologies?
- What changes or effects are associated with individual program components/activities?
- How have the behaviours of targeted actors changed over time?
- What external factors are related to the achievement of observed market effects? What is the strength of those relationships?
- How effective has the program been in reducing market barriers?
- Have desired behavioural outcomes continued over time?

## Task 6d: Cost-Effectiveness Evaluations

Cost-effectiveness evaluations measure the stream of benefits against the costs to achieve those benefits. In general, cost-effectiveness evaluations are implemented at the program level by leveraging industry-established tests. The details of the tests required in Ontario can be found in **Technical Guide 2: Cost-Effectiveness Guidelines**. Cost-effectiveness evaluations may also target measures, program delivery agents, and specific program activities.

**Examples of research questions used in cost-effectiveness evaluations:**

- How much did the verified energy savings and demand reductions cost to achieve?
- What benefits resulted from individual program activities relative to their costs?
- Was the program cost-effective? Does this program pass the cost-effective hurdles established for the Province of Ontario?
- Which delivery channels are working best to achieve program objectives?

## Task 6e: Outcome Evaluations

Outcome evaluations are similar to market effects evaluations except that output evaluations do not link program expenditures to program impacts. Outcome evaluations are used to document causal linkages between program outputs and program outcomes or, to test elements of complex program theory.

Outcome evaluations are used to establish the efficacy of market transformational initiatives, policy directives, social programs and other interventions within a complex environment where direct impacts may be difficult to isolate from influences beyond those resulting from program-sponsored activities.

**Examples of research questions used in outcome evaluations, often the first step in an impact assessment looking at indirect or unintended program impacts:**

- What are the secondary and tertiary benefits resulting from the program under consideration (for example, persistence, delayed implementations, spin-offs)?
- What were the nature and magnitude of non-energy benefits associated with the program?
- What were the nature and magnitude of non-energy benefits associated with individual program activities?
- What were the causes of any unintended program impacts?

### Summary of Actions

- Determine what elements need to be assessed to quantify program impacts
- Identify the type of evaluation used to assess the program impacts
- Verify whether the examples of research questions pertain to the program evaluation

# Step 7: Evaluation Plan Development Guidelines

## Key Points / Highlights

**Evaluation Administrators should consider the following tasks when developing an Evaluation Plan:**

**7a.** EM&V Plan Content and Structure

**7b.** Final Evaluation Plan (FEP)

**7c.** Key Evaluation Consideration

The Evaluation Administrator authors the evaluation planning documents. The first step is development of a *Draft Evaluation Plan*. An evaluation plan results from the steps presented in above.

The Evaluation Administrator uses the logic model to select areas of study and to choose the types of evaluations sought.

Program Managers and Evaluation Administrators use their knowledge of program objectives, delivery mechanisms, and motivations to properly scope the evaluations needed. Evaluation planning includes allocating program resources to monitoring, measurement, verification and evaluation.

**Types of Evaluations and Assessments Typically Included in Draft Evaluation Plans**

- **Impact Evaluations** – these look at behavioural outcomes and their likelihood to generate the intended program impact (typically demand reductions and energy savings). They may also look at both positive and negative unintended impacts. To the extent that unintended impacts have a substantive impact on program outcomes, they should be evaluated.
- **Process Evaluations** – these are used to explore the methods, activities, and expenditures used to generate program outputs. They evaluate things like the effectiveness of promotional campaigns, informational materials, educational seminars, training, financial assistance, technical assistance, etc.
- **Market Effects Evaluations** – these are used to estimate the contribution of program outcomes to market trends. They may also be used to evaluate the converse: how trends in the market place (for example, electricity pricing, rate schedules, legislation, and so on) impact program outputs.
- **Cost-effectiveness assessments** – these are used to quantify and analyze the benefit and cost streams (for example, cost-benefit ratios). These are generally conducted after the impact and process evaluations have been completed.
- **Outcome Evaluations** – these are used to explore how behaviours arise from program-sponsored activities. They seek to explain behavioural choices in the context of desired attitudes and added abilities resulting from program outputs.

Each of these types of evaluations is discussed in detail in **Step 6: Evaluation Classification Protocols**.

## Task 7a: EM&V Plan Content and Structure

An example of a Draft Evaluation Plan Template is provided (p. 29). Unless there is a specific reason for using some other format, using it as such is recommended because it facilitates easy review of plans and approvals from Program Managers and executive management across different programs.

## Task 7b: Final Evaluation Plan (FEP)

A *Final Evaluation Plan* builds on the *Draft Evaluation Plan*. The Evaluation Contractor works with the Evaluation Administrator to formalize all elements and objectives of the evaluation. The Evaluation Contractor submits the *Final Evaluation Plan* to the Evaluation Administrator for final approval. The FEP is detailed enough to ensure the approved evaluation activities yield a high level of confidence in the reported energy savings, demand reductions and program cost effectiveness.

## Task 7c: Key Evaluation Considerations

When planning evaluations, Program Administrators and Evaluation Administrators should consider how the evaluation serves as a management tool. The evaluation provides savings estimates that demonstrate program impact and cost-effectiveness, which may be used for regulatory purposes. Evaluation findings are used to improve both short-term and long-term impacts, allowing mid-course corrections to enhance program achievement. To realize these benefits it is important to keep in mind that evaluations are not meant as mere audits of program performance.

To help ensure the usefulness of evaluations, keep the following in mind:

- **Integration of Evaluation into the Program Implementation Cycle** – Before describing the evaluation planning process, it is important to understand how it is integrated with the program planning-implementation-evaluation cycle. This is necessary to align budgets, schedules, and resources. It is also a way to ensure that data collection supports planned evaluation efforts and is embedded with program delivery
- **Program Design** – The *Draft Evaluation Plan* is prepared as part of the program design and an evaluation budget is assigned at that stage. On completion of the program design, the evaluation plan is implemented to ensure data is collected and reported in a timely manner, allowing for incremental feedback to guide Program Managers.
- **Program Goal Setting** – If the program (or portfolio) goal is to save electricity during peak hours, the evaluation goal is to accurately document how much electricity demand is reduced during the peak hours (gross savings), how much of these savings can be attributed to the program (net savings) and then, in the case of Tier 1 programs, the savings are allocated to individual LDCs.
- **Preparing for Program Launch** – Ideally, the draft evaluation plan should be prepared before the program is launched. If it cannot be developed before program launch, it should be drafted as soon as possible following program launch. Baseline data should be collected before, or soon after, program launch so that market effects resulting from the program offer are documented.

- **Defining the Evaluation Objectives** – Evaluations focus on the linkage between program outputs and the resulting program outcomes. The evaluation should provide guidance to the Program Administrator on ways to enhance program efficacy. To this end, Program Administrators and regulators need to be assured that the evaluations conducted will deliver the type and quality of information needed.
- **Program Implementation** – Some baseline data collection and all program reporting continues throughout program implementation. The incremental data informs and updates program metrics. The Evaluation Administrator should analyze and present performance metrics to Program Managers as findings from Evaluation Contractors. Keep in mind that evaluation activities often continue after the program year is completed.

Evaluations must also be properly scoped. Addressing issues that are not program priorities or issues, or employing unnecessarily complex methods, can waste valuable resources. When faced with limited evaluation resources prioritizing the key activities will ensure the evaluation objective have been met without straining resources.

## Summary of Actions

- Scope of Evaluation deliverables
- Create a draft Evaluation Plan
- Work with Evaluation Contractor to complete the Final Evaluation Plan

## Draft Evaluation Plan Template

### Program Overview

#### Program Description

**Provide a short introduction of the program offer from the perspective of the program manager. It should provide a high-level description of the planned program strategy. Where appropriate include the following descriptions:**

- **Goals and Objectives:** A statement of the goals and objectives for the program and the rationale for the evaluation
- **Target Market:** Profile each market segment targeted by the program offer. Describe the size and characteristics of each target market. The target market should match the segments defined in Program Logic Model.
- **Eligibility Criteria:** Describe the protocols/procedures that will be used to qualify program applicants or markets targeted.
- **Key Program Elements:** Highlight the intended program process flow. Each program element should be identified in the 1-page graphic and annotated in the text that follows. This information should be drawn directly from the program design documents.
- **Program Timing:** A schedule of when the key elements of the program will be in market, including program launch date and program end date.
- **Estimated Participation:** Estimated participation, by measure if applicable, for the program.

#### Program Theory / Program Logic Model (if available)

**Introduce the mechanisms by which the program will function.**

Even when a program manager provides a detailed logic model, the evaluation administrator should investigate independently the causal influence of each program element towards the realization of intended programmatic impacts. The program manager should review the logic model and ensure it is an accurate portrayal of the program theory.

Annotate the program logic model from top (resource allocation) to bottom (intended impacts). Of particular interest are the linkages between program outputs and observed outcomes. Where practical, each connecting line or arrow should be annotated as a researchable programmatic assumption (null hypothesis).

### Previous Program Evaluations

**A brief description of similar program evaluations relevant to the program, including pilots.**

### Evaluation Goals and Objectives

**Introduce the goals and objectives of the planned evaluation and indicate the rationale for the evaluation: administrative (verified savings), experimental (measure effectiveness), qualification (program pilot), or operational (cost-effectiveness).**

#### Overarching Concerns

Provide a list of questions posed by program stakeholders to the evaluation administrator. These should be categorized and refined as necessary to adequately communicate the areas of investigation sought by those sponsoring, operating, or participating in the program offer.

#### Research Questions

From the overarching concerns of program stakeholders, a set of research questions should be developed by the evaluation administrator and presented here. The number of research questions should be limited and prioritized based on reasonable use of resources.

# Draft Evaluation Plan Template

## Evaluation Approach

**Introduce the details of the approach that follows.**

### Evaluation Type (repeat for each type)

Provide a description of the types of evaluations required and summarize the experimental approach anticipated. Include in the title, the frequency of the evaluation type such as, an "Annual Impact Evaluation" or a "Year One Process Evaluation". In the description, highlight the major deliverables needed to complete each study and special methods sought from the evaluation contractor.

| | |
|---|---|
| ☐ **[Frequency] Impact Evaluation.** | Impact evaluation description. |
| ☐ **[Frequency] Process Evaluation.** | Process evaluation description. |
| ☐ **[Frequency] Market Effects Evaluation.** | Impact evaluation description. |
| ☐ **[Frequency] Cost Effectiveness Evaluation.** | Cost-effectiveness evaluation description. |
| ☐ **[Frequency] Outcome Evaluation.** | Outcome evaluation description |

**Study Focus.** Associate the planned approach to applicable research questions. Indicate how the planned evaluation activities contribute to or answer the questions at hand. This is often done in the form of a null hypothesis.

**Data Collection Plan.** Describe the processes deemed appropriate to collect, validate, and audit the data used in the evaluation.

**Analysis Methods.** Describe the specific analytical methods sought for the evaluation. For example, one may wish to normalize weather to a specific year verses a long-term normal average daily temperature.

**Limitations/Caveats.** Describe limitations and restrictions associated with intended approach; thereby, providing evaluation contractors and implementers the ability to improve upon the planned evaluation.

**Study Outputs.** Identify the specific outputs expected by the evaluation administrator of the evaluation contractor. This description may include a report template, presentation requirements, delivery media, ownership of resulting datasets, etc.

### Evaluation Dependencies

Discuss key collaborations essential to the successful implementation of the evaluation. The following are common dependencies associated with industry research, more may be added as appropriate for the planned evaluations.

| | |
|---|---|
| ☐ **Enabling Stakeholders** | Identify and discuss as is appropriate. |
| ☐ **Access Requirements** | Identify and discuss as is appropriate. |
| ☐ **Data Sharing** | Identify and discuss as is appropriate. |
| ☐ **Funding Support** | Identify and discuss as is appropriate. |

**The evaluation activities undertaken as part of the program evaluation should be carried out using the guidelines specified in the Conservation First 2015-2020 EM&V Protocols and Requirements.**

## Draft Evaluation Plan Template

### Special Provisions

**Clarify any atypical considerations associated with the planned evaluations.**
Where necessary and helpful, attach materials necessary to fairly represent the work envisioned.

### Data Collection Responsibilities

**A listing of all the data that must be collected to the support the evaluation of the program and who is responsible to collect it.**

### Evaluation Schedule

**A listing of all the physical deliverables that will be part of the Evaluation, e.g., evaluation plans, memos, interim reports, final reports.**

| Evaluation Deliverable | Date |
|---|---|
| Draft Evaluation Plan | |
| Final Evaluation Plan | |
| Other Deliverable #1 | |
| Other Deliverable #2 | |
| | |
| | |
| | |
| Other Deliverable #3 | |
| Draft Final Evaluation Report | |
| Final Evaluation Report | |

# Step 8: Hire an Independent, Qualified Evaluation Contractor

## Key Points / Highlights

**Hiring an Independent, Qualified and Authoritative Evaluation Contractor involves the following tasks:**

**8a.** Provide for EM&V Contractor Autonomy

**8b.** Request independent Verification of Program Outputs

**8c.** Select an appropriate Methodology

## Task 8a: Provide for EM&V Contractor Autonomy

An independent evaluation requires that unbiased parties with no real or perceived conflicts of interest conduct the planned evaluations. Evaluations conducted by Program Managers themselves are not considered sufficiently "independent" to verify program savings.

An organization can sponsor both a program and its evaluation but in that case, the sponsoring organization must procure a third-party evaluator to implement EM&V plan, drafted by the Evaluation Administrator. In very narrowly prescribed situations, an organization sponsoring a program may appoint an internal review board or specialized program evaluation staff to assess a program offer and implement the approved EM&V plan. In such cases the sponsoring organization must be able to demonstrate autonomy between the groups implementing the program and the groups evaluating the program.

In all cases, whether it is the Evaluation Contractor or an internal review board must be free to report their findings without consequence or retribution.

## Task 8b: Request & Ensure Independent Verified Results

The intended impacts of the programs will always be reduced energy demand and savings. The Draft Evaluation Plan must explore unintended impacts that may result from the intervention. The requirement that the Evaluation Contractor must be exploring both the positive and negative impacts expected from the program must be part of the evaluation scope of work and must set out in the contract with the Evaluation Contractor.

The Evaluation Contractor must be free to present to the appropriate regulatory authority or administrative agency its findings, results, and conclusions without limitation. Under no circumstance may valid findings of fact, substantive conclusions, verified impacts, or program recommendations be censored. Where the sponsoring organization (the Evaluation Administrator or Program Manager) and the Evaluation Contractor disagree about a point, the disagreement should be outlined in footnotes in the EM&V report. The footnotes should clearly outline the opposing arguments, including attribution to the person raising the concern.

## Task 8c: Select an Appropriate Methodology

Depending on the program evaluation different methodologies can be proposed by the Evaluation Contractor. It is the job of the Evaluation Administrator to ensure that the appropriate methodology is selected. Notably, methodologies can vary depending on the data available to conduct the analysis.

Clear and specific capacity/demand reduction targets, as well as energy savings targets, have been established for CDM initiatives. And, thanks to the installation of smart metering technologies, data related to energy use exists.

While hourly load shapes add rigor to EM&V practices, the Evaluation Administrator must not dismiss the basic principles of program impact assessment. Savings calculations require a gross-to-net savings adjustment, either by generally accepted net-to-gross calculations or through net-savings calculations based on experimental or quasi-experimental models.

This task establishes a preference for advanced analytics involving smart-meter data as a key method for the verification of demand reduction and energy savings associated with CDM initiatives.

Refer to **Technical Guide 9: Demand Response Load Impact Guidelines** when procuring for a Demand Response contractor.

### Summary of Actions

- Outline (in a footnote) any disagreements between the Evaluation Contractor's findings and conclusions and those of the sponsoring organization.
- Have an independent review of program monitoring practices carried out.

# Step 9: Vendor Selection Process Guidelines

## Key Points / Highlights

**A competitive procurement process allows the Evaluation Administrator to choose from a number of proposals, which helps the Evaluation Administrator to balance many factors in an effort to meet the evaluation priorities.**



There are a number of reasons why EM&V services should be procured through a competitive process. Second, the contracted values generally associated with EM&V services often exceed the monetary thresholds that trigger competitive procurements within the public sector. Secondly, since varied approaches can often be taken for the provisions of EM&V services, by using a competitive process the Evaluation Administrator may have several options from which to choose. Lastly, a competitive solicitation ensures multi-jurisdictional vendor support for Ontario's EM&V service requirements.

Public procurements in Ontario are expected to comply with the July 2011 Procurement Directive issued by the Management Board of Cabinet. The overall objective of this Directive is to ensure acquisition of goods and services are conducted in the most economical and efficient manner.

A benefit of relying on a competitive procurement process is that the Evaluation Administrator generally will be able to choose from a number of proposals, which helps the Evaluation Administrator balance many factors in an effort to best meet emergent priorities. Vendors often submit proposals that set forth methods that tackle issues and tasks in unanticipated, clever, and meaningful ways; providing a learning opportunity for Evaluation Administrators and Program Managers.

The Draft Evaluation Plan (found in **Step 7: Evaluation Plan Development Guidelines**) forms the basis of the request for consulting services.

### Task 9a: Evaluation Contractor Selection Process

Once a valid RFP process (as described in the section above) has been held, a winning bidder must be selected. It is important that an objective selection process be followed and that appropriate documentation of the selection process is recorded and filed.

The simplest way to avoid bias or the perception of bias in the selection process is to employ an Evaluation Contractor Selection Committee. Generally it is best to form a cross functional team representing the varying interest in the evaluation results.

### Task 9b: Budget Considerations

When issuing an RFP for evaluation services to vendors, information on the program's budget for services will not be included.

There are general guidelines on the appropriate amount to spend on evaluation relative to the size of a program. As detailed in the Protocols, the typical range is 4% to 6%. Small pilot studies where very detailed information will help inform and reduce risk for a potential broader roll-out strategy could justify spending the same amount

as the program itself. In fact, pilots could be considered a form of evaluation. On the other end of the spectrum, a program that has been running consistently for several years and that has no new or unusual activity happening in it may require only a basic level of field verification and audit and so it should not require a significant expenditure. The cost to achieve a successful evaluation is also affected by whether multiple evaluation categories are required (outcome, impact, process, market, cost-effectiveness) or just a selected one.

The second reason not to include budget expectations in an RFP is because Evaluation Contractors will propose alternate methods and approaches to achieve the same end result. And, since there is more than one appropriate and acceptable way to accomplish most energy program evaluation tasks, alternate methods may have different cost implications. It is best to allow the proponents to detail their position as to why the combination of quality and cost they propose should outrank their competitors.

A third reason is that evaluation methodologies and best practices are also evolving. So, at any time, proposals may present a new way to measure performance results. A core purpose of the competitive process is to spur this type of innovation and creative thought process. We want RFP respondents to continually strive to provide the best value proposition.

Lastly, it will be rare that the absolute best quality approach will get selected or even proposed. Energy program evaluation is always a compromise between best practice and available resources. Managing this balancing act and deciding which contractor to select is easier when a truly competitive process is followed for both the substance and cost portions of the job.

## Summary of Actions

- Public procurements in Ontario are expected to comply with the July 2011 Procurement Directive issued by the Management Board of Cabinet.

# Step 10: Coordinate EM&V Activities and Report Findings

## Key Points / Highlights

**Coordinating EM&V Activities and Reporting Findings involves the following tasks:**

**10a.** Detail Research Methodologies Employed

**10b.** Present Evaluation Findings

**10c.** Assess Reasonableness of Conclusions and Recommendations drawn from Evaluation Findings

## Task 10a: Detail Research Methodologies Employed

Evaluation reports must include a detailed statement of the analytical methods used. Such reports should include a description of the evaluation objectives, a list of the research questions addressed, the approach taken to answer the research questions, the experimental model(s) employed and the analytical methods used in the presentation of findings. This will require data collection instruments (i.e. survey work) to be appended to the report.

The descriptions used in the evaluation report must be detailed enough to allow other evaluation professionals to repeat the procedures used by the Evaluation Contractor and to facilitate audits administered by the appropriate regulatory bodies and/or administrative agencies.

## Task 10b: Present Evaluation Findings

The findings of an evaluation report should be presented clearly in either graphical or tabular format. Text must highlight key findings and link the data to the research methods used to analyze data. The Evaluation Contractor must outline in the report instances where the findings confirm or contradict earlier findings, including specific reference to the previous study.

Evaluation Administrators and Program Managers may find regular monthly program reporting and EM&V findings overlap during early stages of a program offer. Such redundancy can be helpful in verifying that critical program outputs and outcomes have been achieved. As programs mature and EM&V efforts focus on downstream behavioural outcomes and program impacts, the frequency of the reports maybe reduced, depending on the regulatory requirements of the jurisdiction.

### Task 10c: Assess Reasonableness of Conclusions and Recommendations Drawn from Evaluation Findings

Evaluation Contractors must reference their conclusions to the key findings upon which the conclusions are based. Furthermore, the conclusions must be based on the data actually collected from the evaluation process versus broad inferences based solely on their experience in other jurisdictions. It should be noted that while inferences from experience in other jurisdictions may be provided, the inferences must be provided within the context of a comparative analysis explicitly requested in the evaluation scope of work.

Because conclusions and recommendations made by the Evaluation Administrator and the Evaluation Contractor often drive policy decisions, it is important that conclusions be drawn from actual findings and that the context be clearly stated. In other words, given the effect of evaluation conclusions and recommendations on organizational priorities and budget allocations, Evaluation Administrators and Evaluation Contractors must ensure the conclusions and recommendations formulated can be supported by the research findings and fall within the scope of the funded evaluation.

#### Summary of Actions

- Provide a detailed statement of the analytical methods used
- Clearly present the evaluation findings graphically or in tabular format
- Ensure conclusions are referenced in key evaluation findings
- Ensure context for evaluation findings is stated

# Step 11: Publication of Evaluation Reports

## Key Points / Highlights

**Publication of Evaluation Reports involves the following knowledge:**

**11a.** Address Timelines and Veracity of Savings Claims

**11b.** Address Comparability of Results Between LDCs

**11c.** Address Use of Utility Billing and Meter Data

**11d.** Address Defensibility of Gross-to-Net Calculations

**11e.** Presentation of Evaluation Results

## Task 11a: Address Timeliness and Veracity of Savings Claims

The appropriate regulatory authority and administrative agencies establish annual energy savings and demand reductions that are credited to each LDC.

The information the LDC provides regarding claimed saving is used to determine portfolio savings estimates. It is important to conduct the evaluation in a timely and efficient manner so that the results can be used by the varying audiences for program enhancements, program design and forecasting etc.

The LDC savings target reconciliation as established by the appropriate regulatory authority and administrative agencies is final. As such, Evaluation Administrators and Evaluation Contractors are encouraged to administer EM&V as outlined within these protocols. By doing so, LDCs have the greatest likelihood of receiving an allocation of savings that best reflect their effort (match the sum of the mean savings reported for each program implemented).

## Task 11b: Address Comparability of Results Between LDCs

Demand reductions and energy savings are considered verified estimates of program impacts. Since point estimates of energy and demand savings may vary in both precision and levels of confidence, the statistical reliability of the reported impacts are considered when comparing impact assessments.

The Evaluation Administrator should prefer a 5% confidence interval around point estimates and ensure a .95 level of confidence for claimed impacts. Where necessary experimentally, exceptions may be used by the Evaluation Contractor. It is helpful if options, including the cost implications, for 5%/0.95 and 10%/0.90 confidence are provided for in Draft Evaluation Plan requests and responses so that Evaluation Administrators can assess the benefit-cost of increased accuracy in the context of their total evaluation budget.

### Task 11c: Address Use of Utility Billing and Meter Data

Evaluation Administrators are strongly encouraged to seek the most robust and direct measurement of energy savings and demand reductions available. Site-specific hourly load shape analysis is the preferred method for calculating achieved results.

Studies using pre/post billing and meter data comparisons are given added weight over studies using prescriptive and quasi-prescriptive estimates of savings based on measure savings assumptions. Evaluated retrofits, for example, must be both measured and verified.

Whole premise measurements should use revenue-grade meters to ensure the most precise estimate of energy use and demand requirements. Where retrofits are isolated and individually metered, meter precision must be addressed when stating the achieved energy savings or demand impacts. If information regarding metered results for both a pre-retrofit and post-retrofit period is lacking, the Evaluation Contractor may use a calibrated simulation. Use of a calibrated simulation should be a method of last resort, but it may be used when evaluating new construction, constant load lighting, re-commissioning projects, and industrial process initiatives. Please refer to **Technical Guide 5: Gross Energy Savings Guidelines** and **Technical Guide 4: Project-Level Energy Savings Guidelines**. Note that use of the International Performance Measurement and Verification Protocol (IPMVP) is an integral part of project-level savings assessments.

### Task 11d: Address Defensibility of Gross-to-Net Calculations

Gross saving estimates are not applied to LDC targets because gross savings estimates do not account for what would have normally occurred absent of program incentives or energy-efficiency upgrades. As a result, net savings are used. Given this, it is essential that the calculations used to establish net savings are defensible.

**Technical Guide 8: Net-to-Gross Adjustment Guidelines** is provided as a reference, but does not replace the expert judgment of the Evaluation Administrator and Evaluation Contractor.

Both the Program Administrator and the Evaluation Administrator must address the calculation of net savings in the development of an EM&V plan. Furthermore, the Evaluation Contractor must be provided with the latitude to adjust gross savings estimates. Where possible, evaluated savings should be normalized to long-term weather and socio-economic trends so that year-over-year savings estimates can be compared.

#### Summary of Actions

- Ensure claimed savings are accurate
- Ensure comparability of study groups
- Choose appropriate cost/confidence level
- Verify type of meter data used
- Specify meter precision information
- Explain how net savings figures were arrived at
- Consider normalizing savings to applicable long-term trends

## Task 11e: Presentation of Results

Evaluation results can be presented in a variety of ways. Evaluation Administrator should apply the preferred method to present results. However, at a high-level comprehensive evaluation report should contain the following information:

### Summary of Impact Evaluation Results

For cross-cutting evaluations, include additional columns for each initiative and a total column

| Program Metric | Initiative 1 | Initiative 2 | Total |
|---|---|---|---|
| Number of Participants | | | |
| Program Realization Rate (%) | | | |
| Gross Verified Demand Savings (MW) | | | |
| Gross Verified Annual Energy Savings (GWh) | | | |
| Gross Verified Lifetime Energy Savings (GWh) | | | |
| Net to Gross Ratio | | | |
| Net Peak Demand Savings (MW) | | | |
| Net Annual Energy Savings (GWh) | | | |
| Net Lifetime Energy Savings (GWh) | | | |

Other key Impact Evaluation findings

### Summary of Process Evaluation Results

Key Process Evaluation findings

| Research Question | Observations | Recommenations |
|---|---|---|
| | | |
| | | |
| | | |
| | | |
| | | |

### Cost Effectiveness Results

| Cost Test | | Initiative 1 | Initiative 2 |
|---|---|---|---|
| **Program Administrator Cost (PAC)** | Benefit ($m) | | |
| | Cost ($m) | | |
| | Net Benefit ($m) | | |
| | Net Benefit Ratio | | |
| **Total Resource Cost (TRC)** | Benefit ($m) | | |
| | Cost ($m) | | |
| | Net Benefit ($m) | | |
| | Net Benefit Ratio | | |
| **Levelized Delivery Cost** | $/MWh | | |
| | $/MW-yr | | |

Other key cost effectiveness results

### Conclusion and Recommendations

# Step 12: Guideline for Managing Program Evaluation Contractors

## Key Points / Highlights

**The responsibilities of an Evaluation Administrator for managing program evaluation contractors include:**

**12a.** Optimizing Resource Utilization

**12b.** Project Coordination

**12c.** Providing Data

**12d.** Quality Assurance

After the evaluation has been planned and an Evaluation Contractor assigned, program evaluation tasks must be implemented and managed. The Evaluation Administrator serves as a liaison with the Evaluation Contractors, coordinating a number of tasks over the course of the EM&V efforts. While the contracted evaluator completes the bulk of the work, the Evaluation Administrator has the following responsibilities:

### Task 12a: Optimizing Resource Utilization

Evaluation Administrators must balance resource commitments within and between multiple projects. Plotting all evaluation activities on a single research calendar helps to identify opportunities to integrate data collection strategies and analysis method, even where the activities cross programs, portfolios, or evaluation disciplines. The proper use of resources avoids sampling fatigue among study populations, maximizes the available funds, and provides valued output.

### Task 12b: Project Coordination

Work, schedules and deliverables must be reviewed daily. The management of evaluations requires the organization of meetings, the establishment of goals, management of stakeholder participation, coordination of evaluation activities among team members, integration of study findings and publishing of results.

### Task 12c: Providing Data

Evaluation requires an exchange of information between planners, implementers, program participants, trade allies, comparison groups, involved organizations, and agencies. Data tracking and warehousing requires an infrastructure for this exchange. Data quality must be ensured before an analysis will meet the reliability standards established by the industry. While this work may be sourced to specialty contractors, the transformation of raw data into consumable and valued information requires significant oversight. As part of the data collection process, the Evaluation Administrator and the Evaluation Contractor should also be familiar with the Freedom of Information and Protection of Privacy Act and privacy laws in general. In particular, a data management plan should be developed for the collection, storage, disclosure and disposal of any personal information as part of the evaluation process.

### Task 12d: Quality Assurance

Administrative agencies and regulatory authorities rely on the quality of the planned evaluations. The Evaluation Administrator is responsible for ensuring quality work has been completed before the results are published and presented to key decision makers. Quality assurance requirements have been established with the Protocols, as well as in the Technical Guidelines. The Evaluation Administrator must ensure information in each published evaluation report, summary of findings, or memo, adheres to the established standards.

#### Summary of Actions

- Ensure the program evaluation contractors are provided with sufficient resources in accordance with the contract
- Ensure the results of the evaluation adheres to the established standards

# Part 2:
# Conducting an Evaluation

Audience: Evaluation Contractor

# Introduction to Part 2

The primary audience for Part 2 is Evaluation Contractors.

This Part is comprised of Technical Guides that relate to different technical processes and techniques that Evaluation Contractors use in conducting evaluations. Because the Technical Guides in this Part cover different topics, each can be read on its own. The Technical Guides provide information on:

- Technical Guide 1: Using Measures and Assumptions Lists
- Technical Guide 2: Program Cost-effectiveness Reporting
- Technical Guide 3: Conducting Process Evaluations
- Technical Guide 4: Determining Project-level Energy Savings
- Technical Guide 5: Determining Gross Energy Savings
- Technical Guide 6: Calculating Demand Savings
- Technical Guide 7: Determining Market Effects
- Technical Guide 8: Net-to-Gross Adjustments
- Technical Guide 9: Determining Demand Response Load Impacts
- Technical Guide 10: Statistical Sampling and Analysis
- Technical Guide 11: Behaviour-Based Evaluation Protocols

## Also Useful to Program Administrators

The work carried out by the Evaluation Contractor involves data collection and analyses that can be relatively technical. To ensure the Evaluation Administrator is able to effectively manage the process and gauge the quality of the work the Evaluation Contractor is doing, the Evaluation Administrator needs a basic understanding of the relevant techniques and methods. This information can be found in Part 2. Unlike the steps set out in Part 1, the guides in Part 2 are stand-alone and provide a high-level understanding of a particular technical process.

# Technical Guide 1: Using Measures and Assumptions Lists

## Key Points / Highlights

Use of accurate and defensible technology assumptions is critical in planning and assessing conservation and demand management (CDM) programs. The assumptions on which CDM programs are found are contained in "measures and assumptions lists" (MALs). The assumptions can be used to screen and assess measures for possible inclusion in a conservation program before the program runs (ex ante). As well, the MALs are used after the program runs (ex post) to evaluate the savings generated by measures and projects undertaken as a result of participation in the program.

The Program Manager reviews input assumptions for measures that are under consideration for inclusion in a program. This information is used to generate energy and demand savings estimates and to provide input into program cost-effectiveness calculations conducted for program design. It is important to use the most recent measures and assumptions list.

Evaluation Managers are responsible for ensuring that information used in evaluations is up to date and accurate.

## Prescriptive and Quasi-Prescriptive Assumptions

Input assumptions are either prescriptive or quasi-prescriptive in nature, depending on whether application-specific information is needed to better reflect variations in how the technology is used or operated.

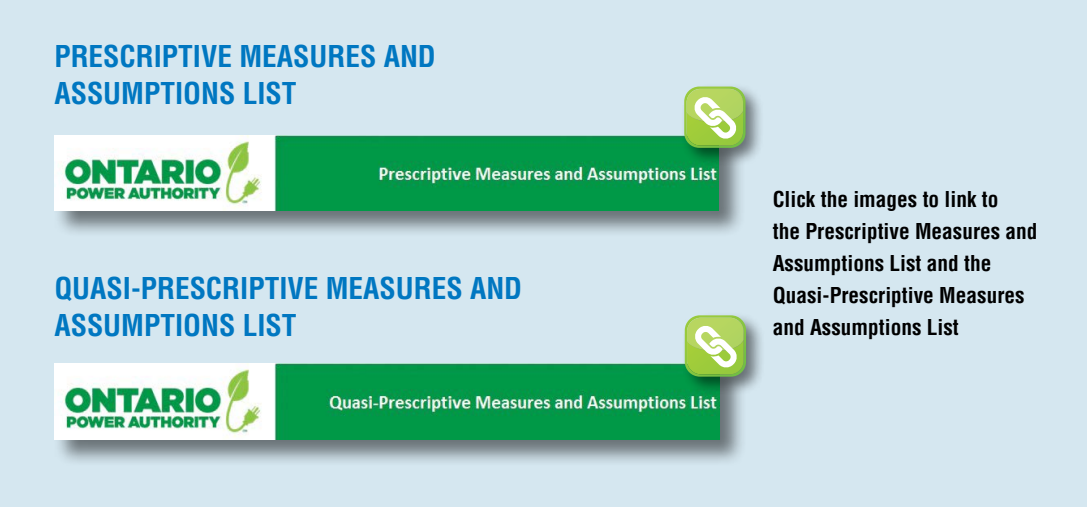
Measures that are included in a MALs are typically substantiated with documented credible results or third-party verification, testing, or certification.

Measure-level assumptions are referred to as **"input assumptions".**

**Conservation and Demand Management (CDM) programs** are programs designed to reduce the amount of electricity participants consume.

**Prescriptive measures** are measures where the energy savings are pre-determined based on how the typical conservation program participant obtains resource savings as a result of implementing the measure (the savings are determined by applying fixed input assumptions into energy and demands savings equations).

**Quasi-prescriptive measures** are measures with resource savings estimates that vary depending on the technology or type of equipment and the context in which the measures are used. Quasi-prescriptive measures provide a methodology that allows for estimating resource savings for various scenarios, rather than relying on a fixed saving value for all scenarios.




Examples of key input assumptions on which measures included in MALs are include:

- Definitions of the baseline and high-efficiency cases or technology
- Energy and demand savings resulting from high-efficiency technology
- Other resource savings (for example, natural gas, water)
- Seasonal and time-of-use (TOU) energy savings patterns (for example, periods emerging from system planning and/or regulatory rate structures such as summer, winter, and shoulder season TOU periods)
- Incremental cost data (for example, the cost differential between baseline equipment and high-efficiency equipment)
- Equipments' useful life and other assumptions about persistence

The Measure-level assumptions are reviewed periodically, and the assumptions are updated as new knowledge, information, or technologies emerge.

## Purpose and Scope of this Guideline

This guideline applies to all CDM programs that support or promote the installation of technologies with prescriptive or quasi-prescriptive assumptions and that are contained, or should be contained on the approved MALs.

This guideline provides information to CDM Program Managers, portfolio managers, and Evaluation Managers with regard to the use of input assumptions included in MALs, and to assist Program Evaluators in data collection, review and updating of measure-level assumptions.

Early in program planning and development Program Designers consult MALs to ensure that measures included in a program:

- are likely to produce reliable energy and/or demand savings
- are cost-effective and provide net benefits to society as demonstrated through the use of the cost effectiveness tests (**Technical Guide 2: Cost-Effectiveness Guidelines**)
- will satisfy other program objectives

Free ridership rates and other net-to-gross adjustment factors are not taken into account in MALs. Such factors are a function of program design and operation and must be determined and accounted for on a regular basis through program evaluation research. In the absence of better information, broad adjustment factor assumptions may be used for program planning and/or portfolio management purposes. But, any free ridership or other net-to-gross adjustment factors should be addressed by the evaluation and program input assumptions and revised as information is gained. These factors are discussed in *Net-to-Gross Adjustment Guidelines*.

## Understanding & Using MAls

All parties involved in the planning, design, implementation and evaluation of resource acquisition CDM programs should be familiar with how MALs are used. When using input assumptions, either those included in MALs or that should be included in MALs, it is important to:

- Understand assumptions and processes used to develop the MALs
- Know of existing measure input assumptions
- Know of, or be able to locate, recent evaluations of comparable programs and assessments of similar technologies
- Have the technical ability to undertake a practical review of measure assumptions, if required
- Understand the need to substantiate measure assumptions and updates

MALs are typically approved by a regulatory board, commission, or authority that is accountable for ensuring that CDM program investments are cost effective and produce real savings.

## Methods of Reviewing Input Assumptions

An input assumptions review is usually one of the first steps in developing a CDM program Evaluation Plan (**Step 7: Evaluation Plan Development Guidelines**). Reviewing input assumption may also be a part of a planning project-level measurement and verification (M&V) activities (**Technical Guide 4: Project-Level Energy Savings Guidelines**) to establish measurement techniques and procedures for calculating savings derived from projects.

Input assumptions for measures included in a typical MAL are:

- Description of the efficient technology
- Description of baseline technology (that is the technology that the efficient technology is replacing)
- Annual energy and demand savings
- Demand savings coincident with summer and/or winter system peak
- Seasonal energy savings patterns
- Effective useful life of the efficient technology (persistence)
- Incremental efficiency technology costs

Caution is required when using a MAL or measure assumption developed for use in other jurisdictions, especially where there are different codes, standards or market conditions. In all cases, the source of the assumptions for measures should be documented.

To provide an appropriate level of confidence in the MAL, periodic reviews of all underlying measure input assumptions are completed by independent research and through program evaluation activities. Any assumption update is based on the best available information.

Where insufficient data exists to complete an update to an assumption, the evaluation should use M&V to verify or re-estimate the assumption. New measure assumptions should be substantiated using literature reviews, program evaluations, case studies or third party testing, verification, or certification relating to the specific measure being investigated.

## Documentation and Reporting

The Evaluator will list the measures covered in the review, the results of the literature search, methods used to identify uncertainties, and methods used to estimate the range of savings specific to the measures in the program.

## Updating the Measures and Assumptions List

The OPA has an open, transparent, and flexible approach for reviewing and maintaining its MALs. Any stakeholder can submit new measures, measure revisions, or other measure considerations.

All requested updates/submissions related to MALs require verification. OPA staff use a standardized Measures and Assumptions Substantiation Form.

## Review of Measures and Assumptions List Update Requests

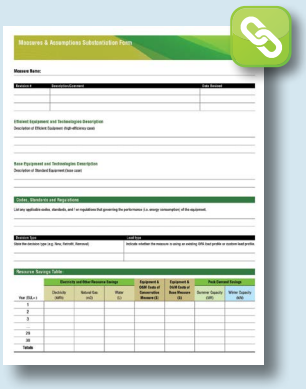
The submissions are reviewed based on the merits of the information provided. Following the review, submissions are either accepted as submitted, accepted with modifications, or rejected on specified grounds.

The review process time and approval is primarily dependent on the quality (relevancy and credibility) of the information provided to the OPA. Information referred to in substantiating the request must be available to, and accessible by, the OPA.

The OPA strongly encourages the inclusion in the submission an hourly (8760) annual load profile created from metered data or from a verified operating schedule. If unavailable, a description of the operating hours during weekdays and weekends for different seasons will be considered.

## The Measures and Assumptions Substantiation Form

The OPA form shows the information that is to be submitted when requesting an update of the OPA's Measures and Assumptions List. External stakeholders are encouraged to use the OPA form, or at least consider it as a guideline when making a submission.



**MEASURES AND ASSUMPTIONS SUBSTANTIATION FORM**

**Click the image to access the form.**

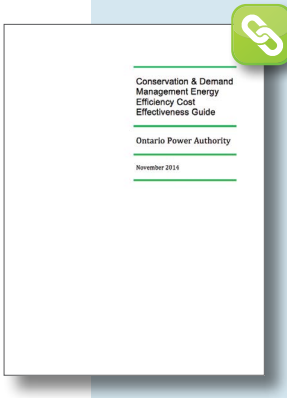
### Summary of Actions

- Consult the MALs to see whether the measures are in them
- Conduct an input assumption review
- Consider whether the correct confidence in values in the MAL
- Consider whether to submit update of MALs

# Technical Guide 2: Cost-Effectiveness Guidelines

## Key Points / Highlights

The *Conservation and Demand Management Cost Effectiveness Guide* sets out the cost-effectiveness policy articulated in the EM&V Protocols. Evaluation Administrators and Evaluation Contractors must follow the requirements of the guide.




### Cost-Effectiveness Guidelines

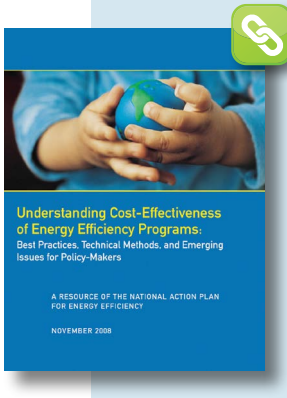
**Click the image to access the guide.**

This Cost Effectiveness Guide ("Guide") describes standard industry metrics to assess the cost effectiveness of conservation and demand management (CDM) resources. The Guide may be updated from time to time. Cost effectiveness assesses whether the benefits of an investment exceed the costs.

## Purpose

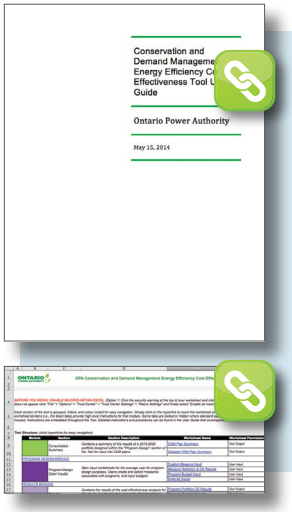
The purpose of the guide is to ensure program cost-effectiveness is considered by a broad range of stakeholders, including and but not limited to:

- Program Administrators
- Regulatory Agencies
- Administrative Agencies
- Policy Makers and
- Ratepayer Advocates




### Reference: Understanding Cost-Effectiveness of Energy Efficiency Programs

**A helpful tutorial on the common CDM-related cost tests can be found in the following document. Click the image to access the document.**




### CE Tool User Guide and CE Tool

These documents are intended to support OPA staff, LDC staff and other external service providers and/or delivery agents to calculate resource savings, budget and cost effectiveness metrics for new and existing conservation programs in Ontario.

## Summary of Actions

- Review the *Conservation and Demand Management Cost Effectiveness Guide*
- Ensure Evaluation Administrators and Evaluation Contractors follow the requirements of the guide to assess program cost-effectiveness.

# Technical Guide 3: Process Evaluation Guidelines

## Key Points / Highlights

A process evaluation is an empirical examination of program design, development, delivery, and administration. Such a systematic assessment of program elements, from resource allocation through program outputs, ensures program stakeholders that the planned offer is realized.

Process evaluations yield both qualitative and quantitative findings on which practical advice can be offered to enhance the program the design and administrative processes and the program service delivery. Unlike audits, process evaluations should provide evidence of outstanding practices and the means by which these practices can be transferred to other program delivery agents.

**Process evaluations gauge the effectiveness and appropriateness of the following:**

- **Program Design** – the linkage between key program elements, as well as the reasonableness of program objectives and resource allocation.
- **Program Development** – the protocols and procedures that form the basic offer to be implemented; the training and technical assistance provided to program delivery agents; and the changes made to the program design.
- **Program Administration** – the controls established for program delivery; the procurement processes for program goods and services; and the mechanisms in place to evolve the program offer.
- **Program Delivery** – the services provided by program agents; the processes used in the field to deliver the program offer; the systems used to track and monitor program outputs; the actual program expenditures over the assessment time horizon; the quality of measure installation; and the levels of participant satisfaction maintained throughout the offer.

## Collaborative Effort

Because of the need for collaboration among program delivery agents, contracted or external Program Managers, and the Program Administrator, process evaluations are complex. The Evaluation Administrator is responsible for fostering a cooperative relationship between the Evaluation Contractors who will be charged with carrying out the work and the program actors.

Experience has shown that attention to the following will help establish strong collaboration between program staff and the evaluation team:

- **Make introductions early:** The Evaluation Administrator should introduce themselves and the Evaluation Contractors to program staff as early as possible within the program development life cycle. Without early involvement, elements of program theory could be missed and the process evaluation could easily turn into, or be perceived as, a process audit.

- **Appreciate that program management and delivery staff are the experts.** Evaluation contractors are experts in assessment, not program operation. Only the program staff can offer the details needed to appreciate the available operational options and the choices made; without this expertise, the process evaluation cannot be developed and meaningful recommendations will not be identified. It is the Evaluation Administrator's responsibility to get the required information from program staff. (Information gathering is an essential competency of any process Evaluation Administrator.)
- **Recognize that observation affects operation.** It is important to remember that an effect cannot be measured without it being affected by the tool used to record the measurement. Process evaluations are a measurement of operational efficiency. As such, the presence of the Evaluation Contractor could affect the efficiency and efficacy of the process being assessed. Evaluation Administrators must be mindful of this when the Evaluation Contractor is formulating conclusions and recommendations.
- **Ensure findings are shared regularly.** After each field visit the Evaluation Contractor should share his/her findings with the Evaluation Administrator, who should then provide the information to the appropriate level of operational management. The responses offered by direct supervisors of those being observed will enlighten Evaluation Contractors about operation constraints and provide the basis for interpreting the evidence collected.

## Process Data Collection

Collecting process evaluation data is relatively straightforward. The evaluation of a process begins by answering the five questions: who, what, when, where, and how.

**What the Evaluation Contractor is looking for with respect to each question:**

| | |
|---|---|
| **Who?** | participant, service provider, Program Manager, etc. |
| **What?** | activity, materials, measures, behaviours, processes, etc. |
| **When?** | frequency, duration, size of interaction, etc. |
| **Where?** | home, office, internet, phone, etc. |
| **How?** | program policies, procedures, protocols, etc. |

Process data should be recorded for each program element or program activity identified within the program logic model (see **Figure 1.0: The Basic Elements of a Logic Model**). The Evaluation Contractor should be confirming whether expenditures match the program budget and if the expected outputs resulted from the activities observed.

The processes evaluated should be readily distinguishable from each other. The process assessment should focus on observable behaviours, the materials leveraged, and how program materials were received by participants.

Each process chosen for assessment should be looked at thoroughly. However, not all processes can be included in the process evaluation. The Evaluation Administrator and the Program Administrator should have already set into place the critical research questions to be answered and the Evaluation Contractor need only examine the processes that fit within the scope of the study.

## Process Evaluation Methods

Process evaluations consist of both quantitative and qualitative methods. Metrics for the quantitative assessment are often tracked by Program Administrators and program delivery agents within tracking systems and management reports. Qualitative data, on the other hand, must be observed or collected through survey/interview techniques.

In deciding who should collect the data, the Evaluation Administrator should balance cost and convenience against potential biases.

**The methods listed below are frequently used when assessing processes, though other techniques may be recommended and used by the Evaluation Contractor:**

- **Reviewing Field Notes:** These are brief records kept by program participants or delivery agents (typically recorded on forms). These forms may be part of the program delivery model or may be forms developed by the Evaluation Contractor. Examples of field notes include: activity logs, diaries, inspection notes, receipts, etc.
- **Creating a Case Study:** Case studies are created based on detailed records, often recorded by the Evaluation Contractor, of a small number of observed program activities.
- **Conducting Ethnographic Analyses:** This is a method of research that involves the Evaluation Contractor's direct observation of a program activity. This may include a "ride-along", which is where the Evaluation Contractor goes into the field with service providers and interacts directly with recipients of program measures and asks questions of program staff regarding their activity.
- **Conducting a Delphi Analysis:** This involves convening a panel of experts to explore a particular process or issue. The objective is to build a consensus opinion around the event or to forecast probable outcomes.
- **Conducting Focus Groups:** Focus groups are small group discussions, generally with the program participants and targeted market actors, aimed at learning about focus groups members experience with a product or service offering of the program.
- **Using Questionnaires:** Using surveys conducted via phone, mail, e-mail, Internet/online or through comment cards with respondents answering questions outlined based on pre-defined questions.
- **Conducting Unstructured Interviews:** This technique is used to elicit information in complex situations where program participation-related motivations are likely to be multi-faceted and behaviours influenced by multiple factors. Unstructured interviews also work well when there is no single decision-maker or the actual decision-maker is not easily determinable (for example, a large industrial customer with significant energy efficiency investment).

## The Process Evaluation Report

Keep in mind that the Process Evaluation Report can never be a compilation of all data recorded. Process evaluation reports should present summary data and should summarize important conclusions, as well as present recommendations based on the evaluation findings. Because there are many processes that get reviewed over the course of a program assessment and the scope of each assessment varies, there is no standard format for such reports. The contents and length of the report should be determined by what is most helpful to the Program Manager and by what meets the research requirements as defined by the Evaluation Administrator.

Determining what to include may not be easy since the Evaluation Administrator will look for detail while the Program Administrator likely wants only actionable items reported.

The Evaluation Administrator should work with the Program Administrator to define the types of information sought and ensure that the information and feedback is provided as quickly as possible and also included in the final process assessment.

### Summary of Actions

- Ensure strong collaboration between program staff and the evaluation team by setting stage for good relationships
- Choose processes for assessment, realizing that not all processes can be assessed
- Decide who should collect process data, balancing cost, convenience and biases
- Consider the appropriate methodology when undergoing the process evaluation
- Ensure Process Evaluation Report contains all that is necessary

# Technical Guide 4: Project-Level Energy Savings Guidelines

## Key Points / Highlights

The objective of measurement and verification (M&V) activities at the project-level is to confirm that energy efficient measures supported by CDM programs are installed and are yielding the desired impacts, such as energy and demand savings.

Two broad categories of projects are covered in this guideline:

- those with program-supplied "deemed" savings assumptions (prescriptive or quasi-prescriptive) and,
- custom projects, which are projects that require M&V to confirm savings.

**Energy efficient measures** (also referred to as "energy conservation measures" (ECMs)) are a single technology, operational change, or action implemented by a customer at the customer's site. Measures can be supported or promoted through a demand-side management program. A "project" can consist of a measure or a combination of measures that, together, are designed to conserve energy. Keep in mind that measures or projects can also be undertaken voluntarily by customers, but this guideline deals with activities that are directly supported by CDM programs.

This guideline assists Program Administrators, as well as program participants, in selecting approaches and methods for estimating energy and demand savings of projects. Results can also be used to support:

- Good energy management practices by program participants
- The determination of cost-effectiveness of projects

This guideline applies to resource acquisition demand-side management retrofits, new construction, and operational change programs that result in direct energy or demand savings at a project level. Programs that produce indirect savings, such as capability building or market transformation programs, are not covered by this guideline. For details on Behavioural Program guidelines, refer to **Technical Guide 12: Behavioural-Based Evaluation Protocols**.

A balance must be found between the needs of the Program Administrator and eventual evaluation requirements and the costs of M&V borne by both participants and the program. On the other hand, the basic reporting needed for the program and evaluation purposes generally overlaps with good basic energy management on the part of energy users.

Under optimal circumstances, the Program and Evaluation Administrators would provide final approval of the program-level plan for project-level M&V. The approval of individual M&V plans, in the context of the operation of the program itself, is within the purview of the Program Administrator (and is subject to evaluation).

At the program-level, it is common to conduct project M&V studies on a representative sampling of projects, particularly for mass market programs, and to extrapolate these findings to estimate aggregate impacts at the program-level. Some programs may require M&V on the full range of projects implemented under the program. Further guidance on estimating savings at the program-level is provided in **Technical Guide 5: Gross Energy Savings Guidelines.**

Projects not directly supported by the efficiency program that are undertaken voluntarily by customers as a result of the program's influence (for example, increased awareness of energy efficiency opportunities) are accounted for in estimates of program "spillover" or other effects (**Technical Guide 9: Net-to-Gross Adjustment Guidelines**). Note that some of these results may need to be sampled for measurement and verification also.

## Purpose and Scope of This Guideline

This guideline provides guidance for Program Administrators in selecting or, in some cases, prescribing evaluation methods to determine the energy savings from program-supported activities. The methods include:

- verifying the installation of energy efficient measures
- identifying factors that may affect prescriptive and quasi-prescriptive savings assumptions for measures
- improving the quality of prescriptive assumptions through technical reviews and,
- ensuring that an appropriate level of rigour is applied to M&V activities.

The Program Administrator is responsible for ensuring that the program design accommodates the need for any post-installation interaction with participants to facilitate project M&V. The Program Administrator also tracks program activity data and ensures that this information is available for the Evaluation Administrator in a usable format. Further, the Program Administrator may have to arrange for meetings or site visits to enable project M&V activities and then EM&V follow-up.

The Evaluation Administrator is responsible for providing oversight in the development of requirements for project M&V during evaluation planning (**Step 7: Evaluation Plan Development Guidelines**). The evaluation plan identifies which program-supported measures or projects will produce savings derived from prescriptive assumptions or through custom M&V methods. The evaluation plan also outlines the methods by which measure installations will be verified, as well as details regarding sampling strategies, data collection and analysis, and documentation of variances in baseline assumptions observed on site. Further, if required, the Evaluation Administrator can provide a technical review of assumptions or savings and, where appropriate, can recalculate the assumption in accordance with approved methodologies.

An Evaluation Contractor needs the following:

- Working knowledge of the International Performance Measurement and Verification Protocol (IPMVP) for energy efficiency projects
- Knowledge of measure-level assumptions (MAL) and use of measures and assumptions lists for prescribed savings (**Technical Guide 1: Using Measures and Assumptions Lists**)
- Knowledge of statistics and sample design methodologies to provide the desired levels of precision and confidence regarding the results
- Familiarity with ASHRAE or other guidelines for the measurement of technology-specific savings and,
- Certified Measurement and Verification Professional (CMVP) status is also highly desirable.

## Methods Applied In Project-Level M&V

The following section outlines methods that are often used Evaluation Contractor in Project-Level M&V:

### Review of Input Assumptions

If the prescriptive assumptions used as program inputs are new, are based on dated research or technologies, or are otherwise considered to be uncertain, a detailed review of the assumptions should be conducted. This review may occur during program planning and design, or during the program evaluation. Subsequent reviews of prescriptive assumptions are typically undertaken at least once every three years.

Detailed reviews or updates of prescriptive assumptions may also be triggered by changes in codes, standards and regulations, or by the natural introduction of more efficient products in the marketplace. A cursory review of all program input assumptions derived from the approved MALs (**Technical Guide 1: Using Measures and Assumptions Lists**) should help determine whether any major changes have occurred since the last detailed review.

When new and existing assumptions for a measure are under review as part of the evaluation or evaluation planning, the following should be considered for inclusion in the M&V study.

For **existing measures**, review input assumptions using:

- Billing, sub-metering, or engineering analyses on a sample of participants and non-participants
- Engineering calculations with M&V related to key assumptions
- Computer simulation models with M&V research related to key assumptions
- Calculations developed for quasi-prescriptive measures (for example, web-based applications) to compute savings based on customer-specific inputs

For **new measures**, determine input assumptions in advance of program implementation using information from:

- M&V study results from any relevant pilot projects
- Billing, sub-metering, and engineering analyses on a sample of potential participants
- Engineering calculations
- Computer simulation models
- Other quasi-prescriptive measure savings calculations, including ones developed by the OPA

During the process of verifying project-level savings, additional data can be collected on participant demographics, such as building or equipment operating characteristics or usage patterns. Further, findings from project-level studies can be used to substantiate differences between the baseline assumptions by improving information on efficiencies of replaced technologies, actual usage patterns, installation location, and so on, identified by participants and Evaluation Administrators.

## Criteria for Selecting M&V Methods

When selecting the methods to use in a project-level M&V it is important to first differentiate the type of project. Keep in mind that programs may involve a blend of several classes of project or may involve situations not contemplated in this guideline. The Protocols should therefore be interpreted as necessary to reflect the spirit of the concepts embodied in this document.

**Types of Projects**

1. **Prescriptive projects** – these are projects where prescribed or "deemed" savings values are derived from the approved MALs (**Technical Guide 1: Using Measures and Assumptions Lists**) with additional documentation and analysis to establish the number of installations.
2. **Custom projects – equipment retrofit only** – these are projects where efficiency gains are achieved by the retrofit or replacement of equipment, without changes in operations.
3. **Custom projects – operational change only** – these are projects where energy consumption (and possibly demand) are reduced by changing the operating periods, settings, or methods, without modifications to equipment.
4. **Custom projects – equipment retrofit and operational change** – these are projects where the combination of equipment and operational changes may impact load and energy separately or energy directly.
5. **Custom projects – multiple energy conservation measures (ECMs)** – these are projects where three or more ECMs are implemented at a single site or facility. Multiple ECM's may enable the use of whole facility metering to determine savings.

## Project Characteristics

Selection of the appropriate M&V method within any project type depends on a number of project characteristics. Five distinguishing characteristics can also be used to assist in selecting the M&V processes. These characteristics should be considered when developing M&V approaches for program-supported measures that do not exactly fit any of the basic project types described in this guideline.

### 1. Project Size

Project size may be based on:

- the incentive level (for example, dollars) for the particular energy conservation measure (ECM), per participant or for the whole program. When considering incentive levels:
  - small is under $10,000,
  - medium is from $10,000 up to and including $50,000 and,
  - large is greater than $50,000
- the participant's investment for the particular ECM, where:
  - small is under $10,000,
  - medium is from $10,000 up to and including $100,000 and,
  - large is greater than $100,000
- the savings (kWh or kW) expected by the participant for the particular measure(s) or project(s) installed.

The definitions of small, medium and large are intended as a guideline only. Program Administrators must provide definitions of the project size classes if these criteria are to be used as determinants of M&V methods.

### 2. Regularity of operating periods

Regularity of operating periods is a characteristic used where operating patterns are driven by routine events and the periods can be estimated with ease and accuracy. If operating periods vary irregularly because of variability in weather or plant production levels, precision must be applied when measuring the operating periods.

### 3. Persistence of savings

Persistence of savings is a characteristic used where the continuing success of the retrofit is uncertain (for example, control changes subject to human interaction). Note that it is inherently risky to base incentive payments and savings estimates on one-time observations. In these situations the reporting period should be extended and projects should be re-evaluated at least once.

### 4. Incentive base

Incentive base is a characteristic used when the basis for incentive payment is demand (kW). In such cases the analysis must consider the fraction of the equipment or the sub-system load that is normally operating when the site utility meter hits its monthly peaks ("diversity factor").[5] Energy savings (kWh) based incentives must consider the load of the equipment and normal annual operating hours.

### 5. Size of savings relative to utility meter total use

Size of savings relative to utility meter total use is a characteristic used where expected savings are small compared to total usage recorded on a meter; sub-meters may need to be added so that savings can be identified with reasonable precision. Suitable accuracy of meters and/or sampling strategy to yield reasonable results. Statistical analysis may be needed to select meters and sample sizes that will yield appropriate precision and confidence in findings

The project characteristics are used to select appropriate M&V strategies from the following list:

- Using the Prescriptive Measures and Assumptions List (**Technical Guide 1: Using Measures and Assumptions Lists**)
- Conducting user survey or site investigation of the number of installations
- Carrying out site measurement by spot readings at representative times, or continuous readings through at least one full cycle of operations
- Estimating interactive effects between the energy efficiency measure and electricity uses not measured as part of the M&V
- Estimating diversity factors, or logging of load patterns and utility meter profiles at times of peak utility usage
- Reporting "Normalized Savings" (under long term "normal" conditions), rather than under actual conditions of the reporting period. Note that adjustments must be made to the baseline period and to the reporting period data to restate it under such normal conditions. The normal set of conditions is defined by each participant for its operations.
- Choosing the most appropriate IPMVP Option when retrofit isolation techniques are not suitable

## Methods for M&V on Prescriptive Measures/Projects

As noted, prescriptive measures/projects are defined as those for which energy or demand savings per item are contained in the MALs (**Technical Guide 1: Using Measures and Assumptions Lists**).

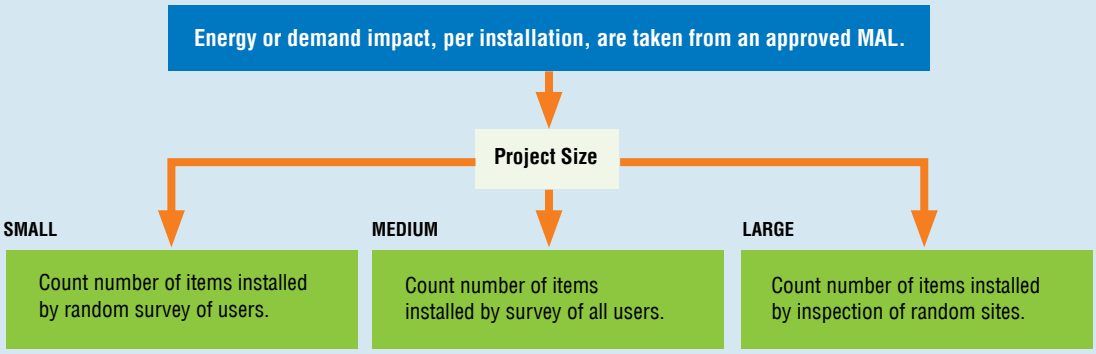
No field measurement is needed to determine the savings per measure or project. Gross impacts are determined by multiplying the per measure values derived from the measures and assumptions list by the number of installations.

The method of counting measures depends upon the size of the overall project:

**Methods used for determining the number of measures in a project generally depend on the size of the project:**

- **Small projects** – for small projects one can use participant self-reporting by questionnaire/survey of randomly selected participants
- **Medium projects** – for medium projects one can use participant self-reporting by questionnaire/survey of all participants
- **Large projects** – for large projects one can inspect randomly selected sites within homogeneous groups of all participant sites. Thus, achieving an overall precision of +/-20% at 90% confidence level

Figure 2.0
Prescriptive Projects



## Methods for M&V on Custom Projects

Four categories of custom projects are considered here:

- projects involving equipment retrofits only
- projects involving operational change
- projects involving equipment retrofit and operational change
- projects involving multiple energy conservation measures

Depending on factors like the amount of anticipated savings, project size, or the incentive amount, the guidance and flow charts that follow are intended to help with the selection of appropriate methodologies for completing M&V on a measure or project basis.

Keep in mind that M&V plans and their reported findings are used to verify that: measures have been installed; are working as planned; and are generating savings. These custom project savings can be assessed by:

- isolating the retrofit,
- measuring the whole facility, or
- using computer simulations.

Installations can be verified through a combination of site visits and participant surveys to ensure reported results match actual impacts.

## Methods for M&V on Equipment Retrofit Only Projects

These are custom projects involving only retrofit or the replacement of baseline equipment with more efficient equipment. In such projects no changes are made to operating periods, settings, or methods. If both retrofit and baseline equipment have load values shown in the MAL (**Technical Guide 1: Using Measures and Assumptions Lists**), these values are used for baseline and reporting period loads.

For equipment not on the MALs, one time measurement(s) must be made using meters of sufficient accuracy to allow the computed raw change in load to be reported with a precision of +/-10% and a confidence level of 90%.

If an incentive is being used to impact energy **demand** (see also **Technical Guide 6: Demand Savings Calculation Guidelines**), the way to undertake M&V depends on the project size:

1. For small projects, multiply the baseline and reporting period loads by an estimated diversity factor.

2. For medium or large projects:

- multiply baseline and reporting period loads by a diversity factor determined by recording the summer and/or winter demand profiles of the particular piece of equipment being retrofitted and the associated utility meter and,
- estimate the interactive effects of the retrofit beyond the boundary of measurement.

If a **consumption incentive** is being used, the change in load is multiplied by the normal operating period. Again, the way to undertake M&V depends on the project size:

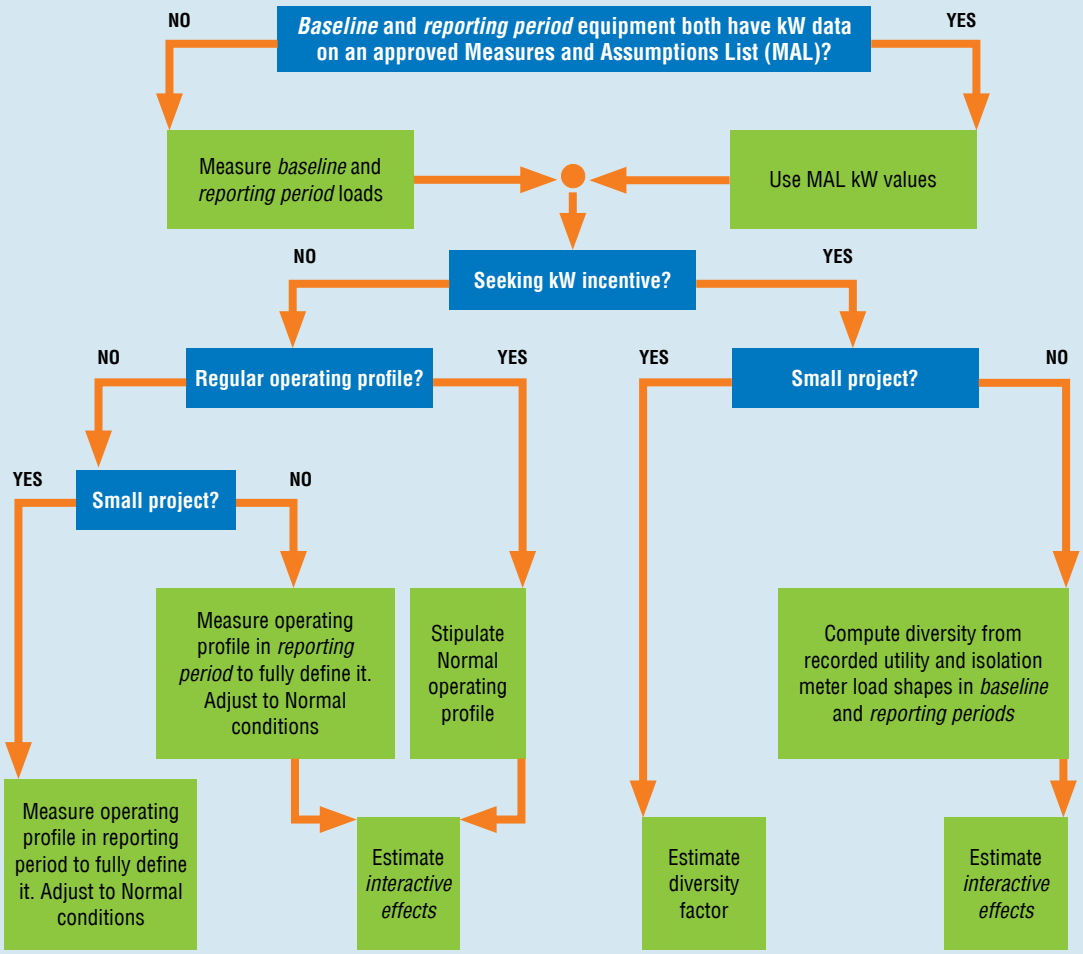
1. For small projects the normal operating period:

- may be assumed, where the operating profile of the equipment before and after retrofit is implemented or,
- Where the operating profile is not regular, M&V should be estimated from measurements taken at two separate points in time (at a minimum) representing the range of the normal operating pattern.

2. For medium or large projects:

- the normal operating period should be estimated from continuous measurement throughout the full range of governing conditions after the retrofit is carried out and,
- an estimate should be made of the interactive effects of the retrofit.

## Figure 3.0:
## Custom Projects: Equipment Retrofit Only



## Methods for M&V on Projects Involving Only Operational Change

Such projects are custom projects that involve only changing equipment operating periods, settings, or methods. No equipment replacements or retrofits are involved. If the equipment whose operation is being changed has load values on a published MAL, the values on the list may be used. Otherwise measure equipment load once with a wattmeter having a precision of +/-5% or better, at a confidence level of 90%.

If a **demand incentive** is being used (see also **Technical Guide 6: Demand Savings Calculation Guidelines**) the way to undertake M&V depends on the project size:

1. For small projects, a diversity factor must be separately estimated for both the baseline and reporting periods and adjusted to normal operating conditions.

2. For medium or large projects:
- determine separate diversity factors for both the baseline and reporting periods by recording the summer demand profiles of the particular piece of equipment and the associated utility meter and,
- estimate the interactive effects of the project, beyond the measurement boundary.

If a **consumption incentive** is being used, the equipment load is multiplied by the change in operating periods between baseline and reporting periods derived as described below:

The baseline period's operating period is determined as follows:

1. If the operating profile is regular, measure it once and project it to normal conditions;

2. Otherwise, if operating profile irregular:
- for small projects, measure the operating profile at two separate points in time representing the range of the normal operating pattern, being sure to adjust the operating profile to normal conditions.
- for medium or large projects, measure the operating profile continuously for one cycle and adjust it to the operating profile of normal conditions.

The reporting period's operating period is determined as follows:

1. If the operating profile is regular, or the project is small, measure the operating profile once and adjust it to normal conditions

2. Otherwise:
- for medium sized projects, measure the operating profile at two separate points in time representing the range of the normal operating pattern. Adjust the operating profile to normal conditions.
- for large projects, measure the operating profile continuously for one cycle and adjust it to an operating profile of normal conditions.
- for medium and large projects, estimate the impact of interactive effects beyond the measurement boundary.

**Figure 4.0**
**Custom Projects: Operational Change Only 1 demand (kW) incentive**

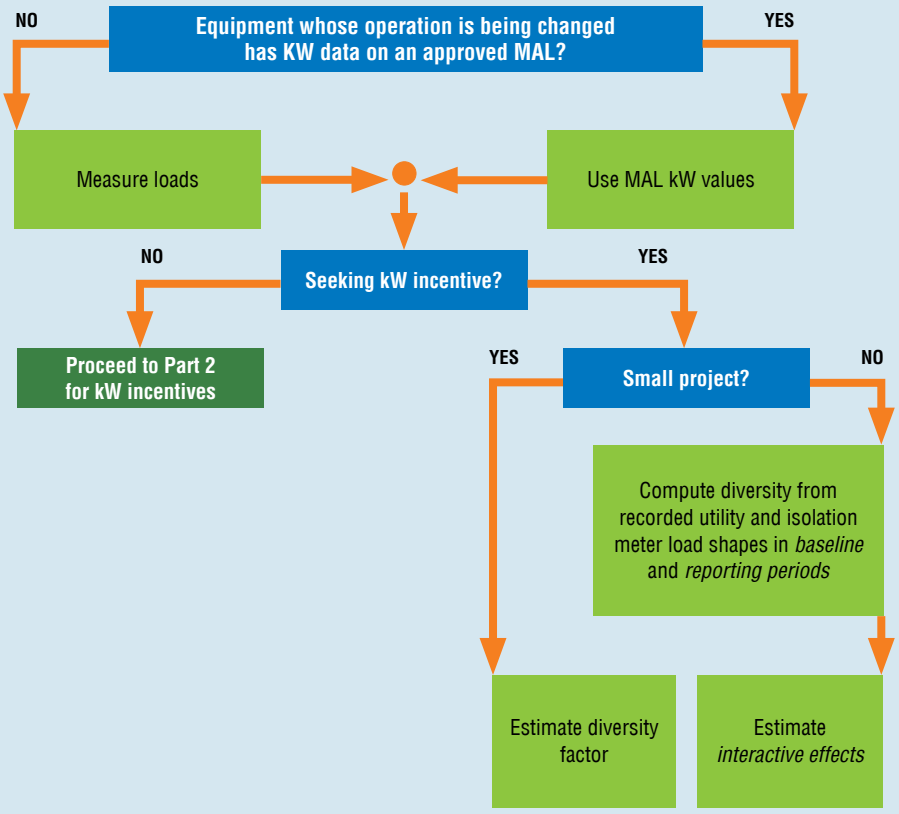
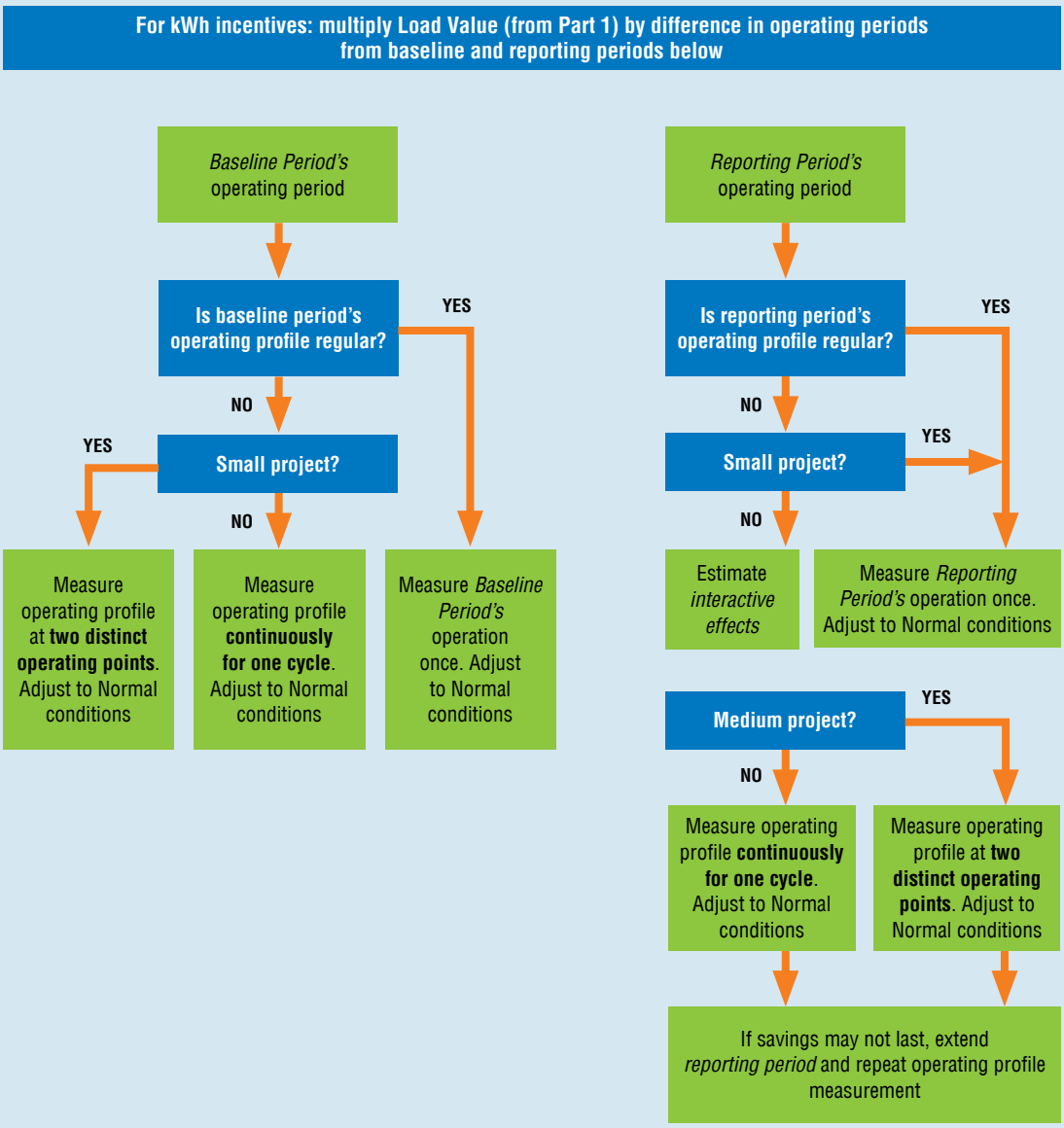


**Figure 5.0**
**Custom Projects: Operational Change Only 2 energy (kWh) incentives**



## Methods for M&V on Equipment Retrofit and Operational Change Projects

These projects are custom projects involving both the retrofit or replacement of baseline equipment and a change in operational periods, methods, or settings.

There are two ways to undertake M&V for such projects:

1. If savings are highly likely to continue over time, or the project is small in size:

- If both baseline and reporting period equipment have load values shown in a current published MAL, use the MAL values to determine the loads; or
- If both values are not in a MAL, take one time measurement(s) using meters that are of sufficient accuracy to allow the computed raw change in load to be reported with a precision of +/-10% and a confidence level of 90%.

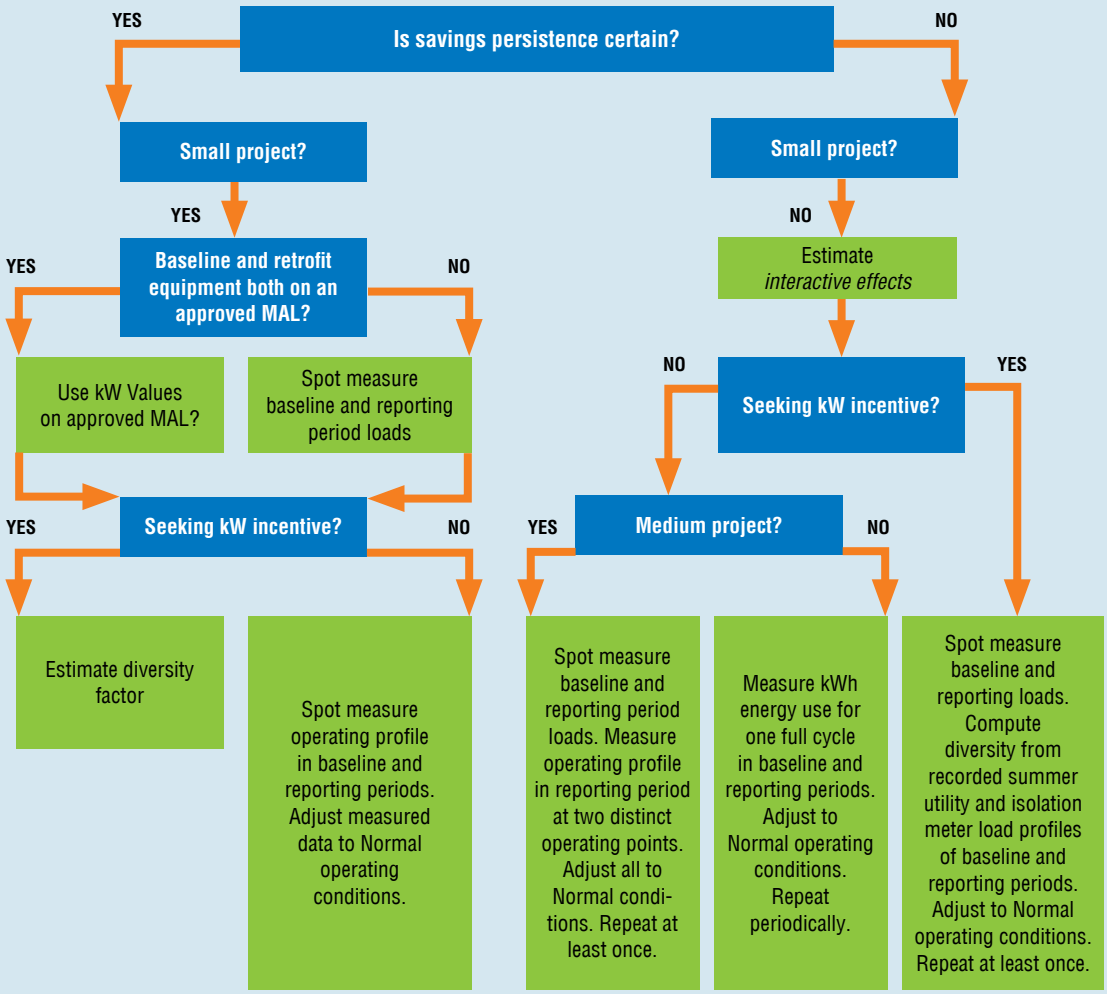
In either case:

- for kW incentives, estimate a diversity factor.
- for kWh incentives, measure operating profiles in baseline and reporting periods. Adjust all measured data to normal conditions.

2. If savings may not continue over time, or the project is medium or large in size:

- estimate interactive effects, and
- for kW incentives
  - take one time measurement(s) of baseline and reporting period loads using meters that are of sufficient accuracy to allow the computed raw change in load to be reported with a precision of +/-10% and a confidence level of 90%. Multiply the loads by diversity factors. Determine the diversity factors by recording the summer demand profiles of the particular piece of equipment being retrofitted and the associated utility meter. Repeat all reporting period measurement and recordings at least once.
- For kWh incentives:
  - for medium sized projects, take one time measurement(s) of baseline and reporting period loads using meters that are of sufficient accuracy to allow the computed raw change in load to be reported with a precision of +/-10% and a confidence level of 90%. Measure operating profiles at two distinct operating points. Adjust all data to normal conditions. Repeat reporting period measurements at least once.
  - for large sized projects, measure energy use for one full cycle of operations in baseline and reporting periods. Adjust all data to normal conditions. Repeat reporting period measurements periodically.

## Figure 6.0
## Custom Projects: Equipment Retrofit and Operational Changes



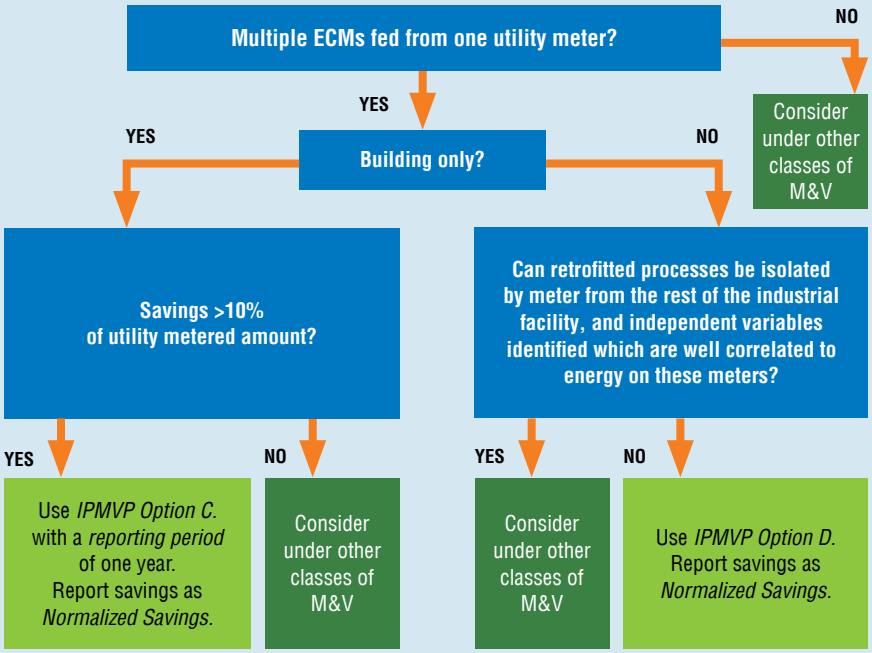
## Methods for M&V on Multiple ECMs or "Blended" Projects

These projects consist of more than one energy efficiency measure. For these custom projects, special approaches can be used in certain circumstances to manage the M&V costs.

1. For buildings, where total expected savings of all ECM's is 10% or more of the affected building's consumption or demand as recorded on the utility meter, use IPMVP Option C Whole Facility . The reporting period should be one year. Savings should be reported under normal conditions ("Normalized Savings").
2. For industrial processes, where the ECMs cannot be isolated by an energy meter or reasonably correlated to independent variables related to production, use IPMVP Option D ("Calibrated Simulation"). This situation is likely to arise where multiple ECMs are installed in complex, integrated process plants. The plant must have a software-based hourly simulation model of the details of plant operations and energy use and a meter that records hourly energy for the portion of the plant being simulated. The software's calculation of energy use must be calibrated against actual hourly metered energy use. Such calibration must present a coefficient of variation of the root mean squared error (CVRMSE) of less than 30%. Savings should be reported under normal conditions ("Normalized Savings").

Site visits are required to verify measure installations for both prescriptive and custom measures. Participants included in this process are chosen using a sampling methodology with the objective of providing confidence and precision levels that meet a 90/10 threshold (evaluation results that are within 10% of the actual result in 90% of cases). The evaluation team should also use the site visits to identify and document any variances in baseline conditions observed on site.

**Figure 7.0**
**Custom Projects: Multiple Energy Conservation Measures (ECMs)**



## Documentation And Reporting on Project-Level Energy Savings Assessments

The M&V report should contain sections and complete descriptions of the processes used and results for all required elements in the M&V plan, including:

- Goals and Objectives of the M&V Plan
- Sampling Plan used to select buildings/participants for examination, including physical and occupancy characteristics of the buildings visited or details (for example, regional) of the participants included
- Description of data collection and analysis procedures
- Estimated accuracy level of proposed assumption
- Verification and data quality procedures used to test the tracking systems
- Summary of the results and discussion of any variances or unexpected findings when the results were compared to the targets
- Documentation of the technical analysis or computer aided assumption reviews undertaken and the associated findings
- Recommendations for how the results should be used to adjust prescriptive/quasi-prescriptive or custom project savings
- Overall summary recommendations for improvement of process of the program in future years

In all cases, the Evaluation Administrator will provide final sign-off on the M&V report and its associated findings.

## Timing of Analysis of Underlying Assumptions

To ensure the viability of the measures included in a program and to ensure their corresponding cost-effectiveness. The analysis of the underlining assumptions used to assess energy and peak demand savings is most beneficial when determined before program launch. However, if this is not feasible (for example, when billing analysis or meter reading is required) these assessments should take place within an appropriate period of time after program launch and the results should be used to update program impact forecasts.

The Program Administrator decides whether a third-party review of prescriptive assumptions. The decision is based on a variety of factors including:

1. Previous independent review(s) of the input estimates
2. The expected magnitude of the programs savings
3. Third-party or intervener concerns over assumptions
4. Issues uncovered during literature review (**Technical Guide 1: Using Measures and Assumptions Lists**)
5. Scope of, and budget for, the evaluation

If an assessment is required, an Evaluation Manager investigates the current assumptions to either verify or re-estimate and the key inputs. Where appropriate, findings from this process should be used to update applicable measure assumptions (**Technical Guide 1: Using Measures and Assumptions Lists**).

### Summary of Actions

- Review the input assumptions
- Select the project level M&V method based on the type of project and project characteristics
- Ensure M&V report contains necessary descriptions of processes used and results for all requirements
- Decide whether third-party review of assumptions is required

# Technical Guide 5: Gross Energy Savings Guidelines

## Key Points / Highlights

Energy savings as a direct result of CDM program activities is a key element to the establishment of energy efficiency as a reliable system resource.

## Purpose and Scope of This Guideline

This guideline provides information about methods that can be used in CDM program evaluations to develop accurate estimates of the energy savings resulting from program activities. The goal is to produce energy savings estimates that are accurate within reasonable levels of precision and confidence (in most cases within 10% of the actual result at a 90% confidence level).

This guideline applies to all CDM programs that have the objective of producing direct energy savings (that is, resource acquisition programs).

The guide expands on the information set out in **Technical Guide 4: Project-Level Energy Savings Guidelines**. Therefore, where possible, we recommended that the same evaluation team perform or provide oversight for the requirements relating to both guidelines.

An Evaluation Administrator is typically responsible for fulfilling the requirements of this guideline through an approved evaluation plan (see **Step 7: Evaluation Plan Development Guidelines**). The evaluation plan details the methods for assessing program-specific energy savings. The Evaluation Administrator also provides the rationale for why the selected methodology has been chosen from the list of approved methodologies or why an alternative method has been proposed. The details of the methodological choices are usually developed in collaboration with the Evaluation Contractor.

Gross savings calculations are based on the difference between energy use and/or demand after the implementation of a program and an assumed set of baseline conditions that estimate what energy consumption and/or demand would have been in the absence of the program. Because there is no way to measure something that did not occur in the first place, there is no direct way to measure gross savings.

Gross savings are not discounted for free ridership or other adjustment/distortion factors (net savings).

This guideline pertains only to estimates of energy (GWh) savings. Demand (MW) savings are covered in **Technical Guide 6: Demand Savings Calculation Guidelines** and net savings are covered in **Technical Guide 8: Net-to-Gross Adjustment Guidelines**.

**Gross energy savings** is the change in energy consumption that results directly from program-related actions program participants take, regardless of the reasons why they participated.

Evaluation Administrators should have the following skills:

- The ability to applying statistical and sample design methodologies
- Ability to calculate, using all relevant adjustment factors, program-specific cost benefit analysis (for example, total resource cost test)
- Strong research skill
- Practical abilities related to technically reviewing input assumptions.

## Selecting an Approach

There are three general approaches for estimating gross savings:

- Deemed savings,
- Large-scale data analysis, and
- Custom M&V.

When choosing the methodology, the following factors should be taken into consideration:

- The program implementation strategy and the types of data that can be collected during the course of program delivery
- The types of measure(s) supported by the program (for example, simple, mass market versus complex, commercial or industrial measures)
- The perceived accuracy of previous evaluations or assumptions, such as those identified in the MALs (**Technical Guide 1: Using Measures and Assumptions Lists**).
- The amount of energy savings expected to result from the program
- The professional judgement of the Evaluation Administrator
- Time and budget available for the evaluation

## Basic Terms and Concepts

If one cannot measure the absence of energy use (savings), as noted, there is no way to directly measure gross energy savings. Energy savings can be estimated by comparing energy use before and after a CDM program is implemented. Equation 1 shows the general formula that applies when calculating energy savings for all energy efficiency programs.

**Equation 1**

**Energy savings** =

baseline energy use – reporting period energy use +/- adjustments

Where:

- **Baseline energy use** is the energy consumption that is estimated to have occurred before the program was implemented. The baseline period is selected to be representative of normal operations.
- **Reporting period energy use** is the energy consumption that occurs after the program is implemented.
- **Adjustments** account for independent variables that are beyond the program implementer or participants control. Adjustments are meant to bring the baseline and reporting periods to the same set of conditions (rather than a simple subtraction of pre- and post-installation energy use). Common independent variables that are adjusted for, include:
  - Weather normalization
  - Occupancy levels and hours (i.e. hours of operations)
  - Production levels (ie. operating cycles, shifts)

## 1. Deemed Savings Approaches

Deemed savings approaches uses agreed upon values for program-supported measures with well-known and documented savings values. Deemed savings are determined using prescriptive and quasi-prescriptive assumptions and standard equations for determining gross savings. Applying deemed savings values to individual measures is addressed in **Technical Guide 4: Project-Level Energy Savings Guidelines**.

For prescriptive and quasi-prescriptive measures, the savings evaluation depends on:

- The technology type
- The number of installations
- The prescribed savings estimates for the technology used

For quasi-prescriptive measures, the savings evaluation depends on:

- Project-specific information generally collected from participants implementing the measures (for example, savings per unit capacity or per hour of operation)
- Other information needed to adjust savings estimates (scalable basis)

For documentation and data collection purposes additional information that should be collected during the evaluation include:

- Customer address or location
- Information on technology being replaced or retrofitted
- Information about operation of new equipment (for example, hours of operation)

### Prescriptive Approach Saving Calculations

Savings are prescribed on a per-participant or per-measure basis and represent an average level of savings that would be achieved by a participant implementing the energy efficient measure. Gross savings are calculated based on the number of participants and/or measures installed multiplied by the prescribed savings per participant or measure. The gross savings are calculated as shown in Equation 2.

**Equation 2**

$$PS_{gross} = N$$

where,

$PS_{gross}$ = Gross program savings (e.g., kWh)
N = Number of tracked participants (or measures installed)
s = Prescribed savings per participant or per measure (e.g. kWh per participant)

### Quasi-prescriptive Approach Saving Calculation

Savings are determined using a prescribed methodology that uses key, project-specific, inputs to estimate the savings for each participant or measure installed. A common quasi-prescriptive methodology is to prescribe energy savings for a measure on a scalable basis (for example, kWh savings per unit of capacity or per hour of operation). If the relationship between the scalable bases and the savings is linear, then gross program savings can be calculated from the number of participants or measures installed multiplied by the average participant value of the scalable basis multiplied by the prescribed scalable savings. The gross program savings are calculated as shown in Equation 3.

**Equation 3**

$$PS_{gross} = N \times SB_{avg} \times s_{scale}$$

where,

$PS_{gross}$ = Gross program savings (e.g., kWh)
N = Number of tracked participants (or measures installed)
$SB_{avg}$ = Scalable basis (e.g., average participant equipment capacity)
$s_{scale}$ = Prescribed savings per participant or measure (e.g., kWh per participant per scalable basis)

Other potential quasi-prescriptive approaches may, as an example, include engineering equations that utilize key participant inputs, prescribed inputs, or default values, to estimate savings estimates or use similar inputs to reference MALs. In these instances, as shown in Equation 4, that the gross program savings are calculated from the sum of the savings calculated for each participant or measure installed.

**Equation 4**

$$PS_{gross} = \sum_{i=0}^{N} (ps_i)$$

where,

$PS_{gross}$ = Gross program savings (e.g., kWh)

N = Number of tracked participants (or measures installed)

$ps_i$ = Savings reported for the $i^{th}$ participant using the quasi-prescriptive methodology

## 2. Large-Scale Data Analysis Approach

Large-scale data analysis applies a variety of statistical methods to measured facility energy consumption meter data (almost always whole-facility utility meter billing data) and independent variable data to estimate gross energy and demand impacts.[1] Meter analysis approach usually involves analysis of a census of project sites, versus a sample.

Most analyses of meter data involve the use of comparison groups. "Quasi-experimental design" has traditionally been used in assessing the impacts of programs. They compare the behavior of the participants to that of a similar group of non-participants – the comparison group – to estimate what would have happened in the absence of the program.

There are three basic large-scale meter data analysis methods employed for energy efficiency programs:

- **Time series comparison** - compares the program participants' energy use before and after their projects are installed. With this method the "comparison group" is the participants' pre-project consumption.
- **Use of comparison group** - compares the program participants' energy use after projects are installed with the energy use of non-participants. This method is used primarily for new construction programs, where there are no baseline data.

## 3. Custom M&V Approaches

Custom M&V approaches are used when no prescribed measures are found on the MALs for the types of measures included in a program. Custom M&V approaches require that gross savings be tracked and estimated on a project-by-project basis. Custom projects tend to be more complex than those using prescriptive measures (for example, building equipment retrofits where equipment load profiles are variable, etc.) and gross savings estimates use specific equations that can change on a project-by-project basis. Therefore, project-level M&V is essential for tracking and reporting savings and should at least be taken into consideration for all situations requiring a custom M&V approach (see **Technical Guide 4: Project-Level Energy Savings Guidelines**).

1 National Action Plan for Energy Efficiency (2008). Understanding Cost-Effectiveness of Energy Efficiency Programs: Best Practices, Technical Methods, and Emerging Issues for Policy-Makers. Energy and Environmental Economics, Inc. and Regulatory Assistance Project. <www.epa.gov/eeactionplan>

For custom M&V approach evaluations, evaluators will need to collect the following information:

- Type(s) of equipment installed
- Type(s) of equipment being replaced
- Customer address or location
- Engineering analyses and/or computer simulations
- Other information needed to determine savings for custom projects

M&V activities consist of some or all of the following:

- Meter installation, calibration and maintenance
- Data gathering and screening
- Development of a computation method and acceptable estimates
- Computations with measured data
- Reporting, quality assurance, and third party verification of reports

At the project-level, the approach is typically outlined in an M&V plan that should be developed before project implementation. Programs that support custom measures are typically targeted to larger customers and are likely to involve fewer projects.

Gross savings can be determined by:

- Selecting a representative sample of projects for review
- Determining the savings generated by each project in the sample using one of the options described in the International Performance Measurement and Verification Protocol (IPMVP) and guidance provided in **Technical Guide 4: Project-Level Energy Savings Guidelines**.
- Applying the savings from the sample of projects to the entire population of projects

## Documentation and Reporting Gross Energy Savings

A final evaluation report related to gross energy savings should include details as to how gross savings were determined. The final report should include information about:

- Methodology or methodologies used to assess gross savings
- Sampling plans and survey instruments used to collect data
- Precision and confidence of data and results
- Total gross savings and sample calculations
- Explanations, where possible, of variances between verified results and forecasted results for the program

The Evaluation Administrator reviews the final estimate of savings demonstrated through the study, which is provided by the Evaluation Contractor.

## Timing of Gross Energy Savings Calculation

Completing a program-level estimate of gross savings takes time, the amount of which will depend on the analytical approaches selected and whether it will be necessary to gather and model a full range of data to complete the analysis (for example, 12 months of pre- and post-implementation electricity bills or one or more full operational cycles). The choice of the data collection period should be an explicit issue identified in the program evaluation plan (**Step 7: Evaluation Plan Development Guidelines**), as it relates to how frequently the calculation is made. Results should be reported in a timely manner to support the objectives of the evaluation.

## Oversight and Responsible Parties

The Evaluation Administrator approves the gross savings methodologies used and is accountable for ensuring the analysis is completed on schedule by the Evaluation Contractor. The analyses are typically carried-out by the Evaluation Contractor and reviewed by the Evaluation Administrator. A broader evaluation team may be part of the review process. It is essential that the Program Administrator establishes a tracking system to facilitate this analysis and provides the Evaluation Administrator and Evaluation Contractor with all requested tracking system outputs and/or read-only access to the tracking system itself.

### Summary of Actions

- Select an approach for estimating gross savings
- Ensure the appropriate equation is used for calculating gross savings
- Ensure the final evaluation report includes details of how gross savings were determined

# Technical Guide 6: Demand Savings Calculation Guidelines

## Key Points / Highlights

The Demand Savings Calculation Guideline establishes the framework for assessing demand savings attributable to specific conservation initiatives.

This guideline applies with regard to all energy efficiency programs designed to achieve energy or peak demand savings (Demand Response programs have a separate procedure).

The Evaluation Contractor is responsible for finalizing the methods used to estimate net demand savings for the program.

The Evaluation Administrator is responsible for reviewing the Evaluation Contractor's proposed plan for calculating demand savings and for signing off on that plan.

The Evaluation Contractor needs the following skills:

- Proficiency with statistical and sample design methodologies
- Familiarity with load shape analysis principles and assumptions
- Market research capabilities
- Technical ability conducive to the understanding of the operational functionality of efficiency measures (for example, peak demand effects)
- Ability to use models to forecast energy usage and ability to translate data into end-use and sector-level load shapes

## Definition of Peak

The concept of peak demand is not simply the highest demand for electricity in a 24 hour period. Instead, the concept relates to energy demanded over the course of pre-defined period of time (i.e., 1 pm-7 pm) during which the overall demand on the province's electricity grid tends to be higher, on average. So, the first step in determining peak demand (and peak demand savings) is determining the pre-defined blocks of hours during which demand is generally at its highest.

In order to maintain consistency from the program design and approvals stage, through to program operations and reporting, and finally to EM&V and verified savings, we use a before the fact (ex ante) definition of peak. Actual (ex post) system demand data is not used for the purposes of defining system peak, (it can, however be used as a reference to ensure that, over time, the ex ante definition of peak is valid.)

The hours that count towards savings targets should be known in advance and remain constant for the full program cycle. It is possible that actual system conditions will vary to a small extent over the framework period.

Though more accurate for in-year savings calculation purposes, normalized system forecasts are used to develop blocks of hours that ensure an extremely high likelihood that the top hour or top-10 hours of system peak will occur within the block(s). Benefits from the clarity and predictability of a block definition include: (a)better ability to track progress-to-target while a program is in-market and (b) greater likelihood that the Program Administrator and participants will comprehend the connection between various measures under consideration and the value they provide to the system (the basis for the cost-effectiveness of the programs).

**Table 1.0**

**OPA EM&V Standard Definition of Peak for Calculating Demand Savings**

Based on analysis of Ontario System Hourly Load data from 2003-2010, the defined summer and winter peak blocks for 2015-2020 are as follows:

**Average Load Reduction over Entire Block of Hours**

| | Time | Months |
|---|---|---|
| SUMMER (Weekdays) | 1pm - 7pm* | June |
| | | July |
| | | August |
| WINTER (Weekdays) | 6pm - 8pm | January |
| | | February |
| | | December |

*Daylight Savings Time-Adjusted

Because of Ontario's unique geography (vast distance from north to south and mid-latitude, full four season climate) and load characteristics, the system peak could occur in either season. Though summer peak has been dominant in recent years, it is not predicted to continue and there is a chance that the system will experience a winter peak.

## Declaration of Peak Savings

Since both summer and winter peak savings have the potential to contribute to reducing the Ontario system peak, Evaluation Administrators should calculate both peaks. For example, automobile block heaters and space cooling/air conditioning provide straightforward examples of winter and summer peak-affecting measures (or initiatives or programs). Street lighting, though used all year around, would be highly coincident with the winter peak block period, but not at all with the summer block. Some measures or programs may be equally suitable for both blocks, so the selection of which one is not particularly important.

Note however, that savings for measures/programs that contribute to both block periods are not double-counted towards system peak. A declaration of the period that savings should be counted towards should accompany program funding approval. Peak demand savings results tracking and program evaluation then flow from that declaration.

Ontario also straddles summer and winter peak in terms of various parts of the province. Depending on the regions, some LDCs may remain summer peaking (for the foreseeable future) and northerly LDCs could remain winter peaking, despite the fact that the Ontario system peak could occur in either of the seasons. An LDC's deployment of summer or winter peak demand reduction initiatives is *not* dependent on its service territory's peaking characteristic, but rather on the initiative's design to target a reduction in either summer or winter peak consumption.

Evaluation Administrators are encouraged to use the standard definition of peak described in **Table 1.0**, since it is the definition that will be used for verified savings calculation and

reporting purposes. Program administrators who choose to use a definition(s) of peak that varies from this one would be advised to employ a methodology to assess the gap between reported program savings and verified/evaluated savings. This gap should be predictable. In other words, a known risk factor that contributes to a gap between reported savings and eventual verified savings should be analyzed and documented so that there are no surprises at the end of the process.

## Estimating Demand Savings during the Peak Period

Peak savings estimates are to be based on the **average demand reduction across the total number of hours** in the appropriate peak summer or winter block (refer to **Table 1.0: Definition of Peak for Calculating Demand Savings**) for block definitions). Note that because impacts across the total number of hours in each block are averaged, the peak blocks for the summer and winter do not comprise the same number of hours. Technologies that provide sustained demand reductions across the entire block have more value to the system than those that are variable. This is by design, since the chance of the actual (ex post) peak occurring in one hour versus another within the defined blocks is roughly equal. Therefore, measures or programs that better sustain savings across the span of the defined block have more value to the electricity system than those that provide a more limited sustained impact.

Maximum monthly demand reduction, typically described as "at design conditions" and/or the top facility hour of the month, in each of the three months (instead of the average of the entire block of hours) is for weather-sensitive measures because their load impact characteristics improve coincident with the system peak, since it is also weather-sensitive. For non-weather sensitive measures, using average impacts ensures that variable impacts are properly accounted for. But weather-sensitive measures are highly likely to produce their maximum impact at the same hour that was the actual top system peak hour (either summer or winter, depending on the measure). Weather-sensitive measures can therefore be properly accounted for their performance relative to periods of electricity system stress by using a much narrower – 3 individual hour in this case – definition of peak.

Other variable loads may also use this approach. Since the weighted average is structured to have no bias (advantage or disadvantage), Program Administrators and those managing M&V plans should feel free to compare and use this alternative approach. If the peak demand

**Table 2.0**

**Alternate Definition of Peak**

An alternative method can be used to calculate peak demand savings for weather-sensitive measures or for facilities with variable load characteristics. Peak demand savings are calculated on the basis of a weighted average of the maximum demand reduction in each of the three months that occurs within the blocks:

**Weighted Average of the Monthly Maximum Load Reduction****

| | Time | Months | Weighting[3] |
|---|---|---|---|
| SUMMER (Weekdays) | 1pm - 7pm* | June | 30% |
| | | July | 39% |
| | | August | 31% |
| WINTER (Weekdays) | 6pm - 8pm | January | 65% |
| | | February | 16% |
| | | December | 19% |

*Daylight Savings Time-Adjusted

**Typically implemented as "at design conditions" and/or for the top facility hour of the month

[3]Weighting is based on the proportion of Top-10 hours that occur in that month

savings credited are higher using one approach versus the other, one should use the approach that produces the higher impact. The higher impact should be used not simply because it is higher – it should be used because it will produce a more accurate assessment of the peak demand savings. For the purposes of preparing verified savings estimates, Evaluation Contractors should promote the method they believe produces the highest confidence result regardless of which approach was taken by Program Administrators.

Please also note that for demand response initiatives that are able to be "called" by the IESO or that operate during real electricity system stress periods, a different process is used to calculate impact. The process discussed here is for energy efficiency initiatives.

### Direct Methods for Computing Peak Demand

1. Collect hourly energy use data from a sample of participants before and after the measure installations, providing an estimate of the peak demand reduction performance of a specific measure
2. Collect hourly energy use data from a sample of locations where the efficiency measure has been installed and compare it to corresponding representative non-participant locations and use the variance to estimate the impacts on peak demand.

### Indirect Methods for Computing Peak Demand

1. Allocate annual energy savings into one or more time of use periods using secondary data on average end use load shapes from past OPA evaluation results, forecasting models, or other relevant studies. Average demand savings can then be determined by dividing the energy use savings allocated to that period by the number of hours in that period.
2. Using the results of energy simulation models, allocate daily or annual energy savings for a measure or set of measures into time of use periods.
3. Estimate total peak savings for prescriptive measures installed based on the per measure values in the most recent MALs.

## Valuation of Peak Demand Reduction

Since the cost-effectiveness of CDM program activity is premised on the avoided cost of generation, the power plant that would theoretically get built in Ontario to serve the marginal peak demand and energy that is being saved by the programs operates in both our summer and winter constrained periods. Setting aside some complexities of winter versus summer system capacity constraint characteristics, heat rates and other technical issues, the same "avoided cost dollars" build the same "peaker" plant that might operate primarily in the summer in years when Ontario's peak occurs in the summer and then switch to the winter if it was a winter peaking year. We don't hypothetically build a second plant to deal with a switch to winter peaking characteristics, either temporary or permanent.

Therefore, as is the case with savings impacts, double-counting avoided cost would be inappropriate. Given the accepted methodology for calculating cost-effectiveness, Program

Administrators must use their earlier declaration of which peak block of hours the initiative is designed to impact. The value of energy savings is unaffected, but the avoided cost of capacity may vary by time period and therefore that value would be applied to the appropriate peak hour used for the avoided capacity cost calculation.

## Report Content and Format

The initial elaboration of peak demand calculation issues should be addressed in the overall Evaluation Plan (Draft and/or Final). The final Evaluation Report should include the following:

- Clarification of program/measure-selected definition of peak demand
- Methodology used to assess program demand savings and program cost effectiveness
- Sampling plan (as well as the survey instrument) used to collect data and discussion of confidence interval
- Peak Demand Savings Results (Summer and Winter), including forecasts, reported energy savings, and verified energy savings levels (where applicable)
- Net Peak Demand Savings Results (Summer and Winter) adjusted for external factor including forecasts, reported and verified energy savings (where appropriate);
- Analysis of variances between forecast, reported, and verified demand savings.

In all cases, the Evaluation Administrator must sign off on the estimation of peak demand savings demonstrated.

The Evaluation Administrator, once they have signed off on the peak demand analysis plan, as outline in the Final Evaluation Plan, is accountable for ensuring the analysis is completed on schedule. Once complete, the Evaluation Administrator must sign off on the estimation of peak demand savings. However, the analysis itself will be carried out by the Evaluation Contractor.

It is essential that the Program Designer establish an appropriate tracking system to facilitate this analysis and provide the Evaluation Contractor with all requested tracking system outputs. Once completed, the Evaluation Administrator informs program designers and delivery agents of key findings from the final demand impact analysis. This feedback is crucial, as it helps the Program Designer:

1. improve on existing program designs
2. develop accurate initial peak demand savings forecasts
3. make decisions about funding and incentive levels provided for the program or similar programs

### Summary of Actions

- Choose method for estimating peak demand savings
- Sign off on Evaluation Contractor's proposed plan for calculating demand savings
- Provide Evaluation Contractor with requested tracking system output
- Ensure report provides required information and details
- Use key finding from report to consider ways of improving program design

# Technical Guide 7: Market Effects Guidelines

## Key Points / Highlights

To be a candidate for a market effects evaluation, the intended market effects should be a distinct part of the program strategy, an intended outcome of the program and have goals or targets forecasted. Ideally, the program administrator should be able to show that a share of the program budget or other resources was allocated with market effects as the intent.

Where substantive market effects are anticipated, simple net-to-gross (NTG) ratios may prove inadequate. In their place, a market effects study should be commissioned to explore changes in market structure and attitudinal changes that contribute to a higher standard of practice.

## Experimental Approach to Determining Market Effects

Evaluation Administrators should conduct in-depth interviews with Program Administrators, trade allies, and program participants to better appreciate the potential outcomes of the planned program design. These interviews should record changes made or changes expected in both the attitudes and abilities of each market actor as a result of the program offer.

The behaviours of market actors should be monitored and all significant changes recorded. The behavioural changes should then be correlated against the variables such as participant activities, perspectives and abilities.

And as a final step, the Evaluation Contractor must ultimately establish causal attribution leading from the program activities, through the realized outputs, accumulated through program outcomes and then to the intended impacts. This attribution pathway provides a foundation that allows Program Designers to assert that a program has broad market effects and creates market transformational savings.

Using the analytical approaches supported by this market transformational model, broad market effects can add significantly to program savings estimates and positively adjust net-to-gross ratios for both measures and programs affected by market changes.

## Analytical Methods Used to Determine Market Effects

Market effects analyses require greater effort than the more typical cross-sectional analyses. Market effects, by their very nature, contribute savings year-upon-year following even a single market intervention.

In the later stages of market transformation, when the market interventions have ceased, the market effects evaluations serve as the program offer and leads to energy savings. The analytical methods applied to measure the market effects selected should take into account this longer-term horizon.

**Methods for Analyzing Market Effects**

- **Longitudinal analyses** –these enable Program and Evaluation Administrators to compare one pre/post period. Therefore, key market externals must be normalized to some comparable base-year or long-term trend.
- **Market characterization studies** – these serve as the data collection instrument for both cross-sectional and longitudinal assessments. These studies effectively capture a snapshot of the market that can be used as a benchmark and/or that can be analyzed to provide normalization factors for key variables and a time series of key program performance metrics.
- **Experimental studies** – these provide valuable explanatory findings that can be used to draw conclusions and formulate program recommendations. Even narrowly focused panel studies and Delphi analysis can help build expert consensus around key issues. In-field metering studies also contribute to establishing program behavioural outcomes (by helping clarify consumer electricity end-uses and the use of energy consuming appliances). Lastly, natural occurring and planned market experiments provide evidence of causal attribution and therefore should not be ignored.

## Asserting the Existence of Market Effects

Evaluation administrators must carefully weigh the potential for market effects. Studies of market effects require a significant investment of both human and capital resources. In the event a potential claim of substantial market effects is absent, a market effects study should be narrowly scoped or avoided altogether.

Still, where substantial and transformational outcomes are expected, the Evaluation Administrator should be prepared to undertake a multi-year, multi-facetted study to capture the breadth of expected market effects. Also, the Evaluation Administrator should work closely with the Evaluation Contractor to ensure the use of methods that take into account causal attribution. Where methods allow for causal attribution, it may be found that the long-term market effects lead to savings greater than the annualized impacts sought directly from the program activities.

### Summary of Actions

- Consider whether market effects are likely relevant; if they are, consider carrying out a market effects evaluation
- Consider what analytical method to use in evaluating market effects

# Technical Guide 8: Net-To-Gross Ratio Adjustment Guidelines

## Key Points / Highlights

The "net-to-gross ratio" (NTG) is an adjustment factor applied to estimates of gross savings (**Technical Guide 5: Gross Energy Savings Guidelines**) to account for those energy efficiency gains that are only attributable to, and the direct result of, the conservation and demand management program in question. The NTG represents the comparison between an estimate of savings achieved as a direct result of program expenditures and an estimate of savings that would have occurred even in the absence of CDM program.

## Purpose of This Guideline

This guideline provides guidance for determining net-to-gross ratios for the estimation of program net impacts. Net savings estimates are the proportion of the gross savings that would have occurred in the absence of the program. Determination is usually done at the program level, but a more refined level of granularity may be warranted in some cases.

Net savings are of most interest to public or ratepayer-funded programs where the responsible party is interested in the influence of the program in producing incremental savings. In contrast, a government or private-sector in-house energy efficiency program or performance contract will be much more interested in total, or gross, savings.

Program benefits used in cost-effectiveness evaluations consider the program's **net savings** as opposed to **gross savings**.

Several factors can reduce or, in some cases, increase the net impacts attributable to a program. Deciding which of these factors to account for in an analysis of net savings is influenced by the objectives of the evaluation. Factors that differentiate net savings from gross savings are also sometimes called "distortion effects", or net-to-gross "adjustment factors", and can include the effects of free ridership, spillover, rebound effects, and transmission and distribution losses (described below). Free ridership is the most commonly evaluated adjustment factor, followed by spillover, and rebound effects.

Participant and non-participant surveys and tracking behavioural changes can help in determining net-to-gross ratio.

Program and Evaluation Administrators of ratepayer or publicly-funded CDM programs will be interested in estimating the net savings attributable to these programs. Program Administrators should consider likely net-to-gross factors during the design and development of a program and in designing the program logic model. Net-to-gross factors should be considered from a risk management perspective because factors such as free ridership, detract from the savings and cost-effectiveness of program investments, while other factors, such as spillover and transmission and distribution losses, can augment savings attributable to program activities.

In selecting an evaluation approach, Program Managers need to consider the level of effort to be devoted to studying net-to-gross factors (**Step 7: Evaluation Plan Development Guidelines**). The approach is tied to the program objectives, size, and scale of the program; the evaluation budget and time available; available resources; and specific aspects of the measures and participants in the program.

## Net-to-Gross Ratio (NTG) Basic Concepts

Energy and demand savings that occur due to CDM program activities are first determined as gross savings. Program net savings are then estimated by adjusting (discounting or augmenting) the gross savings by applying a set of net-to-gross "adjustment factors," such as free ridership rates, spillover effects, and rebound effects. The aggregate effect of these factors in a program impact evaluation is represented by the NTG.

The value of the NTG can vary dramatically depending on the type of program; how the program is implemented in the marketplace; the number of other programs that reach similar customer classes; or other market influences, such as codes and standards. For example, participants in some programs may be largely free riders whereas other programs may have virtually no free ridership.

To determine an estimated NTG value for program design, Program Administrators should incorporate free ridership rates and spillover effects, but may choose to disregard rebound effects. However, we recommend that all net-to-gross factors be considered when estimating the value of the NTG, especially when these factors could be significant.

Some, though certainly not all, of the common net-to-gross factors that are used to calculate the NTG are:

**There are three general categories of free ridership:**

- **Total free riders** – the total of consumers that would have installed the program-promoted measures at the same timeframe, regardless of program's existence
- **Partial free riders** – consumers who would have installed measures that are more efficient than baseline, but less efficient than the program promoted-measures, or who would have installed fewer of the program- promoted measures
- **Deferred free riders** – consumers who would have installed the program-promoted efficient measures, but at a later time

### Free Ridership

Free ridership is a measure of program participants that would have implemented the program measure or practice even in the absence of the program. Savings do occur as a result of free ridership, but they may not be directly attributable to the program being evaluated, and thus these effects reduce the direct impact of the program.

### Spillover Effects

Spillover effect occurs when the presence of an energy efficiency program influences customers to reduce energy consumption or demand, but the incremental savings are not directly a result of the program. Non-participant spillover is sometimes called "free drivership", which is the effect of people or companies that install energy efficiency measures as a result of the effects or influence of a program, but who never collect a rebate or incentive. These behavioural changes increase the effect of the program and can partially offset the effects of free ridership.

### Program Enabled Savings (PES)

Program enabled savings are energy and demand savings resulting from additional energy efficiency actions that program participants or non-participant might have undertaken because of program influence, but for which they received no financial incentives. They are often referred to as "spillover" savings.

Types of program enabled savings can include:

- Operational/process changes
- Additional equipment retrofit
- Behaviour change

### How can Program Enabled Savings be calculated:

For savings to be claimed, they must be quantifiable. Quantification must be transparent, assumptions clearly stated, and back-up documentation must be accessible. The following is a list of documentation that may be requested in order to calculate and validate savings claims:

- Description of the project with contact details
- Description of the Existing Condition/ Baseline
- Description of the Efficient Condition
- Annual Savings Estimate (kW, kWh)
- Persistence estimate
- Input assumptions used (with references), Engineering Calculations
- In service date
- Operating schedules
- Process modifications
- Project cost estimates

### Rebound Effect

A rebound effect is an increase in energy-using behaviour following customer action to increase efficiency. This is sometimes referred to as "snap-back". An example of rebound is when customers increase their use of equipment after they have installed energy efficient equipment, or when customers use more energy when rates are low, such as during off-peak hours.[2] For example, curtailing residential air conditioning load during a set period reduces the consumption during that period, but there is a rebound effect if the customer increases their consumption by running the air conditioner harder and longer in the hours following the curtailment to make up for the increased heat and/or humidity in the home. This rebound effect can potentially offset a major part of the energy savings of a residential air conditioning load control initiative. Of course, in that case, the rebound effect might not be of much concern if the intention is to accomplish demand savings during specified times and there is a greater benefit to reduce demand.

2 This can occur under a time-of-use rate structure or a critical-peak-pricing regime.

## Electricity Transmission and Distribution Losses

Because electricity is lost through the process of transmission and distribution of energy between a power plant and a consumer, when an efficiency project reduces the electricity consumption at a facility, electricity transmission and distribution losses are avoided. As a result, the amount of electricity saved by no longer having to be generated at a power plant is actually greater than the reduction experienced at the site (note that electricity transmission and distribution losses do not come into play in evaluations or net savings calculations because they are accounted for at the public reporting stage (see **Step 12: Provincial Reporting Standards**).

Other influences that come into play when determining gross savings include:

- the effects of multiple programs operating within a utility service area or region
- overlapping effects that can occur when marketing and promotion for energy efficiency programs are broadcast in neighbouring jurisdictions or service territories (through print media, radio or television) and,
- influence of energy efficient codes and standards that reduce the availability of low efficiency equipment can have the effect of increasing free ridership.

## Approaches for Determining Net-to-Gross Ratios

There are three approaches[3] for determining net-to-gross ratios:

1. Self-reported surveys and enhanced self-reporting surveys
2. Econometric methods
3. Agreed on net-to-gross ratios

All three approaches can be used with any type of CDM program, but econometric methods require large numbers of participants. Agreed on net-to-gross ratios are the least costly approach, followed by self-reported surveys and enhanced self-reporting surveys.

### 1. Self-reported surveys

Self-reported surveys ask participants a series of questions to get at what actions they would have taken in the absence of the program. Estimates of spillover effects can be developed by surveying non-participants. Surveys can be web-based, distributed in hard copy, or administered by telephone. Self-reporting surveys are the lowest cost approach to estimating free ridership and spillover rates for specific programs that support particular technologies or measures.

A word of caution about situations where respondents self-select for participation in the survey: self-selection bias can skew the results because those with strong opinions or higher

**Table 3 Sample free ridership survey question matrix** (For illustration purposes only)

| Survey Question | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| Required financial help? | Yes | No | No | No | No | No | No | No | No |
| Previous experience with technology? | – | No | No | No | Yes | No | Yes | Yes | Yes |
| Planned to install measure without program? | – | No | No | Yes | No | Yes | Yes | No | Yes |
| Program influenced install decision? | – | Yes | No | Yes | Yes | No | Yes | No | No |
| Free rider score | **0.0** | **0.0** | **0.17** | **0.33** | **0.33** | **0.67** | **0.67** | **0.67** | **1.0** |

*Source:* Adapted from BC Hydro, Power Smart Partners Program Free Ridership Case Study

3 National Action Plan for Energy Efficiency (2007). Model Energy Efficiency Program Impact Evaluation Guide. Prepared by Steven R. Schiller, Schiller Consulting Inc. www.epa.gov/eeactionplan.

degrees of knowledge about the subject tend to be more willing to take the time to participate in a survey.

A typical self-reporting survey asks a series of questions and may present respondents with an answer scale, rather than allowing for simple yes or no responses. A sample set of survey questions is provided below and **Table 2: Sample free ridership survey question matrix** illustrates an example of how these types of questions can be used in conjunction with a matrix to estimate free ridership.

- Did you require financial assistance in order to go ahead with the install?
- Did you have previous experience with the energy efficient technology?
- Had you already planned to install the measure without the program/incentive?
- Did the program/incentive influence your decision to install the measure?
- Would you have installed the same number of measures without the program/incentive?
- Would you have selected the same level of efficiency without the program/incentive?

### Enhanced self-reporting surveys

Enhances self-reporting surveys are used to improve the quality of information used to provide net-to-gross ratios derived from self-reporting survey methods. Multiple additional data sources and techniques can be used to get at the rationale for decisions to install energy efficiency measures or to adopt conservation behaviours. Some of these techniques include:

- **In-person surveys** – surveys conducted in person can improve the quality of the survey results because personal views and information can assist in understanding the influences and motivations that determine the role of CDM programs in participant and non-participant decision-making processes.
- **Project analyses** – these analyses consider specific barriers to energy efficient measure installations and document participants' rationale for proceeding with the measure or project. For example, since most barriers to energy efficiency are related to the costs of installation, conducting a financial payback analysis on a project may reveal the likelihood that the customer would have proceeded with the project in the absence of the program if the project is shown to have a very short payback period. Feasibility studies, engineering reports, and internal memos are examples of other documentation that may provide insights into whether a customer would have proceeded with a project regardless of the program.
- **Non-specific market data collection** – this involves collecting information from other programs to estimate an appropriate net-to-gross ratio or a reasonable range to apply to the program being evaluated.

### 2. Econometric Methods

Econometric methods are mathematical models that use statistics and energy and demand data from participants and non-participants to derive accurate net-to-gross ratios. Applying econometric methods are the most costly way of estimating net-to-gross factors and require large numbers of participants and comparable non-participants to make accurate estimates.

Any of the above methods can be combined with participant and non-participant surveys to estimate free ridership, spillover, and rebound effects. When non-participants are included in the net-to-gross ratio, care must be taken to select a group that is comparable to the participant group.

### 3. Agreed on Net-to-Gross Ratios

In some jurisdictions, agreed on net-to-gross ratios may be set by regulatory boards or commissions to be used by Program Administrators. Agreed on net-to-gross ratios can be used when the cost of conducting more detailed analyses of program net-to-gross factors is a barrier or when the accuracy of the results is not paramount. Agreed on net-to-gross ratios are often periodically updated based on reviews and evaluations of net-to-gross factors. They are not used in the context of Conservation First 2015-2020 EM&V Protocols.

## Adjusting Gross Savings to Estimate Net Savings

The net program savings are calculated in a similar manner as the gross program savings with the difference being the number of tracked participants and/or measure is discounted (or increased) by net-to-gross adjustment factors determined through the program evaluations. The net program savings are calculated as shown in Equation 1.

## Timing Of Consideration Of Net-To-Gross Factors

Net-to-gross factors should be examined during the evaluation planning stage (**Step 7: Evaluation Plan Development Guidelines**). The evaluation should seek to identify and to clarify, through participant surveys and follow-up activities, the net-to-gross factors and their relative magnitudes. Net-to-gross factors are determined once, at the time of the evaluation.

**Equation 1**

$$PS_{net} = \sum_{i=0}^{N} (NTG_i \times N_i \times S_i)$$

where,

| | | |
|---|---|---|
| $PS_{net}$ | = | Net program savings (kWh / Kw) |
| NTG | = | Net-to-gross ratio (e.g., %) |
| N | = | Number of tracked participants/measures installed |
| $S_i$ | = | Adjusted gross savings for the $i^{th}$ participant/measure |

*Note that adjusted gross savings will vary according to the various types of measures (i.e. prescriptive, quasi-prescriptive, and custom) and should account for adjustment factors (i.e. realization rate, installation rates, etc.).*

### Summary of Actions

- Consider whether the gross savings estimated should be adjusted by a net-to-gross ratio
- Consider whether there might be free ridership, spillover effects, or rebound effects
- If a net-to-gross ratio adjustment is appropriate, consider the best approach for determining the adjustment; for example, consider whether to use an agreed-to ratio, self-reporting or enhanced self-reporting surveys, or econometric methods.

# Technical Guide 9: Demand Response Load Impact Guidelines

## Key Points / Highlights

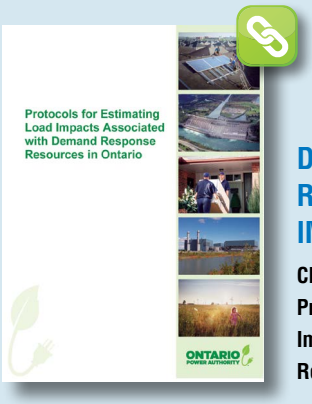
This document is intended to provide the necessary background to understand and implement the protocols, as well as provide guidance on related evaluation issues that go above and beyond the minimum requirements.

The intended audience for this document includes:

- evaluation contractors performing DR impact evaluations
- staff responsible for managing DR program evaluations
- staff responsible for DR program design and resource planning
- staff of other organizations that may administer future DR programs/tariffs and,
- other stakeholders with an interest in DR program evaluation.

For demand response resources, evaluation methods are less well developed. Moreover, several key aspects of DR resources differ from energy efficiency in ways that impact their evaluation. Recognizing these differences, the OPA initiated development of a separate DR evaluation framework that includes the following key elements:

- a white paper that discusses issues and methods for estimating load impacts and cost effectiveness of DR resources and provides recommendations on a DR evaluation framework – completed in November 2008;
- a set of protocols for estimating load impacts;
- a framework for determining the cost-effectiveness of DR programs




**DEMAND RESPONSE LOAD IMPACT GUIDELINES**

**Click image to open the OPA's Protocols for Estimating Load Impacts Associated with Demand Response Resources in Ontario.**

# Technical Guide 10: Guideline for Statistical Sampling and Analysis

## Key Points / Highlights

Generally, when studying the impact of a program it is not viable to study every single program participant. Furthermore, with respect to a comparison group (or control group), it is nearly impossible and most often not feasible to study the entire range of eligible non-participants. Therefore, statistical sampling of the two populations (participants and non-participants) is used to gauge program effectiveness.

**Questions to consider when drawing samples from a population under study:**

1. **What if the sample population looks or behaves nothing like the larger population?** If the sample is not representative of the larger population, then it is not possible to say anything about the larger population by studying the smaller sample. To ensure accurate representation of a population that needs to take steps to avoid bias in the sample. Common biases found during sampling, particularly for evaluations, include: self-selection bias, non-response bias, and voluntary response bias. If researchers are aware of, or perceive that there is, a high likelihood that such biases may impact results, steps should be taken to mitigate such biases during the sample design stage.
2. **Is it ever certain that the sample population would achieve the exact results as the population under study?** In the very best case, the sample only provides an estimate of program effect. There is always a degree of uncertainty embedded in the estimate. Therefore, short of taking a census, there must be a recognition that some degree of uncertainty exists in any statement of program effect.
3. **What if the sample population being studied is affected by influences beyond the scope of the program offer?** Statistically significant effects may be observed even where a program has not been implemented. For example, a commercial building or industrial account may be shown to have reduced energy consumption by 20% following an economic recession. These same accounts may or may not be participants in a program designed to achieve energy savings. The question becomes what portion of the 20% is attributable to the program and what percent is associated with the economic recession and other external factors. As such, it is important to recognize that a correlation does not necessarily indicate causation.

To deal with such questions, the industry relies on a research process known as a sample design. This guideline provides a primer on this subject and provides guidance for determining what design is best suited to serve the research objectives. Consult with a statistics professional before to implementing complex statistical analysis.

## Defining the Study Population

When selecting a sample, the first question that must be asked is what is the population under study? To evaluate Conservation First programs, the first step is to decide whether the savings estimate is to be assigned at the provincial, regional or individual utility level.

This is important because a small rural LDC's customer base may be primarily single family homes, farms, and some small commercial accounts. This population would not be representative of the Greater Toronto Area; nor is it likely to resemble Ontario as a whole. Therefore, it may not be accurate to formulate a provincial savings estimate by studying the program participants from this single LDC. Conversely, it may not be accurate to project savings for this small single LDC from a broadly scoped study used to establish a provincial savings estimate. As such, it is essential to describe the characteristics of the population including, but not limited to, size and variance.

## The Need for Strata

How the study population is defined will determine what conclusions can be drawn from the evaluation. As a result, it is sometimes necessary to stratify (sub-divide) the population. In the example above, a provincial savings estimate is desired plus the means to allocate savings to individual LDCs. Therefore, it may be practical to sub-divide Ontario into strata by individual LDC or by stratum of LDCs with similar characteristics. By dividing the population into distinct and independent strata, researchers can draw inferences about the sub-populations that otherwise would be lost in a more broadly defined sample.

If the following conditions exist, applying stratification is likely appropriate:

- Variability within the defined strata are reduced
- Variability between the defined strata are maximized and,
- Variables used to stratify the population are strongly correlated with the desired dependent variable.

These three criteria may help show that the LDC is not the appropriate differentiating stratum, and it could be something else.

## Advanced Stratification Options

To apply stratification, information about the characteristics of the population is required. Absent prior research, the researcher will have difficulty in defining appropriate strata. If that happens, the researcher may look to more advanced statistical methods to define the appropriate strata.

The two most common advanced approaches are over-sampling and post-stratification. With over-sampling, the researcher intentionally biases the sampling process to represent a known about the population, such that the resulting findings better represent the study population; even when the population itself cannot be appropriately sampled. For example, if it is known that there is a high non-response bias from a particular demographic of participants, the researcher may want to over-sample this population or sub-population to ensure that the actual number of responses received meets statistical requirements. In addition to over-sampling, a technique known as post-stratification may be used to develop estimates about sub-populations after the study is complete and can be used if characteristics about the sub-populations are unknown at the time the study is conducted. An example of this technique may be to simply over-sample a population to develop a provincial savings estimate for a program that can later be stratified to yield LDC savings estimates.

Both over-sampling and post-stratification are advanced research methods and are fraught with potential pitfalls. If applied incorrectly, these two techniques could compromise compliance with the Protocols. These advanced techniques should be reserved for specific situations and used only after careful consideration of other options. In addition, use of the methods should be well documented in the experimental approach of the **Draft Evaluation Plan**.

## Sample Selection

With the population and sub-populations defined, the researcher may turn his or her attention to selecting samples representative of the defined populations. These study populations are often referred to as the sample frame.

The sample frame is simply the pool from which a sample will be drawn; ideally, this will be from the entire study population. The worst-case scenario for a sample frame is to use a population of convenience, such as individuals who have participated in an initiative, to complete a questionnaire if they choose to (the reason using such a population is not a good idea, is because those who complete the questionnaire typically are people with strong opinions or higher degrees of knowledge about the subject and therefore are not necessarily representative of the entire population participating in the initiative). As a result it is important to use the appropriate sampling technique to address such biases during sample selection. Regardless of the sample methodology chosen, it is important to always keep in mind that a sample must be drawn to represent the population under study.

Of course, there are many other sampling techniques that could be employed in the study of conservation and demand management initiatives. The EM&V Protocols allow researchers to draw from the wide array sampling techniques available, however justification and documentation should be provided with regards to the sampling method employed.

**The most common probability sampling techniques used to study energy efficiency and conservation programs are:**

- **Simple Random Sampling:** This involves the random assignment of members from the study population to the study sample. This could be done, for example, using a computer to randomly assign 15% of program participants to the study sample.
- **Systematic Sampling:** This involves the systematic assignment of members from an ordered study population to a sample; for example, every 12th participant entering a program may be selected for the study sample.
- **Matched Random Sampling:** This involves the selection of members from the population based on relevant characteristics and assigning them to a group, then randomly selecting samples from within each group. For example, the researcher may decide to categorize participants by facility size and select a random sample from each group. This technique may be used to select a comparison group when studying a program. Alternatively, the use of a matched control group can be used to normalize estimates obtained for a study population.
- **Quota Sampling:** This is when the researcher is asked to sample a fixed number of members that meet specific criteria and assign them to a study sample; for example, a researcher may be asked to survey 400 middle-aged women and 300 middle-aged men. Quota sampling relies on the researcher's judgement and convenience in sample selection. Because of this, quota sampling is a non-proportional (biased) sampling technique.
- **Panel Sampling:** This involves the longitudinal study of a previously defined sample. For example, this approach may be employed to infer how a population is likely to react to an increase/decrease in energy prices.

## A Situation Requiring a Non-Probability Sample

If the goal is to study electricity use across the whole of Ontario, the broad scope of such an effort would require the population to be stratified. By doing so, several sub-populations could be identified based on similar characteristics and each can be studied independently of the other.

One such stratum could be industrial or manufacturing facilities, for example. Since the sub-population of industrial and manufacturing customers is typically not a homogeneous group, a non-probability sample may be employed for this stratum while using a random probability sample for the remaining strata. Because of the inherent differences between the energy use of the various industrial and manufacturing customers, a random sampling of this stratum could lead to unintended biases, namely, the selection of unusually large or abnormally small customers whose energy use are not representative of the stratum. In this case, a subject matter expert or a sector specialist may be better able to define a representative sample of the population. For example, the sector specialist may be able to isolate from the stratum some of the odd accounts and systematically select a sample from the remaining customers that can represent the group as a whole.

By allowing a sector expert to help with the sample selection, a more accurate study of the industrial and manufacturing sub-population can be realized than would be achieved based on a simple random sample. *Non-probability* samples must be carefully considered to ensure that sampling bias is explicitly identified and kept to a minimum.

## Sizing the Study Sample

Some of the main advantages of sampling are:

- sampling is less expensive than conducting a census of the whole population;
- the data can be analyzed easier and there is greater flexibility in the analytical methods that can be applied; and
- sampling can lead to greater sensitivity for the study of populations and sub-populations (as required).

However, researchers should also be aware that the trade-off to studying a sample as opposed to the entire population can lead to errors and inferences being made about the population that may not be completely accurate. Thus, it is important for researchers to be comfortable with the level of precision that their sampling strategy can provide.

One consideration that must be addressed when sampling any population or stratum is the degree of precision desired for an estimate. Another factor is the confidence level sought. An evaluation contractor may have a requirement for the savings estimate to be ± 5% at a 95% level of confidence. That is to say a repeated sampling of the population would result in a mean savings estimate that is within 5% of the true mean of the population 95 times out of 100.

To determine the required size of the study sample, the researcher must consider the desired levels of precision along with some assumptions about the normal variance around the *population mean*. Generally, the mean of the population is not known; otherwise a study of that population would not be necessary. Where the mean of the population is unknown, the variance around that mean is also unknown.

Therefore, an assumption often has to be made regarding the coefficient of variance, which is the dispersion of a probability distribution. Typically, the coefficient of variance is set at 0.5%, when other studies are not available to inform the likely variance around the population mean sought. The setting of the coefficient of variance at 0.5 is often acceptable because such a coefficient is indicative of neither a weak nor strong *dispersion*.

## Deciding on a Statistical Test

Statistical testing is generally used by researchers to describe a given population, make comparisons against a hypothetical value, or establish predictions based on known values. In this section we outline tests commonly used to make inferences; however this section is not intended to be a step-by-step manual that explains how to perform these calculations, since most situations are unique in terms of inputs and desired outcomes.

As there are several types of statistical test models that can be employed during an experiment, researchers must take care to determine the most appropriate test to answer their particular research question(s). Statistical test selection can be quite a simple exercise or highly complex depending on the nature of the study. Because one or more tests may be suitable, to address a research question we recommend that one consult a statistics professional before finalizing the required test.

To determine the most suitable test, the researcher must first determine the distribution of the population. Populations with a normal (Gaussian) distribution, or close to a normal distribution, will be more suitable to certain tests while unique techniques may make it harder to test populations with a *non-normal* distribution. In this guideline we focus on those tests that are suitable for normally distributed populations; however it is important to note that if the population being studied is not normally distributed, there are alternative testing methods that should be employed. Common examples of where a population may not be normally distributed include purchasers of luxury items and early adopters of new technologies.

Researchers are to determine if they anticipate one possible outcome or two possible outcomes from the test being performed. As well, the researcher must also determine the purpose for the outcome of the test.

Below is a matrix of commonly used statistical tests for normally distributed populations. Keep in mind that the items included are only some of the tests, researchers may wish to use other test models.

Researchers should carefully document in the Draft Evaluation Plan the rationale behind the chosen test method and should outline all calculation methodologies applied.

**Table 4.0 Common Statistical Tests for Normally Distributed Populations**

| Goal | Possible Outcomes | |
|---|---|---|
| | One (Measurement) | Two (Binomial) |
| **Describe a group** | **Mean and Standard Deviation** | **Proportion** |
| Compare a group to a hypothetical value | One-sample t-test | Chi-square Test or Binomial Test |
| Compare two unpaired groups | Unpaired t-test | Fisher's Test or Chi-square Test |
| Compare two paired groups | Paired t-test | McNemar's Test |
| Compare three or more unmatched groups | One-way Analysis of Variance | Chi-square Test |
| Compare three or more matched groups | Repeated Measure Analysis of Variance | Cochrane Q |
| Quantify association between two variables | Pearson Correlations | Contingency Coefficients |
| Predict value from another measured variable | Simple Linear Regression **or** Nonlinear Regression | Simple Logistic Regression |
| Predict value from several measured or binomial variables | Multiple Linear Regression **or** Multiple Nonlinear Regression | Multiple Logistic Regression |

## Summary of Actions

- Define the study population
- Determine whether there is a need for stratification of the population chosen
- Decide on the sampling technique that will be used
- Decide on the sample size
- Decide whether to apply a statistical test
- Ensure the report includes information relating to the test method chosen as well as the rationale for choosing that test

# Technical Guide 11: Behaviour-Based Evaluation Protocols

## Key Points / Highlights

This document sets forth the basic protocols that are to be used in evaluating behavioral programs. Chapters 1 - 3 introduce the protocols, describe the philosophy behind their development and outline the types of programs that are governed by the protocols that are to be applied. Chapter 4 discusses the protocols that are to be used for cost benefit analysis, process evaluations and market effects studies. Chapter 5 introduces the basic research designs that are appropriate for assessing the impacts of behavioral interventions. Chapters 6 through 9, provide protocols for designing impact evaluations for Training/Capacity Building programs, Information Feedback programs and Public Information Programs. Finally, Chapter 10 provides protocols for analyzing data from experiments and other research designed to assess the impacts of behavioral programs.

## When to Use this Guide:

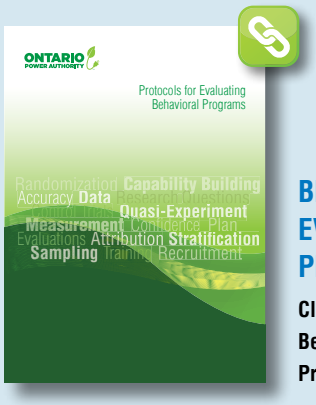
Behaviour-Based Evaluation Protocols should be employed when assessing the impact of behavioural programs on energy consumption. The following are examples of programs intended to alter behavior to achieve energy savings include:

- providing normative comparisons in which consumers are provided with comparisons of their household energy consumption with that of other purportedly similar households
- providing feedback technologies that allow consumers to observe their energy use at websites or from devices installed in their homes
- providing home automation technologies to consumers that help them consume less energy
- providing time varying rates that help consumers lower their energy consumption to reduce demand on the electric system while saving money on their bills
- providing financing for energy efficiency investments designed to encourage consumers to purchase more energy efficient equipment
- providing training to various market actors to enhance the likelihood that they properly size and install energy using equipment
- providing training to building industry professionals to assist them in designing and building energy efficient buildings

## How to Use

These protocols are intended by be used by evaluators and policy makers to plan and carry out evaluations of behavioural programs. They describe best practices for evaluating such programs as well as the minimal information that must be reported regarding the selection of research methods and results. Four basic types of evaluations may be required in assessing the performance of behavioral intervention programs. They include:

- **Impact evaluations** – assessment of the impacts of capacity building programs on energy consumption;
- **Market effects evaluations** – assessments of the impacts of capacity building programs on various aspects of the market including changes in sales and prices of energy efficiency measures, prevalence of behaviors and opinions that influence energy consumption and actions that may be taken by market actors in response to the program;
- **Cost effectiveness evaluations** – assessments of the extent to which cost savings resulting from the program exceed the costs of delivering the program; and
- **Process evaluations** – assessments of the extent to which the process used to deliver the program was efficient and effective in accomplishing its intended purpose.




**BEHAVIOUR-BASED EVALUATION PROTOCOLS**

**Click image to link to the OPA's Behaviour-Based Evaluation Protocols.**

# Tables and Figures

## Figures



## Tables

# Glossary and Bibliography

# Glossary of General Program Evaluation Terminology

The definitions in this glossary are adapted from federal, provincial, and academic sources, many of which are listed in the bibliography at the end of this appendix.

**Accuracy**
The correspondence between the measurements made on an indicator and the actual value of the indicator at the time of measurement.

**Activities**
A term used generically in logic modeling to describe the action steps necessary to produce program outputs.

**Administrative Agency**
An organization tasked with administering electric generation, transmission, distribution, reliability, and conservation programs within the Province of Ontario, such as the OPA, OPG, IESO, etc.

**Bias**
The extent to which a measurement, sampling, or analytical method systematically underestimates or overestimates a value.

**"CDM" Conservation and Demand Management**
Outside of Ontario CDM is often referred to as Demand–Side Management (DSM) and so CDM and DSM are often used interchangeably.

**Comparison Group**
A group of individuals or organizations that have not had the opportunity to receive program benefits and that have been selected because their characteristics match those of another group of individuals or organizations that have had the opportunity to receive program benefits. The characteristics used to match the two groups should be associated with the action or behaviour that the program is trying to promote. In evaluation practice, a comparison group is often used when random selection of recipients of the program benefit and a control group is not feasible.

**Control Group**
A randomly selected group of individuals or organizations that have not had the opportunity to receive program benefits. A control group is measured to determine the extent to which its members have taken actions promoted by the program. These measurements are used to estimate the degree to which the promoted actions would have been taken if the program did not exist.

**Cost-Benefit**
Comparison of a program's outputs or outcomes with the costs. Benefit-cost is an alternate. The comparison of a cost to a benefit is often expressed as a ratio.

**Cost-Effectiveness**
Comparison of a program's benefits with the resources expended to produce them.

**Cost-Effectiveness Evaluation**

Analysis that assesses the cost of meeting a single output, objective, or goal. This analysis can be used to identify the least costly alternative to meet that output, objective, or goal. Cost-benefit analysis is aimed at identifying and comparing all relevant costs and benefits. The analysis is usually expressed in dollar terms. The two terms (cost effectiveness and cost benefit) are often interchanged in evaluation discussions.

**Deemed Savings**

An estimate of an energy savings or energy-demand savings outcome (gross savings) for a single unit of an installed energy-efficiency or renewable-energy measure that:
(1) has been developed from data sources and analytical methods that are widely considered acceptable for the measure and purpose, and
(2) will be applied to situations other than that for which it was developed.

That is, the unit savings estimate is "deemed" to be acceptable for other applications. Deemed savings estimates are more often used in program planning than in evaluation. They should not be used for evaluation purposes when a program-specific evaluation can be performed. When a deemed savings estimate is used, it is important to know whether its baseline is an energy-efficiency code or open-market practice. Besides the OPA's Measures and Assumptions Lists (**Technical Guide 1: Using Measures and Assumptions Lists**), an extensive database of deemed savings is also available in California's Database for Energy Efficiency Resources (DEER). Note that the deemed savings in DEER are tailored to California and should not be used for Ontario initiatives without thought or review. If there are measures on deemed savings lists from other jurisdictions that are not on the official Lists in **Technical Guide 1: Using Measures and Assumptions Lists**, please request that they be analysed and added.

**Defensibility**

The ability of evaluation results to stand up to scientific criticism. Defensibility is based on the assessment by experts of the evaluation's validity, reliability, and accuracy. See also *Strength*.

**Evaluation, Measurement & Verification (EM&V)**

The undertaking of studies and activities aimed at assessing and reporting the effects of a Conservation program on its participants and/or the market environment. Effectiveness is measured though energy efficiency and cost effectiveness.

**Evaluation Administrator**

The person responsible for developing an EM&V plan for a particular program or portfolio. This person is also the point-of-contact for EM&V contract management. This person is sometimes referred to as an Evaluation Manager.

**Energy Conservation Measures (ECM)**

An activity or set of activities designed to increase the energy efficiency of a facility, system or piece of equipment. ECM may also conserve energy without changing efficiency. An ECM may be applied as a retrofit to an existing system of facility, or as a modification to a design before construction of a new system or facility.

**Evaluation Contractor**

The individual(s) or firm(s) selected to implement the EM&V plan developed by the Evaluation Administrator. The Evaluation Contractor could also be referred to as the "Independent, Third-Party Evaluator" or the "Evaluator.

**Ex ante load impact estimate**

A load impact estimate representing a set of conditions or group of customers, or both, that differ from historical conditions (from the Latin word for "beforehand").

**Ex post load impact estimate**

A load impact estimate representing a set of conditions that actually occurred on a specific date or over some period of time for the customers that were enrolled in the program and called on that date or over that period of time (from the Latin word for 'something done afterwards').

**Free driver (free drivership)**

A non-participant who has adopted a particular efficiency measure or practice as a result of the evaluated program.

**Free rider**

A program participant who would have implemented the program measure or practice in the absence of the program. Free riders can be total, partial, or deferred.

**8760s**

Full year hourly consumption loads.

**Impact Evaluation**

The application of scientific research methods to estimate how much of the observed results, intended or not, are caused by program activities and how much might have been observed in the absence of the program. This form of evaluation is employed when external factors are known to influence the program's outcomes in order to isolate the program's contribution to achievement of its objectives.

**Indicator**

An indicator is the observable evidence of accomplishments, changes made, or progress achieved. An indicator is also a particular characteristic used to measure outputs or outcomes; a performance quantifiable expression used to observe and track the status of a process.

**Interactive Effects**

Energy effects created by energy conservation measure but not measured within the measurement boundary.

**Logic Model**

A plausible and sensible diagram of the sequence of causes (resources, activities, and outputs) that produce the effects (outcomes) sought by a program.

**Market Effects**

A change in the structure or functioning of a market or the behaviour of participants in a market that results from one or more program efforts. Typically the resultant market or behaviour change leads to an increase in the adoption of energy-efficient or renewable-energy products, services, or practices. Examples include an increase in the proportion of energy-efficient models displayed in an appliance store, the creation of a leak inspection and repair service by a compressed-air-system vendor, an increase in the proportion of commercial new-construction building specifications that require efficient lighting.

**Market Study Evaluation**

A study that characterize energy markets, assess spatial and temporal changes in market structure and function that result from program interventions and other external influences (i.e., such as codes and standards, fuel price volatility, and environmental concerns).

**Measurement**

A procedure for assigning a number to an observed object or event.

**Measurement and Assumptions Lists**

The OEB-approved electricity-sector "deemed savings" lists to be used for program planning and forecasting purposes. One major goal of EM&V program evaluations is to confirm or update these assumptions. **Technical Guide 1: Using Measures and Assumptions Lists.**

**Normalized Savings**

Savings calculated based on adjustments. The baseline energy use is adjusted to reflect "normal" operating conditions. The reporting period energy use is adjusted to reflect what would have occurred if the facility had been equipped and operated as it was in the baseline period under the same "normal" set of conditions. These normal conditions may be a long term average, or those of any other chosen period of time, other than the reporting period.

**Outcome**

A term used generically with logic modeling to describe the effects that the program seeks to produce. It includes the secondary effects that result from the actions of those the program has succeeded in influencing.

**Outcome Evaluation**

Measurement of the extent to which a program achieves its outcome-oriented objectives. Outcome evaluations measure outputs and outcomes (including unintended effects) to judge program effectiveness and may also assess program process to understand how outcomes are produced.

**Output**

A term used generically with logic modeling to describe all of the products, goods, and services offered to a program's direct customers.

**Peak demand**

OPA defines peak demand as follows:

**Table 1.0**

**OPA EM&V Standard Definition of Peak for Calculating Demand Savings**

Based on analysis of Ontario System Hourly Load data from 2003-2010, the defined summer and winter peak blocks for Conservation First Framework (2015-2020) are as follows:

**Average Load Reduction over Entire Block of Hours**

| | Time | Months |
|---|---|---|
| SUMMER (Weekdays) | 1pm - 7pm* | June |
| | | July |
| | | August |
| WINTER (Weekdays) | 6pm - 8pm | January |
| | | February |
| | | December |

*Daylight Savings Time-Adjusted

**Persistence of savings**

A critical element for many stakeholders is whether energy savings from the ECM and/or behavioral change continue over time. It is important to determine the value of the energy and demand savings beyond the initial program year. There are at least two different situations for which evaluators may assess persistence of savings

**Prescriptive measures**

A prescriptive measure uses defined or fixed input assumptions embedded into the energy and demand savings equations. These input assumptions can include default efficiencies for a type of equipment specified or annual operating hours for the type of building selected.

**Probability Sampling**

A method for drawing a sample from a population such that all possible samples have a known and specified probability of being drawn.

**Process Evaluation (or Assessment)**
of the extent to which a program is operating as its implementation intended. Process evaluations assess program activities' conformance to statutory and regulatory requirements, to program design, and to professional standards or customer expectations.

**Program Administrator**
The persons or organizations responsible for the design, development, and implementation of an energy efficiency, conservation, or demand response initiative. A Program Administrator may also be referred to as a "Program Manager" or a "Program Implementer." An LDC may also be a Program Administrator. Outside of an EM&V context there may be distinctions between Program Administrators and external Program Managers or other subtleties that are ignored in the EM&V context. In the EM&V context a Program Administrator is someone (or an entity) other than the Evaluation-related staff or entities.

**Program Evaluation**
Program evaluations are independent systematic studies conducted periodically on an ad hoc basis to assess how well a program is working and whether the program it is achieving its intended objectives. Program Evaluations are conducted by experts external to the program staff.

**Program Logic Model**
A diagram showing a causal chain with links that go from resource expenditure to long-term outcomes for a program.

**Program Manager**
The individual/group responsible for implementing a program

**Qualitative Data**
Information expressed in the form of words.

**Quantitative Data**
Information expressed in the form of numbers. Measurement gives a procedure for assigning numbers to observations. See *Measurement*.

**Quasi-prescriptive Measure**
A quasi-prescriptive measure has varying resource savings estimates according to the technology or type of equipment and the context in which they are used. It contains key, measure-specific inputs to estimate energy and peak demand savings for each program participant. It provides a methodology that allows estimating resource savings for various scenarios rather than relying on a fixed savings value for all scenarios. A quasi-prescriptive approach will allow different parameters or variables to be assumed to estimate different levels of resource savings for different retrofits in different business segments

**Random Assignment**
A method for assigning subjects to one or more groups by chance.

**Rebound Effect**
A change in energy-using behaviour that yields an increased level of service and occurs as a result of taking an energy efficiency action.

**Regulatory Authority**
The entity with the mandate to oversee the actions of local distribution companies and administrative agencies; in Ontario this could be the Ontario Energy Board (OEB), the Environmental Commissioner of Ontario (ECO), or the Ministry of Energy, or any combination of the three.

**Reliability**
The quality of a measurement process that would produce similar results on: (1) repeated observations of the same condition or event; or (2) multiple observations of the same condition or event by different observers.

**Representative Sample**
A sample that has approximately the same distribution of characteristics as the population from which it was drawn.

**Simple Random Sample**
A method for drawing a sample from a population such that all samples of a given size have equal probability of being drawn.

**Spillover**
Reductions in energy consumption and/or demand caused by the presence of the energy efficiency program, beyond the program-related gross savings of the participants. There can be participant and/or non-participant spillover.

**Strength**
A term used to describe the overall defensibility of the evaluation as assessed by use of scientific practice, asking appropriate evaluation questions, documenting assumptions, making accurate measurements, and ruling out competing evidence of causation.

**Structured Interview**
An interview in which the questions to be asked, their sequence, and the detailed information to be gathered are all predetermined. Structured Interviews are used where maximum consistency across interviews and interviewees is needed. Whereas unstructured interview is an interview used to elicit information in complex situations where questions can be changed or adapted to meet the interviewee's responses. Unlike structured interviews, it does not offer a limited, pre-set range of answers for an interviewee to choose, hence, the lack of consistency and reliability.

**Verified Savings**
The net evaluated energy and demand savings of a program. Verified Savings are used as the base for the allocation of savings to targets or for official reporting purposes.

# Bibliography


1. California Public Utilities Commission, "The California Evaluation Framework," June 2004. www.calmac.org/publications/California_Evaluation_Framework_June_2004.pdf

2. DOE," The Performance-Based Management Handbook, Volume 4, November 2000, Appendix A. www.orau.gov/pbm/pbmhandbook/pbmhandbook.html

3. GAO, "Designing Evaluations," GAO/PEMD-10.1.4, March 1991, pp. 92-94. www.gao.gov/special.pubs/pe1014.pdf

4. GAO, "Performance Measurement and Evaluation: Definitions and Relationships," GAO/GGD-98-26, April 1998. www.gao.gov/special.pubs/gg98026.pdf

5. OMB, "Preparation and Submission of Strategic Plans, Annual Performance Plans, and Annual Program Performance Reports, Circular No. A-11, Part 6, June 2002, section 200-2. www.whitehouse.gov/omb/circulars/a11/churrent_year/parta6.pdf

6. OMB, "Instructions for the Program Assessment Rating Tool," pp. 7-10. www.whitehouse.gov/omb/part/2006_part_guidance.pdf

7. McLaughlin, J.A., and Jordan, G. B., "Logic Models: A Tool for Telling Your Program's Performance Story," Evaluation and Program Planning, Volume 22, Number 1, February 1999.

8. University of Wisconsin Extension, "Planning a Program Evaluation," February 1996, pp. 2-10. www.uwex.edu/ces/pdande/evaluation/evaldocs.html

9. Ernest Orlando Lawrence Berkeley National Laboratory, Edward Vine, Jayant Sathaye, and Willy Makundi, "Guidelines for the Monitoring, Evaluation, Reporting, Verification, and Certification of Forestry Projects for Climate Change Mitigation", Environmental Energy Technologies Division, March 1999. http://ies.lbl.gov/iespubs/41877.pdf