

**Illinois Statewide
Technical Reference Manual
for Energy Efficiency
Attachment A
Illinois Statewide
Net-to-Gross
Methodologies**

January 23, 2015

DRAFT

Effective for Evaluation:

June 1st, 2015

Table of Contents

I.	Illinois Statewide Net-to-Gross Methodologies	3
A.	Policy Context for this Information	3
B.	SAG consensus on deviations from the methods	Error! Bookmark not defined.
C.	Programs Currently Covered in this Document	4
D.	Updating the IL-NTG Methods.....	4
E.	Procedure for Non-Consensus Items.....	5
II.	Attribution in Energy Efficiency Programs in General.....	5
III.	Attribution within the Commercial, Industrial, and Public Sectors.....	8
A.	Standard/Prescriptive and Custom Programs	8
1.	Free Ridership.....	8
2.	Spillover	10
IV.	Attribution within the Residential and Low Income Sectors	10
A.	Appliance Recycling Programs.....	10
1.	Free Ridership.....	10
2.	Secondary Market Impacts.....	11
3.	Induced Replacement	12
4.	Integrating Free Ridership, Secondary Market Impacts, and Induced Replacement	13
a.	Residential Upstream Lighting Programs	Error! Bookmark not defined.
5.	Free Ridership.....	17
6.	Participant Spillover.....	19
7.	Nonparticipant Spillover.....	20
8.	Method Advantages and Disadvantages.....	20
V.	Appendix A: Overview of NTG Methods	22
A.	Survey-based approaches	22
1.	Self Report Approach.....	22
2.	Econometric/Revealed Preference Approach.....	23
B.	Randomized control trials (RCT) and quasi-experimental designs.....	23
C.	Deemed or stipulated NTG ratios.....	24
D.	Common practice baseline approaches	24
E.	Market analyses	24
F.	Structured expert judgment approaches.....	25
G.	Program Theory-Driven Approach.....	25
H.	Case Studies Design	26
VI.	Appendix B: References	27

I. Illinois Statewide Net-to-Gross Methodologies

A. Policy Context for this Information

The Illinois Evaluation Teams (Opinion Dynamics, Cadmus Group, Navigant Consulting, Itron, and ADM Associates) are working with the Illinois Stakeholder Advisory Group (SAG) to create an Illinois Statewide Net-to-Gross (NTG) Methodologies document (IL-NTG Methods). The IL-NTG Methods document is included as an attachment to the Illinois Statewide Technical Reference Manual for Energy Efficiency (IL-TRM). Through five different dockets, the Illinois Commerce Commission (ICC) has directed the Evaluation Teams to compile and formalize standard NTG methods for use in Illinois energy-efficiency (EE) evaluation, measurement and verification (EM&V) work. The ICC EE dockets are shown in the following table.

Table 1. ICC Energy Efficiency Dockets

ICC Order Docket No. and Date	Program Administrator	NTG Discussion - Order Pages	ICC Link
13-0495 (1/28/14)	Commonwealth Edison Company (ComEd)	129-130	ICC Order Docket No. 13-0495
13-0498 (1/28/14)	Ameren Illinois Company (Ameren)	167, 171	ICC Order Docket No. 13-0498
13-0499 (1/28/14)	Illinois Department of Commerce and Economic Opportunity (DCEO)	20, 23, 49	ICC Order Docket No. 13-0499
13-0549 (5/20/14)	Nicor Gas Company	41-42, 78	ICC Order Docket No. 13-0549
13-0550 (5/20/14)	North Shore Gas Company (North Shore Gas) and The Peoples Gas Light and Coke Company (Peoples Gas) (collectively, PG&NSG or Integrys)	54-55, 66	ICC Order Docket No. 13-0550

To provide clarity to the ICC directives, the relevant section on IL-NTG Methods is shown in its entirety from the Nicor Gas Order (Docket No. 13-0549). The Nicor Gas Order provides the most detail on the ICC NTG directive in comparison to the other EE orders. The Nicor language is as follows:

The Commission believes that Staff’s recommendations concerning Commission adoption of consistent statewide net-to-gross methodologies (“IL-NTG Methods”) for use by the evaluators are reasonable and will aid in future evaluation of the energy efficiency programs. To help ensure the independence of the evaluators, to improve efficiency in the evaluation process, and to ensure programs across the state as delivered by the various program administrators can be meaningfully and consistently evaluated, the Commission hereby adopts Staff’s recommendation that consistent IL-NTG Methods be established for use in the evaluations of comparable energy efficiency programs offered by different Illinois program administrators. The Commission notes that Section 8-104(k) of the Act encourages statewide coordination and consistency between the gas and electric energy efficiency programs and Staff’s proposal would help ensure consistency in the evaluation of program performance. The Commission notes that this directive is not to create entirely “new” NTG methodologies for every energy efficiency program, but rather to assess NTG methodologies and survey instruments that have been used to evaluate energy efficiency programs offered in Illinois, and to compile the most justifiable and well-vetted methodologies (or potentially combine certain components from the existing approaches to better represent the most justifiable and well-vetted method consistent with best

practices) in an attachment to the Updated IL-TRM that would get submitted to the Commission for approval. The Commission notes that the IL-NTG Methods will be flexible and adaptable to multiple program designs and budgets and tailored to appropriately assess the specifics of each of the program administrators' energy efficiency programs, consistent with standard NTG methodologies adopted in other states that were filed in this proceeding. The Commission agrees with Staff that in the interest of efficiency, the current program evaluators should take the lead in compiling and formalizing standard methodologies for NTG in Illinois taking into consideration SAG input. Because the existing Plan 1 evaluators are under contract with the Company for the evaluation of the program year three energy efficiency programs, it is appropriate for these existing evaluators to work on and complete the compilation of the IL-NTG Methods over the next year. The Commission recognizes that each year considerable time may be spent vetting NTG methodologies for each program evaluation separately for each utility under the existing evaluation plan review practices; adoption of IL-NTG Methods would save on these limited evaluation resources by having a common reference document for the evaluators to use in estimating net savings for Illinois.

The Commission hereby directs the Company to require its evaluators to collaborate with the other Illinois evaluators and the SAG to use best efforts to reach consensus on the approaches used in assessing NTG in particular markets for both residential and non-residential energy efficiency programs in a manner consistent with the direction described herein. (Pages 41-42)

(16) Northern Illinois Gas Company shall require its evaluators to collaborate with the other Illinois evaluators and the SAG to reach consensus on the most defensible and well-vetted methodologies for assessing net-to-gross ratios in particular markets for both residential and non-residential energy efficiency programs in a manner consistent with the direction provided herein;

(17) ICC Staff shall file the agreed-upon consensus statewide NTG methodologies with the Commission as an attachment to the Updated IL-TRM, and if consensus is not reached on a certain component of the statewide NTG methodologies, that particular non-consensus component should be submitted in a manner consistent with the approach used for non-consensus IL-TRM Updates; (Page 78)

B. Programs Currently Covered in this Document

This document will be updated over time to cover a range of programs. To facilitate completion of part of the IL-NTG Methods sections prior to March 1, 2015, this document includes methods specific for three program types: 1) Commercial, Industrial, and Public Sector Standard/Prescriptive and Custom programs, 2) Appliance Recycling programs, and 3) Residential Upstream Lighting programs.

C. Updating the IL-NTG Methods

This attachment is part of the IL-TRM and follows the timeline for updating of the IL-TRM as specified in the IL-TRM Policy Document.¹ In general, the following will take place:

- Updates will occur annually

¹ Policy Document for the Illinois Statewide Technical Reference Manual for Energy Efficiency. October 25, 2012. <http://www.icc.illinois.gov/downloads/public/IL%20TRM%20Policy%20Document.pdf>

- Updates will be discussed within the SAG and completed by March 1st
- The ICC Staff will then submit a Staff Report (with the consensus Updated TRM attached) to the Commission with a request for expedited review and approval.

D. Procedure for Non-Consensus Items

Non-consensus items that arise during the development and updating of the IL-NTG Methods document will be handled in substantially the same way as non-consensus IL-TRM Updates are addressed. The approach to be used is as follows.

- Once the Illinois NTG Working Group² has progressed as far as they can on the methodology, and it has been found that there is non-consensus on a specific Net-to-Gross Methods topic or procedure, the Illinois NTG Working Group shall submit to the ICC Staff and the Stakeholder Advisory Group's (SAG) Technical Advisory Committee (TAC) a Comparison Exhibit of Non-Consensus Net-to-Gross Methods topics/procedures *within 1 week* after the Illinois NTG Working Group has failed to reach consensus. The TAC will then deliberate on the issue with a goal of reaching consensus.
- If consensus does not emerge in the TAC regarding a particular Net-to-Gross Methods topic or procedure, the Comparison Exhibit of Non-Consensus NTG Methods topics/procedures is then sent to the full SAG for their deliberations and input. The SAG provides a forum where experts on all sides of the contested issue can present their expert opinions in an effort to inform parties of the contested issue and to also facilitate consensus.
- If the full SAG is unable to reach consensus, the non-consensus item will be referred to the ICC for resolution at the time of the IL-TRM Update proceeding. After receipt of the Comparison Exhibit of Non-Consensus Net-to-Gross Methods topics/procedures, the ICC Staff will submit a Staff Report to the Commission to initiate a proceeding separate from the consensus IL-TRM Update proceeding to resolve the non-consensus Net-to-Gross Methods topics/procedures.

II. Attribution in Energy Efficiency Programs in General

One of the most difficult aspects of evaluation, and not just within evaluation of energy efficiency programs, is attributing results to a program. Attribution provides credible evidence that there is a causal link between the program activities and the outcomes achieved by the program. Attribution research estimates the difference between the outcomes and those that would have occurred absent the program (i.e., the counterfactual). Put in research terms, evaluators must reject the null hypothesis of no causality through probabilistic statements (e.g., "strong evidence", "high probability"). As such, it is important to realize that the concept of the counterfactual cannot be proven with certainty. So even though the NTG ratio is a single value, conceptually it is a probabilistic statement³. One of the main academics within evaluation stated that there is a "...total and inevitable absence of certain knowledge [arising] from the methods social scientists use" when assessing the counterfactual. (Shadish, et al. 2002) This statement is not about poor methods, but about the counterfactual itself. Because programs work with people and are not a laboratory experiment that can be replicated over and over, to find out what actions people would have taken absent an intervention, one would need a time machine to take people back in time

² The Illinois NTG Working Group consists primarily of the subset of Evaluators deliberating on NTG methodologies; however, any interested party may participate in the Illinois NTG Working Group.

³ A probabilistic statement is not the same as the confidence and precision information calculated based on sampling theory.

and not provide the program. Since time machines do not exist, evaluators have developed methods that approximate the counterfactual to the best of their ability.

For energy efficiency programs, evaluators differentiate between savings at a “gross” and “net” level as described below in the short set of relevant definitions. Research to determine attribution occurs to allow for a better understanding of the net level of savings.

Relevant Definitions:

Concept	Term	Definition
Consumers	Nonparticipant	Any consumer who was eligible but did not participate in the subject efficiency program, in a given program year.
	Participant	A consumer that received a service offered through the subject efficiency program, in a given program year; also called <i>program participant</i> . The term “service” is used in this definition to suggest that the service can be a wide variety of inducements, including financial rebates, technical assistance, product installations, training, energy efficiency information or other services, items, or conditions. Each evaluation plan should define “participant” as it applies to the specific evaluation.
Impacts	Gross Impacts	The change in energy consumption and/or demand that results directly from program-related actions taken by participants in an energy efficiency program, regardless of why they participated.
Attribution of Impacts	Net Impacts	The change in energy consumption and/or demand that is attributable to a particular energy efficiency program. This change in energy use and/or demand may include, implicitly or explicitly, consideration of factors such as free ridership, participant and nonparticipant spillover, and induced market effects. These factors may be considered in how a baseline is defined (e.g., common practice) and/or in adjustments to gross savings values.
	Net-to-Gross Ratio	A factor representing net program savings divided by gross program savings that is applied to gross program impacts to convert them into net program impacts. The factor itself may be made up of a variety of factors that create differences between gross and net savings, commonly including free riders and spillover. The factor can be estimated and applied separately to either energy or demand savings.
	Free Rider	A program participant who would have implemented the program’s measure(s) or practice(s) in the absence of the program. Free riders can be (1) total, in which the participant’s activity would have completely replicated the program measure; (2) partial, in which the participant’s activity would have partially replicated the program measure; or (3) deferred, in which the participant’s activity would have partially or completely replicated the program measure, but at a future time.
	Spillover	Reductions in energy consumption and/or demand caused by the presence of an energy efficiency program, beyond

Concept	Term	Definition
		<p>the program-claimed gross savings of the participants. There can be participant and/or nonparticipant spillover. <i>Participant spillover</i> is the additional energy savings that occur as a result of the program's influence when a program participant independently installs incremental energy efficiency measures or applies energy-saving practices after having participated in the energy efficiency program. <i>Nonparticipant spillover</i> refers to energy savings that occur when a program nonparticipant installs energy efficiency measures or applies energy savings practices as a result of a program's influence.</p>
<p>Markets</p>	<p>Market</p>	<p>The commercial activity (e.g., manufacturing, distributing, buying, and selling) associated with products and services that affect energy use.</p>
	<p>Market Effects</p>	<p>A change in the structure of a market or the behavior of participants in a market that is reflective of an increase (or decrease) in the adoption of energy efficient products, services, or practices and is causally related to market interventions (e.g., programs). Examples of market effects include increased levels of awareness of energy efficient technologies among customers and suppliers, increased availability of energy efficient technologies through retail channels, reduced prices for energy efficient models, build out of energy efficient model lines, and—the end goal—increased market share for energy efficient goods, services, and design practices.</p>
	<p>Market Assessment</p>	<p>An analysis that provides an assessment of how and how well a specific market or market segment is functioning with respect to the definition of well-functioning markets or with respect to other specific policy objectives. A market assessment generally includes a characterization or description of the specific market or market segments, including a description of the types and number of buyers and sellers in the market, the key actors that influence the market, the type and number of transactions that occur on an annual basis, and the extent to which market participants consider energy efficiency an important part of these transactions. This analysis may also include an assessment of whether a market has been sufficiently transformed to justify a reduction or elimination of specific program interventions. Market assessment can be blended with strategic planning analysis to produce recommended program designs or budgets. One particular kind of market assessment effort is a baseline study, or the characterization of a market before the commencement of a specific intervention in the market for the purpose of guiding the intervention and/or assessing its effectiveness later.</p>

Source: Derived from State and Local Energy Efficiency Action Network. 2012. *Energy Efficiency Program Impact Evaluation Guide*. Prepared by Steven R. Schiller, Schiller Consulting, Inc., www.seeaction.energy.gov.

III. Attribution within the Commercial, Industrial, and Public Sectors

Over thirty programs across 12 types of Commercial, Industrial, or Public Sector programs are expected to be offered in Illinois in electric program year 8 (EPY8) for electric utilities and gas program year 5 (GPY5) for gas utilities (i.e., June 2015 – May 2016). The evaluation team has worked partially through the NTG method for the Commercial & Industrial (C&I) and Public Sector Standard/Prescriptive and Custom programs. Future updates to this document will include a full NTG method for these programs as well as other programs.

A. Standard/Prescriptive and Custom Programs

All C&I and Public Sector Standard/Prescriptive and Custom programs offered in Illinois in GPY5/EPY8 are similar enough in scope and implementation to fall under the consistent methods outlined in this section. The detail drafted below documents agreements reached by the evaluation teams through approximately 10 hours of discussion spread out over five meetings which began in October 2014 and continued through early January 2015. Additionally, evaluators spent considerable amount of time prior to official meetings delving into NTG details. Consensus reached so far pertains to the self-report approach and is documented below.

1. Free Ridership

There have been several core agreements reached by the evaluation teams. These agreements should reduce potential methodological differences employed by different evaluation contractors. Each is bulleted below.

- **Multiple Questions:** Evaluators will use program participant responses to multiple survey questions as inputs to the free ridership calculation algorithm. Evaluators will not use the response to a single question to establish a survey respondent as either a complete free rider or a complete non-free rider.
- **Program and Non-Program Factors:** Evaluators will administer survey questions to obtain respondent ratings on a numeric scale of the impact, influence, or importance on the decision to implement energy efficiency measures or take energy efficiency actions. A series of questions will focus on factors that the evaluator determines are a function of the program. Such program factors may, for instance, include availability of the program incentive, technical assistance from program staff, program staff recommendations, program-administrator marketing materials, and endorsement or recommendation by utility account manager or program partner staff. Previous experience with the program is not a program factor for purposes of obtaining respondent ratings of program impact, influence, or importance on the decision to implement energy efficiency measures. Evaluators will also administer a series of questions to obtain respondent ratings on a numeric scale of the impact, influence, or importance on the decision to implement energy efficiency measures or factors that the evaluator determines are not a function of the program. Such non-program factors may include, for example, age or condition of existing equipment, previous experience with the measure, standard business or industry practice, and organizational policy or guidelines.
- **Mediation of Numeric Scales:** Evaluators will administer survey questions referencing numeric scoring scales for the purpose of quantifying free ridership. The numeric scales shall be based upon 11 points ranging from 0 to 10. Survey respondent numeric scores obtained from the administration of these questions will serve as inputs to the applied free

ridership calculation algorithm. In calculating free ridership, survey respondent numeric scores may be mediated by other algorithmic components.

- **Vendor Recommendations:** Equipment vendor or contractor recommendations may also be a program factor to the extent that such recommendation is a function of the program. The evaluator may administer survey questions to vendors or contractors to verify their involvement with participant projects and to obtain respondent ratings – on a numeric scale – of the impact, influence, or importance of the program on the decision to recommend the energy efficiency measure(s) to the program participant.
- **Counterposing Program and Non-Program Factors:** Evaluators will administer a survey question that asks respondents to quantify the impact, influence, or importance on the decision to implement energy efficiency measures of factors that the evaluator determines are a function of the program relative to factors that the evaluator determines are not a function of the program.
- **Likelihood to Implement:** Evaluators will administer a survey question to obtain respondent ratings on a numeric scale of the likelihood of the respondent, in the absence of the program, to implement specified energy efficiency measures. The evaluator may administer questions to collect respondent self-report data regarding the respondent course of action, in the absence of the program, relating to the likelihood and timing of implementation, project scope, and measure characteristics.
- **Consistency Checks:** Evaluators should administer survey questions as checks on the consistency of responses associated with a core free ridership assessment methodology. Evaluators may also reference available data, including consistency check data, to perform documented modifications to individual free ridership estimates resulting from the application of a core free ridership assessment methodology.

The survey questions referenced above constitute basic guidelines for evaluators to use in the development and application of a core free ridership assessment methodology – these survey questions are not all encompassing and other survey questions may be asked by evaluators.

a) Algorithm

The evaluation teams have not yet reached agreement on the specific algorithm to use. There have been thoughtful discussions around the status quo algorithms, multiplying specific inputs rather than averaging them, and including partial free ridership through a very different approach of time-varying free ridership values⁴.

⁴ Within time-varying free ridership, free ridership may vary over the course of measure life due to respondents' self-reported timing of implementing actions under the counterfactual scenario (i.e., absence of the program). Free ridership may also vary based on project scope and measure characteristics associated with respondents' self-reported actions under the counterfactual no-program scenario. As stated above, evaluators may, on a pilot basis, separately calculate the free ridership rate applicable to annualized first year gross energy savings and the free ridership rate applicable to gross energy savings occurring over the lifetime of implemented measures.

The evaluation teams will continue discussions in 2015 with the intent of using future evaluations to pilot the algorithms.

2. Spillover

Spillover has not yet been discussed by the evaluation teams in terms of reaching consensus on spillover methods. Future methods will be informed by current spillover study results.

IV. Attribution within the Residential and Low Income Sectors

Over 30 programs across 13 types of Residential programs are expected to be offered in Illinois in EPY8 for electric utilities and GPY5 for gas utilities (i.e., June 2015 – May 2016) Appliance Recycling

A. Appliance Recycling Programs

Appliance recycling programs (ARPs) typically offer some mix of incentives and free pickups for the removal of old-but-operable refrigerators, freezers, or room air conditioners. These programs encourage consumers to undertake the following:

- Discontinue use of secondary or inefficient appliances;
- Relinquish appliances previously used as primary units upon their replacement (rather than keeping the old appliance as a secondary unit); and
- Prevent the continued use of old appliances in other households through direct transfers (i.e., giving it away or selling it) or indirect transfers (resale in the used appliance market).

As the program theory and logic for appliance recycling differ significantly from standard “downstream” incentive programs (which typically offer rebates for purchases of efficient products), the free ridership estimation approach also significantly differs.

There are basic and enhanced methods described next.

Basic Method

1. Free Ridership

Free ridership is based on participants’ anticipated plans had the program not been available, thus classifying a free rider as a participant who would have removed the unit from service regardless of the program.

Estimating net savings for ARPs should adopt a multistep process to segment participants into different groups, each with specific attributable savings.

In general, independent of program intervention, participating appliances would have been subject to one of the following options:

1. The appliance would have been kept by the participating household.
2. The appliance would have been discarded in a way that transfers the unit to another customer for continued use.
3. The appliance would have been discarded in a way that would have permanently removed the unit from service.

Only Option 3 constitutes free ridership (the proportion of units that would have been taken off the grid absent the program). Options 1 and 2 both indicate non-free riders. However, these

respondents need to be further classified to account for potential induced replacement and secondary market impacts, both described below.

A participant survey—drawn from a random sample of participants—will serve as the primary source of data collected for estimating NTG for the ARP. To determine the percentage of participants in each of the three options, evaluators will begin by asking surveyed participants about the likely fate of their recycled appliance had it not been decommissioned through the utility program. Responses provided by participants generally can be categorized as follows:

1. Kept the refrigerator
2. Sold the refrigerator to a private party (either an acquaintance or through a posted advertisement)
3. Sold or gave the refrigerator to a used-appliance dealer
4. Gave the refrigerator to a private party, such as a friend or neighbor
5. Gave the refrigerator to a charity organization, such as Goodwill Industries or a church
6. Had the refrigerator removed by the dealer from whom the new or replacement refrigerator was obtained
7. Hauled the refrigerator to a landfill or recycling center
8. Hired someone else to haul the refrigerator away for junking, dumping, or recycling.

Additional, follow-up questions will be included to validate the viability of all responses.

Next evaluators will assess whether each participant's final response indicates free ridership.

- Some final responses clearly indicate free ridership, such as: "I would have taken it to the landfill or recycling center myself."
- Other responses clearly indicate no free ridership, as when the refrigerator would have remained active within the participating home ("I would have kept it and continued to use it") or used elsewhere within the utility's service territory ("I would have given it to a family member, neighbor, or friend to use").

If the respondent planned to have the unit picked up by the retailer and the retailer would likely resell the unit in the secondary market, they are not a free rider. Absent retailer survey primary research described in the Enhanced Options below, the evaluators will utilize data from the most recent research conducted of the ComEd program to determine the proportion of free riders unless another metric is mutually agreed upon.⁵

2. Secondary Market Impacts

In the event that the unit would have been transferred to another household (Option 2 above), the question then becomes what purchasing decisions are made by the would-be acquirers of participating units now that these units are unavailable. These would-be acquirers could:

⁵ Note that such retailer interviews are being conducted annually for the ComEd ARP evaluation, and answers are used directly in the calculation of the NTG ratio in cases where: (1) the respondent planned to have the unit picked up by the retailer; and (2) the retailer was interviewed.

1. Not purchase/acquire another unit
2. Purchase/acquire another used unit.

Adjustments to savings based on these factors are referred to as the program's secondary market impacts.

If it is determined that the participant would have directly or indirectly (through a market actor) transferred the unit to another customer on the grid, the next question addresses what that potential acquirer did because that unit was unavailable. There are three possibilities:

A. None of the would-be acquirers would find another unit. That is, program participation would result in a one-for-one reduction in the total number of refrigerators operating on the grid. In this case, the total energy consumption of avoided transfers (participating appliances that otherwise would have been used by another customer) should be credited as savings to the program. This position is consistent with the theory that participating appliances are essentially convenience goods for would-be acquirers. (That is, the potential acquirer would have accepted the refrigerator had it been readily available, but because the refrigerator was not a necessity, the potential acquirer would not seek out an alternate unit.)

B. All of the would-be acquirers would find another unit. Thus, program participation has no effect on the total number of refrigerators operating on the grid. This position is consistent with the notion that participating appliances are necessities and that customers will always seek alternative units when participating appliances are unavailable.

C. Some of would-be acquirers would find another unit, while others would not. This possibility reflects the awareness that some acquirers were in the market for a refrigerator and would acquire another unit, while others were not (and would only have taken the unit opportunistically).

The evaluators will assume Possibility C unless primary research to assess the secondary appliance market is undertaken as described in the Enhanced Options below. Specifically, evaluators will assume that half (0.5, the midpoint of possibilities A and B) of the would-be acquirers of avoided transfers found an alternate unit.

Once the proportion of would-be acquirers who are assumed to find alternate units is determined, the next question is whether the alternate unit was likely to be another used appliance (similar to those recycled through the program) or, with fewer used appliances presumably available in the market due to program activity, would the customer acquire a new standard-efficiency unit instead.

Again, unless primary research is undertaken as described in the Enhanced Options below for an assessment of the appliance market, evaluators will apply a midpoint approach assuming half (0.5) of the would-be acquirers of program units would find a similar, used appliance and half (0.5) would acquire a new, standard-efficiency unit.

3. Induced Replacement

If, however, the unit would have been kept by the participating household, the next question is whether the appliance was replaced and, if so, whether the household would have replaced the appliance regardless of the program.

The purchase of a refrigerator in conjunction with program participation does not necessarily indicate induced replacement. (The refrigerator market is continuously replacing older refrigerators with new units, independent of any programmatic effects.) However, if a customer would have not purchased the replacement unit (put another appliance on the grid) in the

absence of the program, the net program savings should reflect this fact. This is, in effect, akin to negative spillover and will be used to adjust net program savings downward.

Estimating the proportion of households induced to replace their appliance can be done through participant surveys. As an example, participants could be asked, “Would you have purchased your replacement refrigerator if the recycling program had not been offered?”

Because an incentive ranging from \$35 to \$50 is unlikely to be sufficient motivation for purchasing an otherwise-unplanned replacement unit (which can cost \$500 to \$2,000), it is critical that evaluators include a follow-up question. That question should confirm the participants’ assertions that the program alone caused them to replace their refrigerator. For example, participants could be asked, “Let me be sure I understand correctly. Are you saying that you chose to purchase a new appliance because of the appliance recycling program, or are you saying that you would have purchased the new refrigerator regardless of the program?”

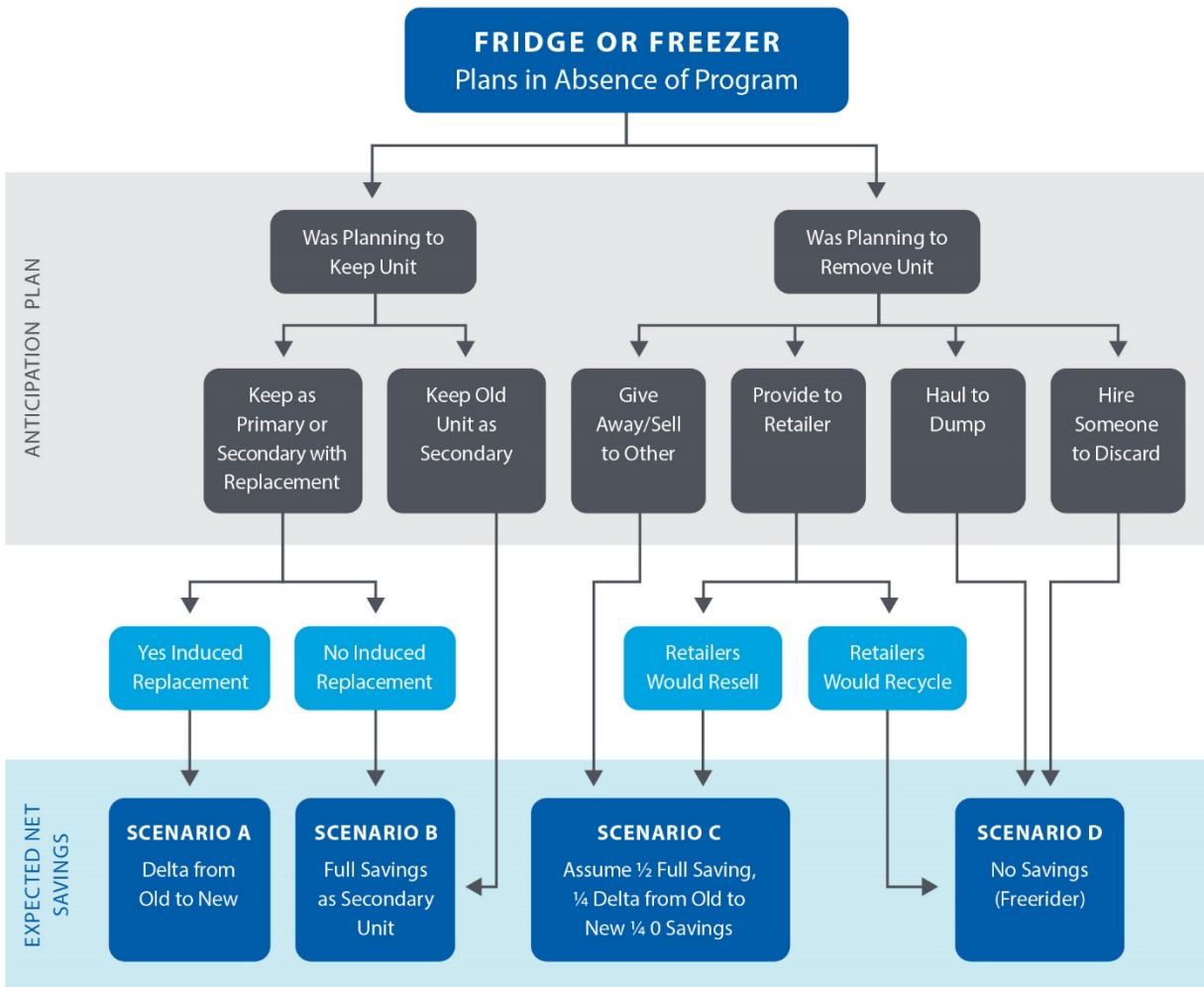
When assessing participant survey responses to calculate induced replacement, evaluators will consider the appliance recycled through the program, as well as the participant’s stated intentions in the absence of the program. For example, when customers indicated they would have discarded their primary refrigerator independent of the program, it is not possible that the replacement was induced (because it is extremely unlikely the participant would live without a primary refrigerator). Induced replacement is a viable response for all other usage types and stated intention combinations.

As one might expect, previous evaluations have shown the number of induced replacements to be considerably smaller than the number of naturally occurring replacements unrelated to the program. Once the number of induced replacements is determined, this information is combined with the energy consumption replacement appliance to determine the total energy consumption induced by the program (on a per-unit basis).

4. Integrating Free Ridership, Secondary Market Impacts, and Induced Replacement

The flow chart shown in Figure 1 illustrates how net savings will be derived for an ARP. As shown, below, expected savings fall into four different scenarios.

Figure 1. Appliance Retirement Scenarios



Algorithm

Net savings will be assigned individually to each respondent, based on responses provided to the questions discussed above. Net savings will be averaged across all respondents to calculate program-level net savings. The following equation will be used:

$$FR = (\text{freeridership and secondary market impacts \%} - \text{induced replacement\%})$$

Table 2 demonstrates the proportion of a sample population classified into each of the seven potential categories and the resulting weighted net savings.

Table 2. Net Savings Example for a Sample Population*

Primary Classification	Secondary Classification	Tertiary Classification	Population (%)	UEC (kWh) w/out Program	UEC (kWh) w/ Program	kWh Savings
Would have kept unit	Scenario A: Kept but Induced Replacement	Non-ES unit	3%	1,026	520	506
		ES unit	2%	1,026	404	622

Primary Classification	Secondary Classification	Tertiary Classification	Population (%)	UEC (kWh) w/out Program	UEC (kWh) w/ Program	kWh Savings
	Scenario B: Kept but NO Induced Replacement	N/A	25%	1,026	0	1,026
Would have removed unit	Scenario C: Transferred	Unit age >= 10 years	12.5%	0	0	0
		Unit age < 10 years	12.5%	1,026	520	506
		No replacement	25.0%	1,026	0	1,026
	Scenario D: Removed from service	N/A	20%	0	0	0
Net Savings (kWh)						604

*The percent values presented in this table serve only as examples; actual research should be conducted to determine the percentage of units falling into each of these categories. Note that UEC values presented in the table represent default values from the 2014 TRM, factoring in part-use.

Enhanced Method

Results can be enhanced by including three additional research efforts. The basic method has defaults where primary research on enhanced approaches cannot be performed:

1. A retailer survey, to determine the quantity and/or proportion of units returned to a retailer, and that the retailer would deconstruct or recycle. Through this survey, one would determine a retailer’s criteria for reselling used units vs. deconstructing them, based on unit age and condition. Results from the survey and analysis would be used to determine the proportion of those who would have returned an old appliance to the retailer that should be included in Scenario D (free riders). This research was conducted for ComEd in PY6 evaluation and those results were applied to Ameren.
2. An appliance market assessment study, to determine the size of the secondary appliance market and whether removal of participating units from the market would cause an otherwise would-be receiver to purchase an alternative used or new unit. Savings attributable to these participants are the most difficult to estimate, as the scenario attempts to estimate what the prospective buyer of a used appliance would do in the absence of finding a program-recycled unit in the marketplace (i.e., the program took the unit off the grid, so the prospective purchaser faced, in theory, a smaller supply of used refrigerators). It is difficult to answer this question with certainty, absent utility-specific information regarding the change in the total number of refrigerators (overall and used appliances specifically) that were active before and after program implementation. In some cases, evaluators have conducted in-depth market research to estimate both the program’s impact on the secondary market and the appropriate attribution of savings for this scenario. Although these studies are imperfect, they can provide utility-specific information related to the program’s net energy impact. Where feasible, evaluators and utilities should design and implement

such an approach. Unfortunately, this type of research tends to be cost-prohibitive, or the necessary data may simply be unavailable.

3. However, it is possible to estimate through nonparticipant surveys which of the disposal responses given by nonparticipants were most likely to have been to an opportunistic would-be-acquirer. Transfers would most likely have been opportunistic are determined primarily based on the cost to the recipient. If the appliance was sold or transferred to a retailer, there would have been a cost to the recipient of that appliance. If the recipient was willing to pay for the appliance or was willing to exert the effort to visit a retail location, which suggests the recipient was actively seeking an appliance. However, if the unit were given away for free there was little cost to the recipient and is a reasonable proxy for the proportion of opportunistic acquirers rather than the 50% split assumed in DOE's Uniform Methods Protocols and previously in this protocol.
4. A nonparticipant survey can be used to assess how nonparticipants acquire and dispose of used units. As nonparticipants do not have the same perceived response bias as participants, they can help offset some of this potential bias in estimating the true proportion of the population that would have recycled their units in program's absence. The evaluators will average the results of the nonparticipant survey with the participant survey if the nonparticipant survey is of sufficient sample size. Otherwise, results may be used for a qualitative characterization of potential bias. Though recommended, use of a nonparticipant survey need not be required, given budget and time considerations. A nonparticipant survey was completed as part of ComEd's PY6 evaluation and used qualitatively to validate participant results.

5. Spillover

Unlike many programs, recycling programs face reduced opportunities for spillover due to the lack of general energy education and the small likelihood of participants having further units to recycle on their own. This program could directly impact decisions to replace refrigerators or freezers with ENERGY STAR units rather than standard efficiency units, given that the program offers marketing and education related to the operating costs of refrigerators and freezers. Reliable methods of conducting this analysis have yet to be developed. One attempted method compared proportions of ENERGY STAR appliances replaced by program participants to proportions of ENERGY STAR new appliance shipments in a similar area. Due to the difficulty in isolating the shipment area to the program area, this has not yielded noticeable spillover in Illinois.

6. Nonparticipant Spillover

The specific approach and method for measuring spillover has not yet been discussed by the evaluation teams to reach a consensus. However, effective program marketing and outreach generates program participation and increases general energy-efficiency awareness among customers. The cumulative effect of sustained utility program marketing (which often occurs concurrently for multiple programs) can affect customers' perceptions of their energy usage and, in some cases, motivates customers to take efficiency actions outside of the utility's program. This phenomenon—called nonparticipant spillover (NPSO)—results in energy savings. Marketing of the Appliance Recycling program specifically may induce nonparticipants to either reduce the use of the secondary refrigerator or freezer that they keep, or when they are purchasing a new refrigerator or freezer, to buy one that is more energy efficient.

B. Residential Upstream Lighting Programs

The Illinois Residential Upstream Lighting programs to date have provided discounts on efficient lighting through retailers at the point of purchase. Such programs often remain transparent to

customers purchasing incentivized lighting. Program administrators also do not know the identity of most customers purchasing the program-discounted lighting; so these customers cannot easily be contacted once they leave the store for a traditional self-report net-to-gross (NTG) evaluation survey (i.e., an after-the-fact, direct solicitation of customers regarding what they would have done in the program's absence). Similar surveys can be conducted with customers within program retailers after they have made their lighting purchasing decision but before they leave the store, however this data collection method presents a number of logistical challenges.

Further, as upstream programs work with multiple market actors and can include wide-reaching marketing campaigns promoting energy efficiency to the general public, they tend to stimulate spillover and "market effects." As a result, estimating NTG for upstream residential lighting programs can be challenging. Multiple methods exist, each with their own strengths and weaknesses.

Ameren and ComEd implement their residential lighting programs comparably, and the evaluation teams have used a consistent primary NTG evaluation method. This section details the consensus NTG methodology, which has been used multiple times for both ComEd and Ameren and is considered the most well-vetted and defensible NTG method that has been successfully used in Illinois.

For EPY5 and EPY6, Ameren and ComEd used a customer self-report methodology to estimate NTG for their upstream residential lighting programs.⁶ Customer self-report data in this method are collected during surveys conducted within program retailers with customers purchasing program bulbs (i.e., in-store intercept surveys). This method separately estimates free ridership, participant spillover, and nonparticipant spillover. Details follow on the primary data collection and scoring algorithms.

1. Free Ridership

a) Definition

Free ridership is the proportion of program bulbs that would have been purchased if the program did not exist. Three alternative scenarios could occur:

1. Full Free rider: The customer would have purchased the same quantity of efficient bulbs (CFLs or LEDs) in the program's absence.
2. Partial Free rider: The customer would have purchased fewer efficient bulbs (CFLs or LEDs) in the program's absence.
3. Non-Free rider: The customer would have not purchased any efficient bulbs (CFLs or LEDs) in the program's absence.

Free ridership is calculated as the average of two distinct scores: a program influence score and a non-program score. These scores are defined as follows:

1. The *program influence score* captures the maximum level of program influence, reported by a survey respondent, of the residential lighting program on their decisions to purchase program bulbs on the day of the survey. This program influence can take a number of forms, such as: the monetary incentive provided to decrease the cost of high-efficiency bulbs; program-sponsored educational materials that explain the benefits of

⁶ ComEd has used this method since PY2. Ameren began using it in PY5.

efficient lighting; in-store product placement of efficient bulbs; and program bulb recommendations provided by retail store personnel.

2. The *non-program* score is used to estimate how many program bulbs a survey respondent would have purchased in the absence of the residential lighting program.

b) Data Collection

To estimate free ridership, the evaluation teams conduct in-store Intercept surveys with customers purchasing program-discounted lighting at participating retailers. Customers are asked questions that are used to estimate a program influence score and a non-program score for each customer and efficient bulb type purchased.

Primary Program Influence Score Questions

1. Light bulb purchasing plans for current shopping trip (Yes/No)
2. If planning to purchase bulbs:
 - a. Bulb type (CFL, LED, Incandescent, Halogen)
 - b. Utility-incentivized bulbs (Yes/No)
3. Influence of various program factors:
 - a. Program incentive
 - b. In-store information (printed materials or information from utility representatives or retail personnel)
 - c. Positioning of discounted bulbs within the store

Primary Non-Program Score Questions

1. Stated preference of light bulb purchases had the utility incentive not been available (purchase all, some or none of efficient bulbs)
2. Quantity of light bulbs purchased absent the utility incentive

c) Scoring Algorithms

Using the data collected from program participants during the in-store intercept surveys, program influence and non-program scores are calculated for each survey respondent and then combined to estimate a respondent-specific free ridership score.

Calculation of the Program Influence Score:

Survey respondents purchasing one or more program-discounted bulbs are assigned a preliminary program influence score based on the maximum program influence level (on a 0 to 10 scale) they assigned to one or more program factors (e.g., monetary incentive/informational materials (printed or from store personnel)/product positioning). The influence level assigned to the monetary incentive was increased for survey respondents (using a linear decreasing function⁷) who indicated that absent the incentive they would not have purchased any of the program bulbs they were purchasing that day.

⁷ The function, adjusted monetary score = (monetary score + 10)/2, increased the monetary score using a decreasing linear function. This function resulted in an increase in the monetary influence score of between 0 and 5 points depending on their original monetary score (i.e. an original score of 0 would become a 5, a 5 would become a 7.5, and a 10 would remain a 10. In the past years this adjustment has typically changed less than 10% of all monetary scores.

After the preliminary program influence score is assigned, a secondary algorithm is run that adjusts the preliminary program influence based on survey data regarding the customers purchasing plans when they entered the store. Survey respondents who indicated they planned to purchase high-efficiency bulbs prior to entering the store and had who not come to the store specifically to buy utility-incentivized program bulbs, have their program influence score cut in half. This adjustment makes the final program influence score reflective of their stated planned intention to purchase efficient bulbs in the program's absence.

Calculation of the Non-Program Score:

The non-program score is based on whether a respondent stated they would have purchased all, some, or none of the program-discounted bulbs in the absence of utility incentives. Respondents reporting they would have purchased all of the efficient bulbs without the incentive should be considered free riders and receive a non-program score of zero. Those reporting they would have purchased none of the efficient bulbs without the incentives should be classified as non-free riders and receive a non-program score of 10, the maximum. Respondents reporting they would have purchased some of the CFLs without the incentive should be assigned a non-program score between 0 and 10, reflective of the percentage of efficient bulbs they would not have purchased absent the program.

Respondents reporting they would have purchased all of the program-discounted bulbs in the program's absence, but in-store materials provided by the utility had a moderate to high influence on their decision should have their non-program scores adjusted to equal the level of influence they attributed to these program-sponsored informational materials.

Calculation of Free ridership:

$$\text{Free ridership} = 1 - (\text{Program Influence Score} + \text{Non-Program Score})/20$$

Using the calculated program influence and non-program scores, free ridership is calculated as one minus the sum of the two scores (program influence score plus non-program score), divided by 20. Dividing the sum of scores by 20 results in a ratio (between 0 and 1) that is representative of the average of the two zero to 10 scores. Subtracting this ratio from one reverses the score, thus representing the free ridership level. If either the non-program or program influence scores are missing, free ridership can be calculated using the single available score divided by 10. Evaluators may also reference available data to perform documented modifications to individual free ridership estimates resulting from the application of a core free ridership assessment methodology

2. Participant Spillover

a) Definition

Participant spillover results from purchases of non-discounted efficient bulbs by program bulb purchasers' who are influenced by their participation in the residential lighting program to purchase additional non-discounted efficient bulbs.

b) Data Collection

Data collected during in-store intercept surveys with customers purchasing program bulbs are used to estimate participant spillover. During these surveys, customers purchasing program-discounted and non-discounted efficient bulbs are asked questions to determine whether the residential lighting program influenced their purchases of non-discounted efficient bulbs.

c) **Scoring Algorithms**

To estimate participant spillover, the number of program-influenced, non-discounted efficient bulbs purchased by program participants is divided by the total number of program bulbs purchased by these program participants. This resulted in the participant spillover rate.

3. Nonparticipant Spillover

a) **Definition**

Nonparticipant spillover results from purchases of non-discounted efficient bulbs by customers who are not purchasing program-discounted bulbs, but report that the residential lighting program influenced their decision to purchase non-discounted efficient bulbs.

b) **Data Collection**

Data collected during in-store Intercept surveys with customers purchasing efficient bulbs not discounted by the program was used to estimate nonparticipant spillover. During these surveys, customers purchasing non-discounted efficient bulbs are asked questions to determine whether the residential lighting program influenced their purchases of non-discounted efficient bulbs.

c) **Scoring Algorithms**

To estimate nonparticipant spillover, one must first calculate the number of program-influenced, non-discounted efficient bulbs purchased by the population of program nonparticipants surveyed. This yields a survey nonparticipant spillover rate. This rate is then extrapolated to the estimated population of nonparticipating utility customers to determine the estimated total quantity of non-program efficient bulbs being purchased within the utility service territory. Dividing this result by the total number of program bulbs results in the nonparticipant spillover rate.

4. Method Advantages and Disadvantages

The in-store intercept method described above has certain advantages and disadvantages.

Advantages: This approach catches customers at their point of purchase, before they leave the store and can no longer be contacted directly. Given the interview's timing, customers can more easily recall price factors leading to their purchase choices. Also, as customers are intercepted at the store rather than surveyed by telephone, a higher cooperation rate results.

Disadvantages: Customers may not fully connect the impact that in-store education, product placement, and advertising have on their decision making. While many consumers believe they are not influenced by advertising, retailers know advertising and product placement work. Further, store intercepts typically must be coordinated with education events, and many retailers do not allow interviews to take place in their stores. Consequently, results are not based on random samples of customers purchasing program-discounted lighting throughout the year and across all participating retailers, which could bias the results.

Additionally, there are several alternative methods that can be used to estimate NTG for upstream lighting programs. Evaluation teams can use these alternative methods in addition to the method outlined above to provide additional information on program performance. Prior to the use of these alternative methods, evaluation teams should be in agreement on the use and execution of the method.

V. Appendix A: Overview of NTG Methods

The evaluation team presents information in this appendix to provide a relatively quick overview of NTG methods for readers unaccustomed to the possible methods that evaluators may deploy. It is not meant to be a complete or deep discussion about each of the methods presented. However, the evaluators in Illinois considered the inclusion of this appendix to be very important in acknowledging the current suite of methods deployed by evaluators throughout the U.S. and giving a framework for work within Illinois.

Much of the information shown below is taken directly a single source — the national Uniform Methods Project, Chapter 17: Estimating Net Savings: Common Practices. (Violette and Rathbun, 2014) This document has done a nice job of summarizing the eight most common attribution methods currently in use across the U.S. The evaluation team recommends that readers go first to this reference for further information. Additionally, while there are slightly over 100 references within the Violette and Rathbun document, other non-duplicative references are included where reasonable as additional resources for those interested in further research into any specific method.

A. Survey-based approaches

Virtually all Illinois based evaluations use a survey-based approach for programs where primary data is used to determine net savings. (The main exception is for Behavioral programs which use statistical analysis based on a randomized control trial program design.) Survey based approaches obtain data from program participants and nonparticipants using a structured data collection instrument implemented via phone, in person or on-line. At times, evaluators create and use an unstructured depth-interview guide to collect information about attribution and this provides both contextual data and quantitative data about a given project.

1. Self Report Approach

The self-report approach relies on the abilities of customers to discuss the program influence as well as the somewhat abstract ideas of the counterfactual (i.e., what would have occurred absent the program) after making a choice to purchase an energy efficient item or take an energy efficient action unrelated to a purchase. For program participants, this could include doing nothing (i.e., leaving the existing equipment as-is), installing the same energy efficient equipment as they did through the program, or an intermediate step of installing equipment that is more efficient than what they had in place previously, but less efficient than what they installed through the program. Evaluators also use this approach when collecting information from trade allies or distributors. This self-report approach is not new, nor is it exclusively used by the energy efficiency industry. An important attribute of this approach is its reliance on well-designed and fielded survey questions, so that the data underlying subsequent analyses are accurate and complete.

The output of this approach is a NTG ratio which can be considered an index of the program's influence on the decision to install energy efficient equipment. The NTG ratio is applied to gross savings in order to obtain an estimate of net savings. The NTG ratio may include free ridership, spillover, or market effects, depending on the survey and analytical design. NTG ratios may be calculated at the measure, suite of measures, or program level and are typically average values weighted by savings. If sufficient information is available, analysis of NTG ratios among certain customer segments may be done to further inform changes to program design.

References

- Sudman, 1996
- Stone, et al. 2000

- Bradburn, et al. 2004

2. Econometric/Revealed Preference Approach

The econometric/revealed preference approach, while still considered a survey approach due to how data is collected, moves beyond asking people about the counterfactual and instead uses the observations of the evaluator to collect information for analysis of a NTG ratio. Within this approach, evaluators typically deploy similar sampling designs as for the self-report approach to collect data, but actively gather what a person is doing (i.e., what is being purchased in a store) to determine attribution.

B. Randomized control trials (RCT) and quasi-experimental designs

As mentioned earlier, evaluators deploy an RCT for estimating savings from the Behavioral programs within Illinois. Additionally, quasi-experimental designs (QED) have been used in the past in Illinois to estimate net savings from the upstream CFL program, and CFL, insulation and air sealing measures within the Home Performance with Energy Star program

RCT and QED use statistical analysis to determine regularities within the data that reveal net savings due to a program intervention⁸. The analytical design attempts to control for factors that can confound net analysis.⁹ When estimating net savings within both an RCT and QED, two groups are included within the analysis: 1) a group that has been exposed to (i.e., treated by) a program and 2) a group that has not been exposed to the program. Evaluators must carefully consider the choice of the non-exposed group (called a control group for RCTs or comparison group for QEDs).

RCT – This design must be integral to a program’s implementation. Without the ability to randomly assign customers to one group or another (or at least randomly encourage customers to participate in a program), the ability of the design to yield unambiguous estimates of net impacts is compromised. Evaluators often help design how a program is implemented and, if not involved at the outset, carefully review choices made by the implementation team.

QED – A QED may be designed after a program has been implemented. It relies on determination of an equivalent comparison group, which is often chosen based on energy use. QED is difficult to perform well within the commercial sector due to the heterogeneity of end uses within the sector.

The output of an RCT or QED is the average net savings for the population within the statistical model. Evaluators may also analyze the data to help understand the savings within specific known segments if sufficient information and data points are available.

References

- Mohr, 1995
- Shadish, Cook, Campbell 2002
- Scriven, 2008
- Donaldson, 2009

⁸ Net savings are calculated when a comparison or control group of non-treated customers are part of the design. Statistical analyses can also obtain gross savings.

⁹ Economists strongly support this approach, but among program evaluators, the idea that an RCT is a “gold standard” for attribution research has been hotly debated for decades.

C. Deemed or stipulated NTG ratios

A deemed (or stipulated) NTG ratio is a value known prior to implementing a program and applied to estimate net savings for that program in a certain year. The current NTG framework¹⁰ within Illinois means that at any given year, deemed NTG values are present for the programs implemented by the Investor Owned Utilities (IOUs). For DCEO programs starting in PY6, the deemed approach is not used; rather, their evaluation team collects data for participants within a specific year and calculates and applies a NTG ratio based on the results of that primary data collection effort.

Deemed or stipulated NTG ratios may be based on previous primary data collection, review of secondary data, or agreed to among stakeholders.

D. Common practice baseline approaches

For this method, the evaluation team estimates what a typical consumer would have done at the time of the project implementation. Essentially, what is “commonly done” becomes the basis for baseline energy consumption and calculation of net savings. No gross impacts are calculated in this approach. This baseline is defined as the counterfactual “i.e., what would have occurred absent the program” and has been referred to as current practice, common practice or industry standard practice. Evaluators determine these practices through multiple methods, but often can be from self-report or on-site audits. The difference between the energy use of measures installed in the program and the energy use associated with current practice is considered by some to be sufficiently close to the net savings.

This approach is not in use in Illinois, but is used elsewhere in the country such as the Pacific Northwest and Delaware.

E. Market analyses

Market analyses can be done in several ways. One example is an approach previously used in Illinois that relied on sales data for the CFL program offerings from several different jurisdictions and applied design and analysis from the QED method. For Illinois, evaluators applied this cross-sectional approach in PY3 in addition to the Self-Report method to calculate a NTG value for the residential upstream lighting program. However, they have not deployed this specific sales data method since then. Market analyses are often used in theory-driven evaluations of market transformation programs.

Other non-sales data market analyses can be postulated on changes specified in program logic such as: 1) changes in the number of energy efficient units manufactured, 2) changes in market actor behavior around promotion or stocking of energy efficient items, or 3) reduction in prices. The analyses involving non-sales data must make a clear link between the program intervention and the changes found in the market. Additionally, outside of Illinois, while evaluators have extrapolated the market changes to specific energy or demand reductions, this activity may be viewed as tenuous due to assumptions that evaluators must make within the analysis.

Illinois is in a position to begin to discuss market analyses and how specific research may be able to interpret changes that have occurred (or may occur in the future) because of the IOU interventions over the past six years. Market analyses can be backward looking through historical tracing, but is best used when the logic of an intervention is described and specific market metrics are tracked over time. This is a switch from the current annual evaluation of programs and has

¹⁰ See for discussion of the NTG Framework.

challenges that stakeholders would need to discuss and reach a consensus on an approach that works for Illinois.

F. Structured expert judgment approaches

Closely tied to market analysis, this approach is a way for evaluators to gather credible evidence of changes that arise due to the intervention of a program. When deployed, it is often used as a cost-effective approach to estimate market effects or reach agreement on a NTG value when several different types of evidence are available. . The key premise of this approach is the use of a select group of known experts that all stakeholders agree can provide unbiased information as well as having sufficient knowledge to judge what may have occurred absent a program intervention.

A Delphi Panel is an example of this approach where data is collected from two or more rounds of data collection (which can occur via email, internet, or in-person). A round is when experts make their thoughts known about a specific subject, the evaluation team synthesizes the data and provides this collated data back to the group to discuss again. Allowing the full experts to see how their peers think about a topic helps to move the group towards consensus.

To date, in Illinois, there has been little need for this approach. However, if more market analyses occur in the future, this is a valuable tool that can be deployed.

References

- Mosenthal, et al 2000
- Powell 2002

G. Program Theory-Driven Approach

This approach is not included in the Violette and Rathbun (2014) document as a high level method, but is discussed by the authors under the historical tracing method. The Illinois evaluators believe that it deserves at least a short discussion within this framework.

A program theory is the written narrative about why the activities of a program are expected to bring about change. Typically associated with this approach is the direct graphical explication of the linkages between activities, outputs, and outcomes through an impact logic model.¹¹

A theory-driven evaluation denotes “[A]ny evaluation strategy or approach that explicitly integrates and uses stakeholder, social science, some combination of, or other types of theories in conceptualizing, designing, conducting, interpreting, and applying an evaluation.” (Coryn 2011) Within this approach, the ultimate conclusions regarding the efficacy of a program are based on the preponderance of the evidence and not on the results of any single analysis. Coryn and colleagues systematically examined 45 cases of theory-driven evaluations published over a twenty-year period to ascertain how closely theory-driven evaluation practices comport with the key tenants of theory-driven evaluation as described and prescribed by prominent theoretical writers. One output from this analysis was the identification of the core principles and sub-principles of theory-driven evaluation. If interested, please review the reference under Coryn 2011.

As an approach, it is best used for complex programs and/or causal mechanisms that extend far into the future. . Evaluators collect evidence that supports or rejects hypotheses that are explicit in the logic model. The case for program attribution is strengthened based on the extent to which an evaluation shows that the expected changes occur. Additionally, the evaluation team may be able

¹¹ Evaluators may use logic models to show program processes as well, but this is a program flow chart, not an impact model.

to collect data collection that will answer questions about the longer term outcomes of a program. This type of data collection may be very similar to market tracking activities described briefly above in Section O.

This approach does not specifically estimate a NTG value, but program administrators can choose to keep, drop or change a program based on intermediary data. Regulators must be convinced that the logic of a program is sound and that the intermediary outcomes are causally linked to expected savings. It has not been used to date in Illinois.

References

- Weiss 1997
- Chen 2000
- Coryn 2011

H. Case Studies Design

Case studies are used extensively in social sciences as well as many other disciplines or practice-oriented areas such as political science, economics, education, and public policy. Case studies help to understand the how and why of a situation and typically retain a holistic aspect of real-life events. As such, they may be a useful approach to determine attribution. As with program theory design, though, the data collected and analyzed within a case study approach will not typically yield a specific NTG value, but can provide credible evidence and insight that supports or refutes the changes brought about by program intervention.

To be used to assess attribution, evaluators must carefully design case studies to assure they account for the threats to causality (i.e. internal validity) that arise in any design. While not typically thought of in this manner, case study design can address multiple types of validity such as construct, internal and external validity as well as assuring reliability. When establishing construct validity and reliability, evaluators must use multiple sources of evidence, create and maintain a study database, and maintain a “chain of evidence” within the analysis. Internal validity is shown through analytic tactics such as pattern matching, explanation building, addressing rival explanations, or using logic models. External validity centers on the ability to generalize the analytical findings to other similar situations. External validity may be shown through replication of findings.

This approach has not been used to date in Illinois.

References

- Yin, 2003
- Stake, 2006

VI. Appendix B: References

Bradburn, Norman, Sudman, Seymore, and Wansink, Brian. 2004. *Asking Questions. The Definitive Guide to Questionnaire Design – For Market Research, Political Polls, and Social and Health Questionnaires.* San Francisco, CA: Jossey-Bass.

Chen, Huey-Tsyh. 1990. *Theory-Driven Evaluations.* Sage Publications.

Coryn, Chris L.S., Norakes, Linday A., Westine, Carl D., Schröter, Daniela C. 2011. "A Systematic Review of Theory-Driven Evaluation Practice from 1990 to 2009". *American Journal of Evaluation*, 32(2): 199-226.

Donaldson, Steward I., Christie, Christina A., Mark, Melvin M. 2009. *What Counts as Credible Evidence in Applied Research and Evaluation Practice?* Los Angeles, CA: Sage Publications, Inc.

Mohr, Lawrence B. 1995. *Impact Analysis for Program Evaluation, Second Edition.* Thousand Oaks: SAGE Publications.

Mosenthal, Philip, Prah, Ralph, Neme, Chris, and Cuomo, Robert. 2000. *A Modified Delphi Approach to Predict Market Transformation Program Effects.* Proceedings from the 2000 ACEEE Conference.

Powell, Catherine. 2002. "The Delphi technique: myths and realities". *Journal of Advanced Nursing*. 41(4), 376-382.

Scriven, Michael. "A Summative Evaluation of RCT Methodology and an Alternative Approach to Causal Research". March 2008. *Journal of MultiDisciplinary Evaluation, Volume 5, Number 9. ISSN 1556-8180.*

Shadish, William R., Cook, Thomas D., and Campbell, Donald T. 2002. *Experimental and Quasi-Experimental Designs for Generalized Causal Inference.* Boston: Houghton Mifflin Company.

Stake, Robert E. 2006. *Multiple Case Study Analysis.* The Guilford Press.

Stone, Arthur A., Turkkan, Jaylan S., Bachrach, Christine A., Jobe, Jared B., Kurtzman, Howard S., Cain, Virginia S. editors. 2000. *The Science of Self-Report. Implications for Research and Practice.* Lawrence Erlbaum Associates, Inc.

Sudman, Seymour, Bradburn, Norman, Schwarz, Norbert. 1996. *Thinking about Answers. The Application of Cognitive Processes to Survey Methodology.* Jossey-Bass.

Violette, Daniel M., Rathbun, Pamela. 2014. "Chapter 17: Estimating Net Savings: Common Practices. The Uniform Methods Project: Methods for Determining Energy Efficiency Savings for Specific Measures". Available electronically at <http://www.osti.gov/scitech>.

Weiss, Carol H. 1998. *Evaluation Methods for Studying Programs and Policies, 2nd Edition.* Upper Saddle River, New Jersey: Prentice Hall.

Yin, Robert K. 2003. *Case Study Research. Design and Methods.* Thousand Oaks, CA: Sage Publications, Inc.