2019 Illinois Statewide Technical Reference Manual for Energy Efficiency

Version 7.0

Volume 4: Cross-Cutting Measures and Attachments

FINAL

September 28, 2018

Effective:

January 1, 2019

Table of Contents

VOLUME 1: OVERVIEW AND USER GUIDE

VOLUME 2: COMMERCIAL AND INDUSTRIAL MEASURES

VOLUME 3: RESIDENTIAL MEASURES

VOLUME 4: CROSS-CUTTING MEASURES AND ATTACHMENTS

6 CROSS-CUTTING MEASURES 6

6.1 Behavior 6

6.1.1 Adjustments to Behavior Savings to Account for Persistence 6

ATTACHMENT A: ILLINOIS STATEWIDE NET-TO-GROSS METHODOLOGIES 19

2 ATTRIBUTION IN ENERGY EFFICIENCY PROGRAMS IN GENERAL 23

2.1 Definitions 23

2.2 Spillover-Specific Issues 25

2.2.1 Measure Costs 26

3 COMMERCIAL, INDUSTRIAL, AND PUBLIC SECTOR PROTOCOLS 27

3.1 Core Non-Residential Protocol 28

3.1.1 Core Non-Residential Free Ridership Protocol 28

3.2 Core Non-Residential Spillover Protocol 34

3.2.1 Core Participant Spillover Protocol 34

3.2.2 Core Nonparticipant Spillover Protocol 37

3.3 Small Business Protocol 41

3.3.1 Free Ridership 41

3.4 C&I New Construction Protocol 42

3.4.1 Free Ridership 42

3.5 Study-Based Protocol 43

3.5.1 Free Ridership 43

3.6 Technical Assistance Protocol 45

3.6.1 Free Ridership 45

4 RESIDENTIAL AND LOW INCOME SECTOR PROTOCOLS 47

4.1 Residential Cross-Cutting Approaches 48

4.1.1 Survey Design Issues 48

4.1.2 Participant Spillover 48

4.1.3 Nonparticipant Spillover Measured from Customers 50

4.2 Appliance Recycling Protocol 52

4.2.1 Basic Method 53

4.2.2 Enhanced Method 56

4.3 Residential Upstream Lighting Protocol 57

4.3.1 Basic Method 57

4.3.2 Participant Spillover 60
4.3.3 Nonparticipant Spillover 61
4.4 Prescriptive Rebate (With No Audit) Protocol 63
4.4.1 Basic Method 63
4.5 Single-Family Home Energy Audit Protocol 65
4.5.1 Basic Method 65
4.6 Multifamily Protocol 68
4.6.1 Basic Method 69
4.7 Energy Saving Kits and Elementary Education Protocol 71
4.7.1 Basic Method 71
4.8 Residential New Construction Protocol 72
4.8.1 Basic Method 72
4.8.2 Participant Spillover 74
4.8.3 Builder Nonparticipant Spillover 75
5 CROSS-SECTOR PROTOCOLS 77
5.1 Combining Participant and Trade Ally Free Ridership Scores 77
5.1.1 Trade Ally Free Ridership Calculation 77
5.1.2 Triangulation 77
5.2 Spillover Measured Through Trade Allies 78
5.2.1 Spillover from Active Trade Allies 79
5.2.2 Spillover from Inactive Trade Allies 81
5.3 Consumption Data Analysis Protocol 82
5.3.1 Nonparticipant Spillover 84
5.3.2 Participant Spillover 84
5.3.3 Free Ridership 85
5.3.4 Consumption Data Analysis Designs with a Comparison Group 86
5.3.5 Randomized Controlled Trials 86
5.3.6 Random Encouragement Designs 86
5.3.7 Quasi-Experimental Designs 87
5.4 Code Compliance Protocol 91
5.4.1 Data Collection 91
5.4.2 Attribution Assessment 92
6 APPENDIX A: OVERVIEW OF NTG METHODS 93
6.3 Survey-Based Approaches 93
6.3.1 Self-Report Approach 93
6.3.2 Econometric/Revealed Preference Approach 94
6.4 NTG with Consumption Data Analysis 94

6.5 Deemed or Stipulated NTG Ratios 94
6.6 Common Practice Baseline Approaches 94
6.7 Market Analyses 94
6.8 Structured Expert Judgment Approaches 95
6.9 Program Theory-Driven Approach 95
6.10 Case Studies Design 96
7 Appendix B: References 97
Attachment B: Effective Useful Life for Custom Measure Guidelines 99

Table of Figures

Figure 2-1. Example - Types of Spillover and Methods for Assessment 26
Figure 3-1. Core Free Ridership Algorithm 1 29
Figure 3-2. Core Free Ridership Algorithm 2 30
Figure 3-3. Small Business Free Ridership 42
Figure 3-4. Study-Based Free Ridership—Overview 44
Figure 3-5. Study-Based Free Ridership—No-Program FR Score Option #1 44
Figure 3-6. Study-Based Free Ridership—No-Program FR Score Option #2 45
Figure 4-1. NPSO Question Logic 51
Figure 4-2. Appliance Retirement Scenarios 55
Figure 4-3. Residential Upstream Lighting Free Ridership via In-Store Intercept 58
Figure 4-4. Residential Upstream Lighting Participant Spillover Determination 60
Figure 4-5. Residential Upstream Lighting Nonparticipant Spillover Determination 62
Figure 4-6. Residential Prescriptive Rebate (With No Audit) Free Ridership 64
Figure 4-7. Single-Family Home Energy Audit Free Ridership—No Cost Measures 67
Figure 4-8. Single-Family Home Energy Audit Free Ridership—Discounted Measures 68
Figure 4-9. Multifamily Free Ridership—Non-CFL/Non-LED Measures 70
Figure 4-10. Multifamily Free Ridership for Property Managers—CFL/LED Measures 70
Figure 4-11. Energy Saving Kits and Elementary Education Free Ridership 71
Figure 4-12. Residential New Construction Free Ridership 73

Table of Tables

Table 1-1. ICC Energy Efficiency Dockets 19
Table 2-1. Definitions 23
Table 3-1. Commercial, Industrial, and Public Sector Programs 27
Table 4-1. Residential and Low Income Programs 47
Table 4-2. Estimation of Respondents' NPSO Savings 52
Table 4-3. Calculation of Total NPSO Generated 52
Table 4-2. Net Savings Example for a Sample Population* 55
Table 4-8. IECC 2015 Building Energy Code 76
Table 5-1. Example Triangulation Weighting Approach 78
Table 5-3. NTG Summary for Consumption Data Analysis 83

6 Cross-Cutting Measures

6.1 Behavior

6.1.1 Adjustments to Behavior Savings to Account for Persistence

DESCRIPTION

Energy efficiency program administrators are increasingly including behavior programs as part of their portfolios. These programs are characterized by various kinds of outreach, education, and customer engagement designed to motivate increases in conservation and energy management behaviors, and most commonly include participant-specific energy usage information. Savings impacts are evaluated by ex-post billing analysis comparing consumption before and after (or with and without) program intervention, and require M&V methods that include customer-specific energy usage regression analysis and randomized controlled trial (RCT) experimental designs, among others (see Behavioral protocol set forth in the IL-TRM Attachment A: Illinois Statewide Net-to-Gross Methodologies for more information). As such, initial calculation of savings is treated as a custom protocol[1].

An important issue for many stakeholders is whether energy savings from behavior programs continue over time (i.e., whether they persist beyond the initial program year). Behavior programs have now been delivered for a number of years in many jurisdictions. The weight of evaluation evidence indicates that the energy-saving behaviors influenced through these programs can persist beyond the initial period of program intervention, even without continued program participation[2]. This post-treatment savings persistence has implications for calculations of first-year savings, measure life, and cost-effectiveness testing. Accounting for persistence will yield savings and cost-effectiveness estimates that more accurately reflect the true benefits of these programs. Because annual goals are based on first-year savings, programs should only count savings attributable to first-year spending. The effect of persistence of savings beyond the first year should be included in any lifetime savings calculations (including cumulative persistent annual savings) and cost-effectiveness testing.

The protocol below was developed to outline the adjustments that should be made to account for the persistence of savings beyond the year of program delivery. This protocol is applicable to behavior programs of any type, delivered to residential or C&I customers, that have evaluated evidence of program persistence; however, the persistence values in this version of the protocol are specific to residential home energy reports (HERs)-type programs[3]. This general protocol should be used for any type of behavior program once supportable assumptions for persistence exist as measured by multi-year, rigorous evaluation studies; persistence factors for those behavioral programs may differ from the specific factors provided in this measure for HERs-type programs.

Currently, evaluations calculate a custom value on an annual basis to estimate yearly savings. Evaluators typically use a regression analysis to estimate program effects. These regression analyses provide what is called an average treatment effect on the treated (ATT) estimate of program savings. The ATT approach takes advantage of the presence of a randomly assigned control group for each cohort that received reports in the service territory. These regressions use various methods to account for household-specific usage patterns[4]. Because of the experimental design, we can assume that the treatment and control groups experienced similar historical, political, economic, and

---

[1] The protocol outlined here assumes that adjustments to remove the effects of savings from program lift (participation in other utility programs), including legacy uplift, to account for move-outs and opt-outs, to normalize for effects of weather, and any other appropriate adjustments, have been made as part of the custom calculation of savings – this final savings value is referred to as “Measured Savings” in the calculations below.

[2] Long-Run Savings and Cost-Effectiveness of Home Energy Reports Programs, Cadmus, October 2014. Also see additional sources in the REFERENCE TABLE below.

[3] Residential HERs-type programs: programs that regularly deliver home energy reports to residential customers through direct mail or email channels using a random control trial (RCT) experimental design. At a minimum, the reports include customer-specific usage information used for a comparison to similar households and individualized energy savings tips.

[4] For example, a linear fixed-effects regression (LFER) model includes a household-specific intercept to account for time-invariant, household-level factors affecting energy use, and a post program regression (PPR) model uses energy use lags to account for household-specific usage in the year prior to the program.

other events that had comparable effects on their energy use. Moreover, because these groups experienced generally similar weather conditions, it is not necessary to measure or include weather in the RCT model specification to calculate initial annual savings related to the program.

However, in the case of comparing and summing savings year over year, exogenous factors, such as weather, are likely to make annual estimates non-equivalent. In particular, weather is likely to play an important role in driving behavioral effects, affecting savings magnitude (e.g., a constant percentage change in consumption will result in more cooling savings during a hotter-than-average summer), as well as savings rate (e.g., the percentage change in consumption is likely to be higher during hotter-than-average summers. As such, for this framework, evaluators will adjust for effects related to weather as part of the custom inputs to this protocol. Each evaluator will choose the most appropriate method for weather normalization. For example, one method would be to provide savings using a model specification that incorporates standard weather year inputs (e.g., HDD and CDD), to be used as the initial input into the calculation of annual savings, as well as inputs for cost effectiveness, as outlined below. This input will approximate average savings for a standard weather year based upon historical data.[5] Adjusting savings to a standard weather year is consistent with how other weather-sensitive TRM measures are specified, and will remove weather risk from performance goals and cost-effectiveness testing.[6]

The protocol will become effective for residential HERs-type programs as of January 1, 2018. All ongoing programs will undergo a "reset" upon institution of this protocol[7]. Regardless of any previous history of behavior program delivery, the program year ending December 31, 2018 will be assumed to be Year 1 for all HERs-type programs underway at that time for the purpose of the incorporation of multiyear measure life/savings persistence into cost-effectiveness calculations and for the application of the adjustments to annual savings as outlined below. Should any additional new programs (referred to as "waves" in the calculations below) be established in 2018 or in subsequent years, their first year will be assumed to be Year 1 for that wave – that is, each wave is tracked separately and savings are calculated separately using the approach outlined here. Waves that existed prior to the program year ending December 31, 2018 will continue to be tracked separately for each wave. All residential HERs-type programs implemented prior to January 1, 2018 will assume a one-year measure life; the assumptions and protocols outlined below will not be applied retrospectively to any utility programs. Updates to persistence factors from future evaluations, once incorporated into the IL-TRM, will be used when available for calculation of annual savings values for applicable program years but will not be applied retrospectively to previous years' first-year savings calculations. All other types of behavior programs will continue to use a one-year measure life until supportable evidence exists for savings persistence, at which time this adjustment protocol can be used with appropriate persistence factors.

Determination of Efficient Behavior

Behavior programs focus primarily on reducing electricity and natural gas consumption through behavioral changes; this reduction is generally measured through ex-post billing analysis after program intervention. Specific energy conservation and management behaviors are not usually directly observable. The specific definition of the efficient case is part of the design of behavioral programs and is included as part of the custom saving protocol, which will include any adjustment necessary to remove effects of program-related investments in efficient equipment.

---

[5] In the future, this approach could be empirically tested by comparing actual savings calculated in future program years against standard weather year results, producing a 'realization rate' between planned and actual savings results. Standard weather years could potentially be enhanced to better reflect these differences.

[6] We acknowledge that this approach is a proxy for estimating actual savings to allow for prospective calculation of lifetime savings. However, a substantial limitation to this approach is the issue of unobserved behavioral ramp-up that is likely to occur for future waves of participants.

[7] It is understood that this approach does not accurately take into account that programs have been in place prior to this date, and the fact that customers at that time will have been receiving reports for variable amounts of time, with varied associated actual savings persistence from these earlier program efforts. The difficulties of trying to "phase in" persistence adjustments to reflect this history have been recognized, and the approach outlined here has been recommended by the Illinois TAC members as a reasonable approximation.

DETERMINATION OF BASELINE BEHAVIOR

The ideal baseline for behavior programs is the energy usage without the program intervention. Various types of experimental, quasi-experimental, and/or regression-based EM&V approaches are used to present statistically valid approximations to this without-program baseline[8]. The specific definition of the baseline case is part of the design of behavioral programs and is included as part of the custom saving protocol.

DEEMED LIFETIME/PERSISTENCE OF SAVINGS

Evaluations in Illinois have shown that savings from residential HERs-type behavior programs can persist into at least the first and second year following discontinuation of program delivery[9], though on-going savings levels decay in the second year. For other residential RCT programs evaluated to date, savings have been shown to persist for at least 3 years year following program delivery[10], and industry expectations are that savings likely persist beyond that. We assume here that savings persist at some level for 5 years[11]. On-going savings over those 5 years are not equal, however; it is preferable that actual levels of ongoing savings should be calculated by future year as outlined below (see Application of Persistence for Cost-effectiveness) and used in cost-effectiveness and lifetime savings calculations[12]. For other behavior program types without evaluations that quantify levels of persistence, measure life is assumed = 1 year.

DEEMED MEASURE COST

It is assumed that most behavior changes in residential settings can be accomplished with homeowner labor only and without investment in new equipment; therefore, without evidence to the contrary, measure costs in such residential programs focused on motivating changes in customer behavior may be defined as $0[13]. Costs for C&I programs may include additional staffing, software purchases, etc. Cost for such programs is therefore program specific and is determined on a custom basis.

LOADSHAPE AND COINCIDENCE FACTOR

While there is evidence from analysis of AMI data that the savings loadshape for residential HERs-type programs

---

[8] See the Illinois Behavioral protocol set forth in the IL-TRM Attachment A: IL-NTG Methods for more information concerning randomized control trials and quasi-experimental evaluation methods for non-randomized designs for behavior programs.

[9] ComEd Home Energy Report Opower Program Decay Rate and Persistence Study DRAFT-Navigant, presented to Commonwealth Edison Company, January 29, 2016; ComEd Home Energy Report Program Decay Rate and Persistence Study, Year Two DRAFT - Navigant, Presented to Commonwealth Edison Company, July 20, 2016; Behavioral Energy Savings Programs: Home Energy Reports Persistence Study Part 2 – April 2015 to September 2015 FINAL – Navigant, Prepared for Nicor Gas, September 21, 2016.

[10] Long-Run Savings and Cost-Effectiveness of Home Energy Reports Programs, Cadmus, October 2014. Also see additional sources in the REFERENCE TABLE below. Given the limited persistence studies available, we acknowledge that using an average of these studies by fuel type may be the best approximation of persistence rates. However, moving forward, the TAC will incorporate additional study values and develop the most appropriate persistence factors, taking into account participant characteristics, such as the duration of exposure, the frequency of reports, baseline usage, as well as the amount of time that has persisted since receiving their final report, and the shape of the persistence curve.

[11] Determined as a reasonable preliminary assumption by Illinois TAC members. This assumption should be updated as additional research is conducted on these types of programs, and additional evaluation should be undertaken to assess the reasonableness of this assumption for Illinois-specific programs.

[12] This method of applying calculated values for future year benefits is preferred. Alternatively, an effective measure life can be calculated as Effective Measure Life = Total Discounted Lifetime Savings / First Year Savings.

[13] Future evaluation of costs of behavior change is encouraged to help clarify this assumption. In addition, as noted earlier in this measure characterization, in order to ensure double counting of savings does not occur, the protocol outlined here assumes that adjustments to remove the effects of program lift have been made as part of the custom calculation of savings. In a similar manner, given the savings accounted for by other utility programs are removed from the savings claims and cost-effectiveness for the behavior program, the incremental costs associated with such utility program incentivized measures should also be excluded from the behavior program cost-effectiveness analysis, so as to help ensure double counting of costs does not occur in the utility portfolio cost-effectiveness analysis.

mirrors the whole-house electric energy load pattern, there are not yet enough data to develop a behavior-specific loadshape. Indications from several unpublished analyses[14] show that these behavior savings occur in a general pattern most closely approximated by the Residential Electric Heating and Cooling Loadshape (R10) than any other current residential measure loadshape; this is therefore recommended as the most reasonable approximation for use until more-specific data are available. Loadshapes and coincidence factors will need to be determined for other types of behavior programs once sufficient data are in hand.

Algorithm

CALCULATION OF SAVINGS

Throughout these protocols, Year T refers to the current reporting year for which annual savings are being determined[15].

ELECTRIC ENERGY SAVINGS

The algorithm shown below for this measure was developed to calculate the annual persistence-adjusted electric savings in to be reported in year T after adjustment to account for the proportion of the measured savings for that program year that actually reflects any persistent savings from prior years' program activities (Years T-1, T-2, T-3, and T-4)[16].

$$\Delta kWh_{T\ Adjusted} = \Delta kWh_{T\ Measured} - (\Delta kWh_{T-1\ Adjusted} * RR_{T-1,T} * PFE_1) - (\Delta kWh_{T-2\ Adjusted} * RR_{T-2,T} * PFE_2) - (\Delta kWh_{T-3\ Adjusted} * RR_{T-3,T} * PFE_3) - (\Delta kWh_{T-4\ Adjusted} * RR_{T-4,T} * PFE_4)$$

Where:

$\Delta kWh_{x\ Adjusted}$ = total program annual savings for year X after adjustments to account for persistence (calculated value)

$\Delta kWh_{x\ Measured}$ = measured kWh savings: total program savings as determined from custom calculation/billing analysis[17] of participants in program during year X (input value)

$RR_{Y,X}$ = Program retention rate in year X from year Y participation[18]

= % of program participants in year Y that are still in program in year X (input value:

[14] Based on communication from Mathias Bell based on (currently unpublished) studies done by Opower, Cadmus, and LBNL. Also see DTE Energy: Behavior Program Measures for Submission to 2015 MEMD - Year Three Energy Savings - Demand Savings. Energy Optimization, April 15, 2014. http://www.michigan.gov/documents/mpsc/memd_2015_453673_7.pdf

[15] Calculation algorithms account for attrition of customers out of the service territory, as well as persistence decay. It has been noted that there may also be a need to adjust for cross-year effects of large differences in weather conditions or economic impacts. Custom savings inputs therefore are adjusted for standard year weather.

[16] This calculation should be carried out separately for each "wave" of behavior programs, where a wave is defined as a newly launched program. For simplicity, any new wave is assumed to start at the beginning of a program year (Year 1) and may include multiple different treatment types such as usage groups, report frequency, etc. For example, any wave added after 2018, will be considered Year 1 in the year they are launched.

[17] All appropriate adjustments to remove effects of participation in other utility programs, move-outs, opt-outs, to normalize for effects related to weather, and other adjustments as determined by the program experimental design, are assumed to have been made to result in this value for "measured savings". This value has been adjusted for standard year weather terms.

[18] It is possible that some savings related to behavioral programs persist even after participants move and are therefore dropped from the program. Such persistent savings could potentially occur in two ways. First, some proportion of program savings likely comes from efficient measures installed on the premises and not otherwise identified through other direct program participation; this component of saving would likely persist even under new building ownership. Second, participants who move might exhibit changes in energy usage even in a new setting; this could continue to provide savings to the program administrator if the move was within the same utility territory. As of this time, no definitive information exists as to the level of program savings related to installed measures vs. behavioral changes, making determination of these effects highly uncertain. As such, this protocol assumes no persistent savings related to customers who move. Future evaluation related to this assumption is encouraged in order to make this determination more precise.

calculated as # participants still in program in year X / # participants in year Y))

PFE_Z = Persistence factor - electric (deemed value)

= % savings that persist Z years after savings were initially measured, where Z is a number from 1 - 4

= use table below to select the appropriate value

Electric Persistence Factors[19]

Program Type	Program Year T - record 100% of adjusted savings ($\Delta kWh_{TAdjusted}$ above)	Percent adjusted savings from Year T activities that persist 1 year after year T	Percent adjusted savings from Year T activities that persist 2 years after year T	Percent adjusted savings from Year T activities that persist 3 years after year T	Percent adjusted savings from Year T activities that persist 4 years after year T
		PFE_1	PFE_2	PFE_3	PFE_4
Residential HERs-type (RCT)	100%	80%	54%	31%	15%

[19] See REFERENCE TABLES below for sources.

Example of Adjusted Annual Savings Calculations:

Assume the following information on participation and measured savings for the following program years (all adjustments have been made to remove effects of program lift, weather, etc. within the custom savings calculations). Assume 2018 is the first year of all programs (or is the "reset" year).

	Reporting Year					
	2018	2019	2020	2021	2022	2023
Input data from program information and custom savings analysis						
# Participants (households)	120,000	109,000	103,000	99,000	94,000	90,000
kWh per participant (household)	200	250	245	250	250	265
Measured kWh savings (custom)	24,000,000	27,250,000	25,235,000	24,750,000	23,500,000	23,850,000

Calculation of Retention Rates:

For use in 2019:

$RR_{2018, 2019}$ = 109,000/120,000 = 0.908

For use in 2020:

$RR_{2018, 2020}$ = 103,000/120,000 = 0.858

$RR_{2019, 2020}$ = 103,000/109,000 = 0.945

For use in 2021:

$RR_{2018, 2021}$ = 99,000/120,000 = 0.825

$RR_{2019, 2021}$ = 99,000/109,000 = 0.908

$RR_{2020, 2021}$ = 99,000/103,000 = 0.961

For use in 2022:

$RR_{2018, 2022}$ = 94,000/120,000 = 0.783

$RR_{2019, 2022}$ = 94,000/109,000 = 0.862

$RR_{2020, 2022}$ = 94,000/103,000 = 0.913

$RR_{2021, 2022}$ = 94,000/99,000 = 0.949

For use in 2023:

$RR_{2019, 2023}$ = 90,000/109,000 = 0.826

$RR_{2020, 2023}$ = 90,000/103,000 = 0.874

$RR_{2021, 2023}$ = 90,000/99,000 = 0.909

$RR_{2022, 2023}$ = 90,000/94,000 = 0.957

Calculation of Adjusted Annual Savings:

$\Delta kWh_{2018\ Adjusted}$ = 24,000,000 kWh

$\Delta kWh_{2019\ Adjusted}$ = 27,250,000 – (24,000,000 * 0.908 * 0.80)
= 9,816,400 kWh

$\Delta kWh_{2020\ Adjusted}$ = 25,235,000 – (9,816,400 * 0.945 * 0.80) – (24,000,000 * 0.858 * 0.54)
= 6,694,122 kWh

$\Delta kWh_{2021\ Adjusted}$ = 24,750,000 – (6,694,122 * 0.961 * 0.80) – (9,816,400 * 0.908 * 0.54) – (24,000,000 * 0.825 * 0.31)
= 8,652,382 kWh

$\Delta kWh_{2022\ Adjusted}$ = 23,500,000 – (8,652,382 * 0.949 * 0.80) – (6,694,122 * 0.913 * 0.54) – (9,816,400 * 0.862 * 0.31) – (24,000,000 * 0.783 * 0.15)
= 8,188,837 kWh

$\Delta kWh_{2023\ Adjusted}$ = 23,850,000 – (8,188,837 * 0.957 * 0.80) – (8,652,382 * 0.909 * 0.54) – (6,694,122 * 0.874 * 0.31) – (9,816,400 * 0.826 * 0.15)
= 10,303,561 kWh

Apply the same approach to calculate adjusted annual kW and Therms.

SUMMER COINCIDENT PEAK DEMAND SAVINGS

Coincident peak demand savings in year T should also be adjusted to account for persistence from previous years using a similar algorithm[20].

If peak demand is measured directly by the custom savings analysis:

$$kW_{T\ Adjusted} = \Delta kW_{T\ Measured} - (\Delta kW_{T-1\ Adjusted} * RR_{T-1,T} * PFE_1) - (\Delta kW_{T-2\ Adjusted} * RR_{T-2,T} * PFE_2)$$

[20] While there are no current studies that evaluate the persistence of peak savings, without more-specific information on the actual behaviors undertaken by program participants and their corresponding peak savings, it seems reasonable to assume that peak savings will also persist in a similar pattern; both of the approaches given assume persistence in peak savings. Further evaluation should be undertaken to clarify this point and determine appropriate peak-specific persistence values.

$- (\Delta kW_{T\text{-}3\ Adjusted} * RR_{T\text{-}3,T} * PFE_3) - (\Delta kW_{T\text{-}4\ Adjusted} * RR_{T\text{-}4,T} * PFE_4)$

Where:

$\Delta kW_{X\ Adjusted}$ = total program demand savings for year X after adjustments to account for persistence (calculated value)

$\Delta kW_{X\ Measured}$ = total program demand savings as determined from custom calculation /billing analysis[21] of participants in program during year X (input value)

Other variables as defined above

If peak demand is not measured directly by the custom savings analysis, peak demand should be calculated as follows:

$\Delta kW_{T\ Adjusted} = (\Delta kWh_{T\ Adjusted\ Summer}$ / #summer hours) * peak adjustment factor

Where:

$\Delta kWh_{T\ Adjusted\ Summer}$ = average adjusted electric energy savings (calculated above) for peak summer months

$= \Delta kWh_{T\ Adjusted} * 0.42 * (3/5)$

$= \Delta kWh_{T\ Adjusted} * 0.25$

Where:

0.42 = Summer Loadshape % for May – Sept

3/5 = proportion of May-Sept hours that fall in June, July, and Aug

summer hours = # hours in June, July, and Aug

= 8760 / 4

Where: 8760 = Hours per year

peak adjustment factor = adjustment for peak k/w over average kW for these hours

= 1.5[22]

NATURAL GAS ENERGY SAVINGS

The algorithm shown below for this measure was developed to calculate the annual persistence-adjusted Therm savings in to be reported in year T after adjustment to account for the proportion of the measured savings for that program year that actually reflects any persistent savings from prior years' program activities (Years T-1, T-2, T-3, and T-4).[23]

$\Delta Therms_{T\ Adjusted} = \Delta Therms_{T\ Measured} - (\Delta Therms_{T\text{-}1\ Adjusted} * RR_{T\text{-}1,T} * PFG_1) - (\Delta Therms_{T\text{-}2\ Adjusted} * RR_{T\text{-}2,T} * PFG_2)$
$- (\Delta Therms_{T\text{-}3\ Adjusted} * RR_{T\text{-}3,T} * PFG_3) - (\Delta Therms_{T\text{-}4\ Adjusted} * RR_{T\text{-}4,T} * PFG_4)$

Where:

[21] All appropriate adjustments to remove effects of participation in other utility programs, move-outs, opt-outs, to normalize for effects related to weather, and other adjustments as determined by the program experimental design, are assumed to have been made to result in this value for "measured savings". This value has been adjusted for standard year weather terms.

[22] Based on an approach used in Michigan that gives resulting values supported by evaluation claims. Also see DTE Energy: Behavior Program Measures for Submission to 2015 MEMD - Year Three Energy Savings - Demand Savings. Energy Optimization, April 15, 2014. http://www.michigan.gov/documents/mpsc/memd_2015_453673_7.pdf

[23] This calculation should be carried out separately for each "wave" of behavior programs, where a wave is defined as a newly launched program. For simplicity, any new wave is assumed to start at the beginning of a program year (Year 1) and may include multiple different treatment types such as usage groups, report frequency, etc.

$\Delta Therms_{x\ Adjusted}$ = total program annual savings for year X after adjustments to account for persistence (calculated value)

$\Delta Therms_{x\ Measured}$ = total program savings as determined from custom calculation/billing analysis[24] of participants in program during year X (input value)

PFG_Z = Persistence factor - gas (deemed value)

= % savings that persist Z years after savings were initially measured, where Z is a number from 1 - 4

= use table below to select the appropriate value

Other variables as defined above

Gas Persistence Factors[25]

Program Type	Program Year T - record 100% of calculated savings ($\Delta Therms_{TAdjusted}$ above)	Percent adjusted savings from Year T activities that persist 1 year after year T	Percent adjusted savings from Year T activities that persist 2 years after year T	Percent adjusted savings from Year T activities that persist 3 years after year T	Percent adjusted savings from Year T activities that persist 4 years after year T
		PFG_1	PFG_2	PFG_3	PFG_4
Residential HERs-type (RCT)	100%	45%	20%	9%	4%

APPLICATION OF PERSISTENCE FOR PROSPECTIVE CALCULATIONS

For determination of prospective savings related to programs delivered in year T (including cost-effectiveness, lifetime savings, and cumulative prospective annual savings (CPAS)), future years' savings related to the current year activities should be recorded for this measure as savings for each specific year calculated using the table below. Because of the potentially confounding effects of differences in weather in future years, the savings inputs used ($\Delta kWh_{TAdjusted}$, $\Delta kW_{TAdjusted}$, $\Delta Therms_{TAdjusted}$) for these future-year savings calculations have been determined using weather normalized inputs. This input (to be provided by program evaluators) will approximate average savings for a standard weather year based upon historical data.[26]

Program Year T - record 100% of adjusted annual savings as calculated above	Percent savings from Year T activities that persist 1 year after year T	Percent savings from Year T activities that persist 2 years after year T	Percent savings from Year T activities that persist 3 years after year T	Percent savings from Year T activities that persist 4 years after year T
$\Delta kWh_{TAdjusted}$ $\Delta kW_{TAdjusted}$ $\Delta Therms_{TAdjusted}$	$\Delta kWh_{TAdjusted} * PFE_1 * RR_{Utility}$ $\Delta kW_{TAdjusted} * PFE_1 * RR_{Utility}$ $\Delta Therms_{TAdjusted} * PFG_1 * RR_{Utility}$	$\Delta kWh_{TAdjusted} * PFE_2 * RR_{Utility}^2$ $\Delta kW_{TAdjusted} * PFE_2 * RR_{Utility}^2$ $\Delta Therms_{TAdjusted} * PFG_2 * RR_{Utility}^2$	$\Delta kWh_{TAdjusted} * PFE_3 * RR_{Utility}^3$ $\Delta kW_{TAdjusted} * PFE_3 * RR_{Utility}^3$ $\Delta Therms_{TAdjusted} * PFG_3 * RR_{Utility}^3$	$\Delta kWh_{TAdjusted} * PFE_4 * RR_{Utility}^4$ $\Delta kW_{TAdjusted} * PFE_4 * RR_{Utility}^4$ $\Delta Therms_{TAdjusted} * PFG_4 * RR_{Utility}^4$

[24] All appropriate adjustments to remove effects of participation in other utility programs, move-outs, opt-outs, to normalize for effects related to weather, and other adjustments as determined by the program experimental design, are assumed to have been made to result in this value for "measured savings". This value has been adjusted for standard year weather terms.

[25] See REFERENCE TABLES below for sources.

[26] In the future, this approach could be empirically tested by comparing actual savings calculated in future program years against standard weather year results, producing a 'realization rate' between planned and actual savings results. Standard weather years could potentially be enhanced to better reflect these differences.

Where:

$RR_{Utility}$ = a utility-specific estimated retention rate for the program[27,28]

Other variables as defined above

[27] This retention rate should be an historical average, based on multiple years of data, that applies across all program waves for a given utility. The retention rate should be updated on a regular basis (for example, with the program planning cycles) to make sure it remains reflective of current program and economic conditions. Evaluators will decide for each utility what population the retention rate should be based on (for example: all residential customers; the entire population eligible for the program; the current program population). In making this decision, evaluators should consider data availability, expected changes in the program population in the planning cycle, and the eligible population for the program.

[28] It is possible that some savings related to behavioral programs persist even after participants move and are therefore dropped from the program. Such persistent savings could potentially occur in two ways. First, some proportion of program savings likely comes from efficient measures installed on the premises and not otherwise identified through other direct program participation; this component of saving would likely persist even under new building ownership. Second, participants who move might exhibit changes in energy usage even in a new setting; this could continue to provide savings to the program administrator if the move was within the same utility territory. As of this time, no definitive information exists as to the level of program savings related to installed measures vs. behavioral changes, making determination of these effects highly uncertain. As such, this protocol assumes no persistent savings related to customers who move. Future evaluation related to this assumption is encouraged in order to make this determination more precise.

Example of Calculation of Cost-effectiveness Inputs – for Electric Savings:

Assume the same information as was used in the Example of Adjusted Annual Savings Calculations.

	Reporting Year T					
	2018	2019	2020	2021	2022	2023
Annual Savings = Adj. kWh savings (previously calculated) = $\Delta kWh_{TAdjusted}$	24,000,000	9,816,400	4,634,922	5,384,166	4,683,858	11,741,354
$RR_{Utility}$ = 0.908						

<u>For calculating cost effectiveness in 2018:</u>

Cost-effectiveness benefit of 2018 savings in 2019 = $\Delta kWh_{2018\ Adjusted}$ * PFE_1* $RR_{Utility}$ = 24,000,000 * 0.80 * 0.908 = 17,433,600 kWh

Cost-effectiveness benefit of 2018 savings in 2020 = $\Delta kWh_{2018\ Adjusted}$ * PFE_2*$RR_{Utility}^2$ = 24,000,000 * 0.54 * 0.908^2 = 10,685,053 kWh

Cost-effectiveness benefit of 2018 savings in 2021 = $\Delta kWh_{2018\ Adjusted}$ * PFE_3*$RR_{Utility}^3$ = 24,000,000 * 0.31 * 0.908^3 = 5,569,683 kWh

Cost-effectiveness benefit of 2018 savings in 2022 = $\Delta kWh_{2018\ Adjusted}$ * PFE_4*$RR_{Utility}^4$ = 24,000,000 * 0.15 * 0.908^4 = 2,447,067 kWh

<u>For calculating cost effectiveness in 2019:</u>

Cost-effectiveness benefit of 2019 savings in 2020 = $\Delta kWh_{2019\ Adjusted}$ * PFE_1*$RR_{Utility}$ = 9,816,400 * 0.80 * 0.908 = 7,130,633 kWh

Cost-effectiveness benefit of 2019 savings in 2021 = $\Delta kWh_{2019\ Adjusted}$ * PFE_2*$RR_{Utility}^2$ = 9,816,400 * 0.54 * 0.908^2 = 4,370,365 kWh

Cost-effectiveness benefit of 2019 savings in 2022 = $\Delta kWh_{2019\ Adjusted}$ * PFE_3*$RR_{Utility}^3$ = 9,816,400 * 0.31 * 0.908^3 = 2,278,093 kWh

Cost-effectiveness benefit of 2019 savings in 2023 = $\Delta kWh_{2019\ Adjusted}$ * PFE_4*$RR_{Utility}^4$ = 9,816,400 * 0.15 * 0.908^4 = 1,000,891 kWh

<u>For calculating cost effectiveness in 2020:</u>

Cost-effectiveness benefit of 2020 savings in 2021 = $\Delta kWh_{2020\ Adjusted}$ * PFE_1*$RR_{Utility}$ = 6,694,122 * 0.80 * 0.908 = 4,862,610 kWh

Cost-effectiveness benefit of 2020 savings in 2022 = $\Delta kWh_{2020\ Adjusted}$ * PFE_2*$RR_{Utility}^2$ = 6,694,122 * 0.54 * 0.908^2 = 2,980,294 kWh

Cost-effectiveness benefit of 2020 savings in 2023 = $\Delta kWh_{2020\ Adjusted}$ * PFE_3*$RR_{Utility}^3$ = 6,694,122 * 0.31 * 0.908^3 = 1,553,506 kWh

Cost-effectiveness benefit of 2020 savings in 2024 = $\Delta kWh_{2020\ Adjusted}$ * PFE_4*$RR_{Utility}^4$ = 6,694,122 * 0.15 * 0.908^4 = 682,540 kWh

Etc.

Apply the same approach to calculate cost-effectiveness inputs for kW and for Therms.

WATER IMPACT DESCRIPTIONS AND CALCULATION

N/A

DEEMED O&M COST ADJUSTMENT CALCULATION

N/A

REFERENCE TABLES

Persistence studies done to date for HERs-type programs capture effects only through a limited time frame and only for the specific program characteristics of the programs studied. They may not accurately represent conditions in Illinois or those for all Illinois programs. The Illinois TAC has determined that an average annual persistence rate across the studies done to date (Table 1 below) is the best currently available data to approximate persistence for the first year for the general class of residential HERs-type programs. Additional information about the rate of decay in the following years is limited. Most studies done to date that assess decay after more than one year do not specifically evaluate after each individual year and instead just calculate an average annual decay across the years studied. This is true of persistence studies for gas HERs-type programs. For them, this protocol assumes a linear on-going rate of decay for five years based on the average annual persistence in Table 1.

Navigant has recently undertaken an evaluation of the ComEd electric HERs program specifically designed to determine the first and second year persistence rate separately for each individual year. The results, shown in Table 2 below, indicate an average increase in the year-over-year persistence factor from year 1 to year 2 of 15%. This level of non-linear increase in the persistence factor is assumed to hold for the five years of electric savings persistence for HERs-type programs and is used to calculate persistence factors used in this protocol. The average annual persistence rate from Table 1 is used for the first year.

It is recommended that the persistence values and the shape of the decay function used in this protocol continue to be updated regularly as further longer term and Illinois-specific evaluations are undertaken.

Table 1: Annual Persistence Rate for Residential HERs-type (RCT) Programs: Reference Studies

Utility/Location	Frequency of Reports when in program	Number of Months in Program Before Terminated	Number of Post-Treatment Savings Analysis Months	Average Annual savings decay	Persistence (= 100% - decay)	Source	Electric or Gas
Upper Midwest	Monthly & quarterly	24-25	26	21%	79%	1	Electric
West Coast	Monthly & quarterly	24	29	18%	82%	1	Electric
West Coast	Monthly & quarterly	25-28	34	15%	85%	1	Electric
SMUD	Monthly & quarterly	27	12	32%	68%	1	Electric
Puget Sound Energy	Monthly & quarterly	24	36	11%	89%	1	Electric
MASS	Monthly & quarterly	26	15	33%	67%	2	Electric
Illinois (ComEd): First Year	Bimonthly	16-52	12	10%	90%	3	Electric
Average Annual Electric Savings Persistence:					**80%**		
MASS	Monthly & quarterly	15	17	64%	36%	2	Gas
Illinois (Nicor)	Bimonthly	12	12	46%	54%	4	Gas
Average Annual Gas Savings Persistence:					**45%**		

Sources:
1: http://www.cadmusgroup.com/wp-content/uploads/2014/11/Cadmus_Home_Energy_Reports_Winter2014.pdf
2: http://ma-eeac.org/wordpress/wp-content/uploads/Home-Energy-Report-Savings-Decay-Analysis-Final-Report1.pdf
3:http://ilsagfiles.org/SAG_files/Technical_Reference_Manual/Version_5/Sources/ComEd_HER_Opower_Persistence_and_Decay_Study_DRAFT_2016-01-28.pdf
4:http://ilsagfiles.org/SAG_files/Technical_Reference_Manual/Version_6/Evaluation_Documents/Nicor_Gas_HER_Persistence_Study_Part_2_Final_2016-09-21.pdf

Table 2: Year-over-Year Persistence Factors for ComEd Residential HERs Programs

	Annual Persistence Factor				Implied Year-over-Year Persistence	Change in Year-over-Year Persistence
	Wave 1	**Wave 3**	**Wave 5 Non-AMI**	**Average**		
Year 1: 11/2013-10/2014	96%	98%	78%	90%	90%	
Year 2: 11/2014-10/2015	85%	83%	40%	69%	77%	15%

Source:
http://ilsagfiles.org/SAG_files/Evaluation_Documents/Draft%20Reports%20for%20Comment/ComEd_EPY7/ComEd_HER_Year_

Two_Persistence_and_Decay_Study_2016-07-20_Draft.pdf This evaluation extends the analysis of the ComEd program waves reviewed in the 2016 study (#3 above) to the second year after reports were terminated. The study shows an increased rate of decay in year two, indicating that a linear decay rate assumption may not be accurate, at least for the first two years. This assessment of a non-liner decay rate will be reviewed, and the rate as it extends beyond the first two years, will be revisited when there have been additional studies designed to explicitly assess the shape of the decay curve across several years.

MEASURE CODE: CC-BEH-BEHP-V03-190101

REVIEW DEADLINE: 1/1/2021

2019 Illinois Statewide Technical Reference Manual for Energy Efficiency Version 7.0

Attachment A

Illinois Statewide Net-to-Gross Methodologies

Effective for Evaluation

All NTG data collection and analysis activities for the program types covered by this document shall conform to the NTG methods set forth herein.

Attachment A: Illinois Statewide Net-to-Gross Methodologies

Policy Context for this Information

The Illinois Evaluation Teams (ADM Associates, Cadmus Group, Itron, Navigant Consulting, Opinion Dynamics, Ridge & 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 1-1. 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	http://www.icc.illinois.gov/downloads/public/edocket/367591.pdf
13-0498 (1/28/14)	Ameren Illinois Company (Ameren)	167, 171	http://www.icc.illinois.gov/downloads/public/edocket/367603.pdf
13-0499 (1/28/14)	Illinois Department of Commerce & Economic Opportunity (Department of Commerce)	20, 23, 49	http://www.icc.illinois.gov/downloads/public/edocket/367581.pdf
13-0549 (5/20/14)	Nicor Gas Company (Nicor)	41-42, 78	http://www.icc.illinois.gov/downloads/public/edocket/378494.pdf
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)	54-55, 66	http://www.icc.illinois.gov/downloads/public/edocket/378495.pdf

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)

1.2 Programs Currently Covered in this Document

This document is intended to cover the majority of residential and non-residential programs offered in Illinois.[29] Programs covered as of the writing of this document are listed in tables at the beginning of Section 3: Commercial, Industrial, and Public Sector Protocols and Section 4: Residential and Low Income Sector Protocols. If the design of a given program changes significantly, then it may mean that the NTG protocol listed for that program in this document is no longer appropriate. If that happens, the evaluator should follow the procedures outlined below under Section 1.4: Diverging from the IL-NTG Methods.

This document will be updated over time to incorporate new programs and to reflect recommended changes to existing methodologies. All NTG data collection and analysis activities for the program types covered by this document shall conform to the NTG methods set forth herein.

1.3 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 Illinois Energy Efficiency Policy Manual. In general, the following will take place:

- Updates will generally occur annually.
- Any changes to the IL-NTG Methods document will be circulated to the full SAG, and SAG participants will have a ten business day review process.
- Updates may be discussed within the SAG throughout the year but will be completed annually.
- Annually, the ICC Staff will submit a Staff Report (with the consensus Updated IL-TRM attached) to the Commission with a request for expedited review and approval.

[29] Evaluation reports on those programs can be found at http://www.ilsag.info/evaluation-documents.html.

- Updated NTG methods go into effect upon SAG approval, which may be before the annual TRM update or before the effective date of the updated TRM.

1.4 Diverging from the IL-NTG Methods

The NTG methods for the programs outlined in this document are partially binding. The criteria for deviating from the IL-NTG Methods document are set forth below. In all cases, the evaluators (or any interested stakeholder) submits the proposed deviation to the full SAG for a ten business day SAG review and comment period. In the event of an objection by a SAG participant, efforts may be made to see if consensus can be reached on the proposed deviation in a subsequent monthly SAG meeting. In this case, a final opportunity for SAG review and comment to the proposed deviation will be provided following the SAG meeting.

Evaluators may modify the approaches described in this document if the following three conditions have been satisfied:

1. Evaluators must explicate within the annual evaluation research plan (or another document) how specific items in the proposed modified NTG method will diverge from what is written in this document. Evaluators must justify why the divergence is appropriate.
2. Prior to the use of the modified NTG method for a particular program, evaluation teams must be in agreement on the use and execution of the modified NTG method.
3. Any objection from SAG participants regarding the proposed modified NTG method is resolved.

Evaluators may test alternative methods of estimating NTG for a particular program in addition to the NTG methods outlined in this document, if the following three conditions have been satisfied:

1. Evaluators must explicate within the annual evaluation research plan (or other document) the proposed alternative NTG method. Evaluators must explain why the proposed alternative NTG method might be superior to the NTG methods outlined in this document for the particular program. Evaluators must discuss the foundation for expecting that the proposed alternative NTG method is likely to produce meaningful results.
2. Prior to the use of the alternative NTG method for a particular program, evaluation teams must be in agreement on the key details of the approach for implementing the alternative NTG method.
3. Any objection from SAG participants regarding the proposed alternative NTG method gets resolved.

When performing alternative NTG methods for a particular program, the choice of methods may vary across the state. For example, if ComEd’s evaluator chooses to test Methods 1 and 2 for a particular program, Ameren’s and Department of Commerce’s evaluators do not also have to perform Methods 1 and 2 for a similar program.

Several sections of this attachment provide example questions that can be used to collect the data required in the NTG algorithms. Adjustments to refine specific question wording, e.g., to better reflect the design of the evaluated program, do not constitute divergence from the IL-NTG Methods.

1.5 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 Group[30] 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 SAG’s Technical Advisory Committee

[30] 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.

(TAC) a Comparison Exhibit of Non-Consensus Net-to-Gross Methods topics/procedures within two weeks 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.

2 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 statement.[31] 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 usually not a laboratory experiment that can be replicated over and over[32] to find out what actions people would have taken absent an intervention, one would need a time machine to take people back in time 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.

2.1 Definitions

For energy efficiency programs, evaluators differentiate between savings at a "gross" and "net" level as described below in the short set of relevant definitions. These definitions are not all encompassing or meant to restrict evaluation in any way, but to provide context before additional detail is provided in later sections. Research to determine attribution occurs to allow for a better understanding of the net level of savings.

Table 2-1. 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 who received a service offered through the subject efficiency program, in a given program year; also called program participant. 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.
Gross 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

[31] A probabilistic statement is not the same as the confidence and precision information calculated based on sampling theory.
[32] However, a small number of program designs do lend themselves to experimental or quasi-experimental designs that allow for regression analysis of net impacts.

Concept	Term	Definition
		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. Note that the net-to-gross ratio (NTGR) = ((1-Free Ridership) + Participant Spillover + Nonparticipant Spillover).
	Core NTGR	1-Free Ridership
	Free Rider	A program participant who would have implemented the program's measures or practices 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. There can be participant and/or nonparticipant spillover. *Participant spillover* (PSO) 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. Evaluated savings associated with Program Administrator Training programs will also be considered Participant spillover. There are several general categories of participant spillover: • *Inside spillover* (ISO): Occurs when program participants implement additional program-induced energy efficiency measures at the program project site. • *Outside spillover* (OSO): Occurs when program participants implement program-induced efficiency measures at other sites within the Program Administrator's service territory at which program project measures were not implemented. • *Like spillover:* Occurs when program participants implement program-induced efficiency measures of the same type as those implemented through the program. Like spillover can occur at the program project sites (ISO) or at other sites within the Program Administrator's service territory (OSO). • *Unlike spillover:* Occurs when program participants implement program-induced efficiency measures of a different type from those implemented through the program. Unlike spillover can occur at the program project sites (ISO) or at other sites within the Program Administrator's service territory (OSO). *Nonparticipant spillover* (NPSO) 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.
Markets	Market	The commercial activity (e.g., manufacturing, distributing, buying, and selling) associated with products and services that affect energy use.
	Market Effects	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

Concept	Term	Definition
		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 shares for energy-efficient goods, services, and design practices.
	Market Assessment	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 factors 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 (or whether continued or even increased intervention is necessary). 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.

Sources: 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; Violette and Rathbun 2014. The Uniform Methods Project: Methods for Determining Energy Efficiency Savings for Specific Measures, Chapter 23: Estimating Net Savings: Common Practices, http://www.nrel.gov/docs/fy14osti/62678.pdf.

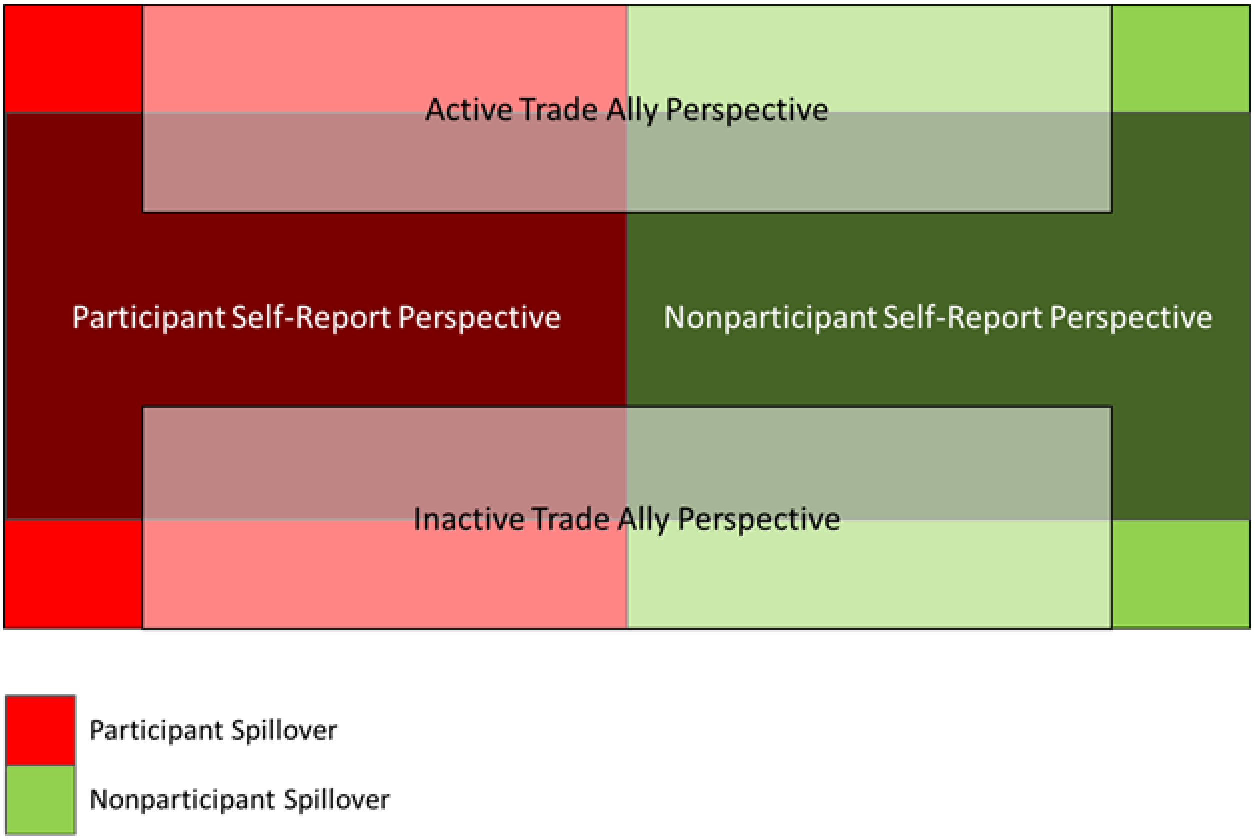
2.2 Spillover-Specific Issues

Some issues related to spillover are applicable for both residential and non-residential programs and are discussed in this section.

Spillover is generally categorized into two broad categories – participant spillover and nonparticipant spillover (see Table 2-1). These protocols include two general methods of assessing spillover, one through end-user (or participant/nonparticipant) research and the other through trade ally research. Estimates of participant and nonparticipant research are mutually exclusive, as long as only one of these two general methods is used for a given evaluation period. For example, there is no danger of double-counting spillover if an evaluation includes end-user research with both participants and nonparticipants. Similarly, there is no danger of double-counting spillover if an evaluation includes research with both active and inactive trade allies (see definitions in Section 5.2). However, once end-user research is combined with trade ally research, there is a potential for overlap in the resulting spillover estimates, and care must be taken to avoid double-counting.

Figure 2-1 provides a visual depiction of how the four methods (or "perspectives") for estimating spillover included in these protocols (participant and nonparticipant self-report, Sections 3.2 and 4.1; and active and inactive trade ally spillover, Section 5.2) can be used to assess both participant (red) and nonparticipant spillover (green). This figure illustrates that (a) different spillover methods can overlap in the spillover they cover, leading to potential double-counting, and (b) some spillover may not be measured by these methods (as represented by the four corners in the diagram).

Figure 2-1. Example - Types of Spillover and Methods for Assessment



2.2.1 Measure Costs

In order to facilitate analysis of program Total Resource Cost (TRC), estimates of the total incremental measure cost (IMC) at the program level must be developed. IMC values are available for most IL-TRM measures and can be summed to the program level. However, the IMC values for spillover measures could also be estimated and added to this total. The problem is that IMC values for spillover measures can be difficult to estimate. When the magnitude of the savings justifies the effort to estimate the total IMC for spillover measures, the following approaches should be used.

- In cases where the evaluator believes the spillover measure incremental costs are not materially different from the rebated measure incremental costs, the evaluator may multiply the IMC for the rebated measure by the spillover rate to derive the IMC for the spillover measure.
- In cases where the evaluator believes the spillover measure incremental costs are materially different from the installed measure incremental costs (e.g., installation of measures that have no efficiency levels), the evaluator should use the estimated incremental project costs as the IMC for the spillover measure.

Normally, the sample-based estimates of IMCs for spillover measures should be extrapolated to the program level using sample weights. Then the total IMCs for rebated measures and the total IMCs for spillover measures should be summed and used in the TRC calculation.

For measures characterized by the IL-TRM, measure effective useful life (EUL) estimates should be based on the IL-TRM. For measures not characterized by the IL-TRM, evaluator can use either the EUL for similar measures or best professional judgment. In either case, the evaluator must provide the rationale for their choices.

3 Commercial, Industrial, and Public Sector Protocols

The table below lists Illinois non-residential programs and the free ridership protocol applicable to each program.[33] If the design of a given program changes significantly, then it may mean that the NTG protocol listed for that program in this document is no longer appropriate. If that happens, the evaluator should follow the procedures outlined in Section 1.4: Diverging from the IL-NTG Methods. Note that the Core Non-Residential Spillover protocol described in Section 3.2 is generally applicable to most of these programs.

Table 3-1. Commercial, Industrial, and Public Sector Programs

Program Administrator	Free Ridership Protocol	Program Name
Ameren Illinois	3.1 Core Non-Residential Protocol	Standard Initiative – Core Program Custom Initiative Streetlighting Initiative Standard Initiative – Instant Incentives
	3.3 Small Business Protocol	Standard Initiative – Small Business Standard Initiative – Online Store
	3.5 Study-Based Protocol	Retro-Commissioning Initiative
ComEd	3.1 Core Non-Residential Protocol	Incentives (Standard, Custom) Business Instant Discounts
	3.3 Small Business Protocol	Small Business Air Care Plus Rural Small Business Kits
	3.4 C&I New Construction Protocol	C&I New Construction
	3.5 Study-Based Protocol	Incentives (Data Centers) Enhanced Building Optimization Program Industrial Systems Retrocommissioning Strategic Energy Management Operational Savings
	3.5 Study-Based Protocol or 5.3 Consumption Data Analysis Protocol	Power TakeOff
	5.3 Consumption Data Analysis Protocol	Business Energy Analyzer
	4.6 Multifamily Protocol	Public Housing Retrofits
	3.1 Core Non-Residential Protocol	LED Streetlighting
Nicor Gas	3.5 Study-Based Protocol	Strategic Energy Management
	3.3 Small Business Protocol	Small Business Program
	3.1 Core Non-Residential Protocol	Business Energy Efficiency Rebates
	3.1 Core Non-Residential Protocol	Business Custom Rebates
	3.4 C&I New Construction Protocol	Commercial and Industrial New Construction
	3.1 Core Non-Residential Protocol	Combined Heat and Power
	3.5 Study-Based Protocol	Retro Commissioning
Peoples Gas/ North Shore Gas	3.1 Core Non-Residential Protocol	C&I and PS Custom
	3.6 Technical Assistance Protocol	C&I and PS Direct Install and Assessment
	3.1 Core Non-Residential Protocol	C&I and PS Prescriptive
	3.3 Small Business Protocol	SB Custom SB Direct Install & Assessment

[33] The "Free Ridership Protocol Name" in the second column of the table refers to the numbered sections in this document, e.g., "3.3 Small Business Protocol."

Program Administrator	Free Ridership Protocol	Program Name
		SB Partner Trade Ally SB Prescriptive
	3.4 C&I New Construction Protocol	C&I and PS New Construction (Joint)
	3.5 Study-Based Protocol	C&I and PS Gas Optimization MF Gas Optimization C&I and PS Retro-Commissioning (Joint)
All	5.2 Code Compliance Protocol	Statewide Codes Collaborative

3.1 Core Non-Residential Protocol

3.1.1 Core Non-Residential Free Ridership Protocol

Key considerations and guidelines for estimation of free ridership under this Core Non-Residential Free Ridership (FR) protocol are listed 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 the availability of the program incentive, technical assistance from program staff, program staff recommendations, Program Administrator marketing materials, and an endorsement or recommendation by a Program Administrator, account manager or program partner staff. 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, of different factors that the evaluator determines are not a function of the program. Such non-program factors may include, for example, previous experience with the measure, standard business or industry practice, and organizational policy or guidelines.
- **Vendor Recommendations:** Vendor recommendations may also be a program factor to the extent that such recommendations are a function of the program. Vendors include trade allies, contractors, distributors, suppliers, and other market actors involved in the selection and installation of program-incented equipment on behalf of the participant. The evaluator may administer survey questions to vendors to verify their involvement with participant projects and to obtain their ratings—on a numeric scale—of the impact, influence, or importance of the program on the decision to recommend the energy efficiency measures to the program participant.
- **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 quantitative and qualitative 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.
- **Quality Control Review:** For programs involving large, complex projects and decision-making, after all the survey data collection has been completed and preliminary NTGRs have been computed using the standard calculation procedures, a quality control review is completed. All quantitative and qualitative data is systematically and independently analyzed by a researcher who is familiar with the program, the individual site and the social science theory that underlies the decision maker survey instrument. They make an independent determination of whether the additional information justifies modifying the previously calculated NTGR score, and present any recommended modifications and their rationale in a

well-organized manner, along with specific references to the supporting data. Circumstances that may justify a revision of the previously calculated NTGR score include: (1) significant inconsistencies exist between one of the scores that may lead to elimination of the score that is an outlier; (2) the emerging "story" from the qualitative data is in conflict with the quantitative data, thereby requiring a callback to the customer to resolve the inconsistency and a revision to the original scoring based on the new information; or (3) the entire set of results for an interview are inconsistent, the data are too disparate and would not be helped with a callback. In such cases, a recommendation is made to remove that sample point and replace it with a back-up point.

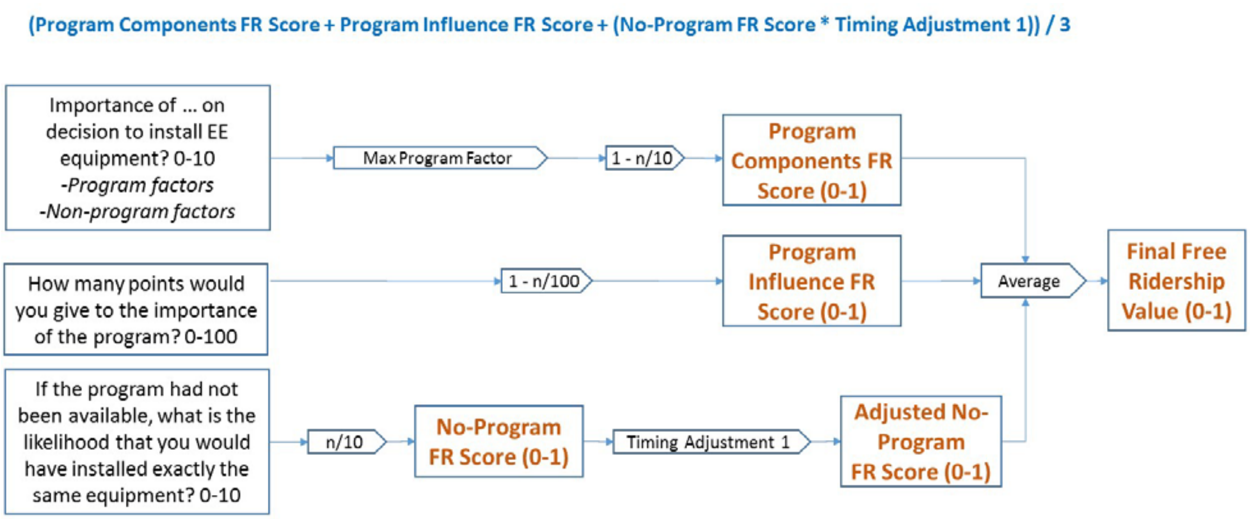
3.1.1.1 Core Free Ridership Scoring Algorithm

The Core Non-Residential FR protocol combines three scores that test different ways of approaching free ridership: the Program Components FR Score, the Program Influence FR Score, and the No-Program FR Score. The three scores are combined to calculate the final free ridership value.

Two options for combining the three scores are shown graphically in Figure 3-1 and Figure 3-2. These two options use different specifications to account for the impact of the program on project timing (referred to as "deferred free ridership"; see also discussion in Section 3.1.1.1.4). Evaluators will calculate free ridership using both options and will select one option for purposes of calculating the annual incremental energy savings for comparing to the legislated goal.[34] To select the appropriate option for use, we recommend that evaluators examine the various components of the free ridership scores to understand the differences between the options and justify their choice. Evaluators may also choose to use Cronbach's alpha to examine the internal consistency of the various options (but evaluators are not *required* to select the option with the highest Cronbach's alpha if they have justification for a different choice).

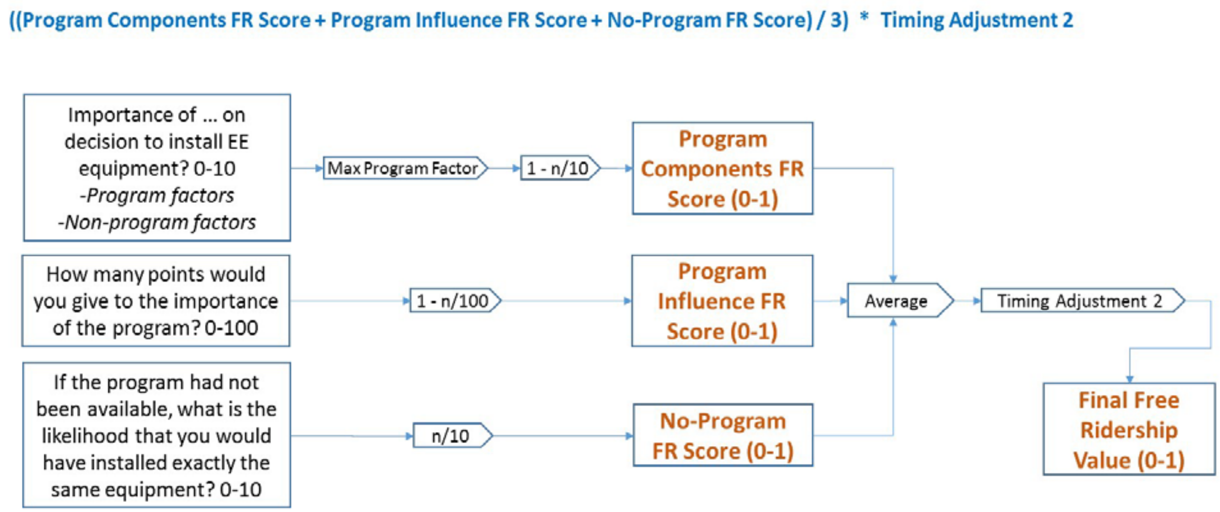
Evaluators will submit participant survey and net savings analysis data to the Illinois NTG Working Group. The group will analyze these data for the purpose of further refining the protocol and potentially reducing the number of alternative algorithm input specifications.

Figure 3-1. Core Free Ridership Algorithm 1



[34] As defined in 220 ILCS 5/8-103 and 220 ILCS 5/8-104.

Figure 3-2. Core Free Ridership Algorithm 2



3.1.1.1.1 Program Components FR Score

Evaluators will administer survey questions to obtain participants' rating of the importance of various factors on the decision to implement energy efficiency measures. The numeric scales shall range from 0 to 10, where 0 means "not at all important" and 10 means "extremely important". The various factors referenced in the survey will include those that the evaluator determines are program factors and non-program factors that could potentially impact the participant decision making process. A participant rating shall be obtained for each relevant program and non-program factor.

Evaluators will calculate the "Program Components FR Score" for each survey respondent using the following equation:

Program Components FR Score = 1 - ([Maximum Program Factor Rating]/10).

These scores can range from 0 (no free ridership) to 1 (full free rider). Since the algorithm uses the numerical rating for the Program Component receiving the highest score, it is important that such scoring be accurate. To facilitate this, the scores feeding into the Program Components FR Score calculation can be enhanced by adjusting survey wording and adding consistency checks around specific program components to seek clarification on how they influenced decisionmaking. For those program components receiving scores of 8, 9 or 10, additional questions can be included to determine why that specific score was given, and further, how that Program Component specifically influenced the participant's decision to upgrade to energy efficient equipment.

Evaluation reports should list all factors considered program and non-program factors. Evaluators must document why factors were treated as program factors or non-program factors.

3.1.1.1.2 Program Influence FR Score

Evaluators will administer a survey question that asks respondents to quantify the importance of the program on the decision to implement energy efficiency measures relative to the importance or impact of non-program factors. Respondents will be asked to allocate a total of 100 points to the program and to non-program factors. The points allocated to the program by the participants are the "Program Points." Evaluators will calculate the "Program Influence FR Score" as 1 - (Program Points/100). This score can range from 0 (no free ridership) to 1 (full free rider).

3.1.1.1.3 No-Program FR Score

Evaluators will administer a counterfactual likelihood survey question to obtain respondent ratings on a 0 to 10-point numeric scale (where 0 means "not at all likely" and 10 means "extremely likely") of the likelihood of the respondent to implement the exact same energy efficiency measures in the absence of the program. Evaluators will calculate the "No-Program FR Score" as the numeric score of the likelihood of the respondent to implement specified energy efficiency measures in the absence of the program divided by 10. This score can range from 0 (no free ridership) to 1 (full free rider).

Note that under one of the two deferred free ridership specifications (see next subsection), a timing adjustment is applied to the "No-Program FR Score." Under this specification, the resulting score is referred to as the "Adjusted No-Program FR Score."

3.1.1.1.4 Timing and Deferred Free Ridership

Evaluators will ask about the likely timing of measure installation in the absence of the program in two different ways. This is referred to as the counterfactual timing question since the evaluators are asking the respondent to speculate on what might have happened within a particular timeframe.

The first question will present a series of date ranges (e.g., within one year, between 12 months and 2 years, etc.) and ask the respondent to pick one representing their best estimate of when the measure would have been implemented in the absence of the program. The free ridership algorithm uses the midpoint of each date range, referred to as "Number of Months Expedited" below. For respondents that report accelerated adoption due to the program, this variable can take on values from 6 to 48 months.

The second question will prompt the respondent to use a 0 to 10-point numeric scale to report the likelihood, in the absence of the program, of implementing the same measure within 12 months of when it was actually implemented. This is the "Likelihood of Implementing within One Year" in the formulas below.

Evaluators will use the Likelihood of Implementing within One Year and/or the Number of Months Expedited variables to calculate two alternative ways of accounting for deferred free ridership:

1) Calculate *Timing Adjustment 1* as equal to:

   1 - (Number of Months Expedited - 6)/42

   Timing Adjustment 1 is multiplied by the No-Program FR Score; it can range from 0 (full deferred free ridership) to 1 (no deferred free ridership). The application of Timing Adjustment 1 is shown in Figure 3-1.

2) Calculate *Timing Adjustment 2* as equal to:

   1 - ((Number of Months Expedited - 6)/42)((10 - Likelihood of Implementing within One Year)/10)*

   Timing Adjustment 2 is multiplied by the average of the Program Components FR Score, the Program Influence FR Score, and the No-Program FR Score; it can range from 0 (full deferred free ridership) to 1 (no deferred free ridership). The application of Timing Adjustment 2 is shown in Figure 3-2.

How these timing adjustments are accounted for in the calculation of the Final FR Value is described below in the subsection "3.1.1.2 Construction of Core Free Ridership Value."

3.1.1.1.5 Consistency Checks

Respondents may be asked one or more questions to facilitate understanding and potentially reconcile apparently inconsistent responses. Some questions may be asked of all respondents; others may be asked when previous answers appear inconsistent. Evaluators should report on the amount of inconsistency encountered and on the resolution to inform future protocol revisions. Three consistency checks are outlined below.

Program Influence/Program Components Consistency Check

A Program Influence/Program Components consistency check is triggered when the following conditions are met:

1) The number of Program Points (supporting calculation of the Program Influence FR Score) is greater than 70; and
2) No program factor is rated greater than 2.

A Program Influence/Program Components consistency check is also triggered by the following conditions being met:

1) The number of Program Points (supporting calculation of the Program Influence FR Score) is less than 30; and

2) At least one program factor is rated greater than 7. In this instance, the highest-rated program factor(s) with a rating of greater than 7 will be referenced in the consistency check question.

Program Components/No-Program Consistency Check

A Program Components/No-Program consistency check is triggered when the following conditions are met:

1) The likelihood of installing the exact same equipment without the program (supporting calculation of the No-Program FR Score) is greater than 7; and

2) At least one program factor is rated greater than 7.

A Program Components/No-Program consistency check is also triggered when the following conditions are met:

1) The likelihood of installing the exact same equipment without the program (supporting calculation of the No-Program FR Score) is less than 3; and

2) No program factor is rated greater than 2.

Timing of Installation Decision/Level of Program Attribution Consistency Check

The survey should contain a question to ask whether the respondent learned about the program after finalizing project specifications, including, where applicable, equipment efficiency level and number of units. The Timing of Installation Decision/Level of Program Attribution consistency check is triggered by the following conditions being met:

1) A respondent learned about the program after finalizing project specifications; and

2) Any of the following occur:

a) The number of Program Points (supporting calculation of the Program Influence FR Score) is greater than 70;

b) The likelihood of installing the exact same equipment without the program (supporting calculation of the No-Program FR Score) is less than 3; or

c) At least one program factor is rated greater than 7.

When the Timing of Installation Decision/Level of Program Attribution consistency check is administered, if the respondent rating of the importance of the vendor on the decision to implement the project is greater than 7, then an open-ended question will be triggered to obtain information regarding the role the vendor played in the participant decision to implement the project.

3.1.1.2 Construction of Core Free Ridership Value

This protocol designates two options of constructing the core free ridership value. Evaluators will calculate free ridership using both options and will select one option for purposes of calculating the annual incremental energy savings for comparing to the legislated goal. Evaluators will present the results of both estimates of free ridership in EM&V reporting.

Evaluators will calculate free ridership values in the following two ways:

1) Core FR Algorithm 1 = AVERAGE([Program Components FR Score], [Program Influence FR Score], [No-Program FR Score*Timing Adjustment 1])
2) Core FR Algorithm 2 = AVERAGE([Program Components FR Score], [Program Influence FR Score], [No-Program FR Score]) * Timing Adjustment 2

The two Core FR Algorithms listed above are graphically presented in Figure 3-1 and Figure 3-2, respectively.

3.1.1.3 Vendor Influence in the Free Ridership Calculation

3.1.1.3.1 Treatment of Participant's Rating of Vendor in the Program Components FR Score of the Core FR Algorithm

The Program Components FR Score of the participant Core FR algorithm is based on participant ratings of program and non-program factors. Vendors[35] often receive a high rating for their influence on the participant's decision to install the efficient measure. To implement the Core FR algorithm, the evaluator needs to decide whether the vendor rating should be considered a program factor or a non-program factor. This section outlines three scenarios for the treatment of the participant's rating of a vendor in the Program Components FR Score of the Core FR algorithm.

Scenario #1: Vendors are automatically considered a program factor

The vendor is considered a program factor in the calculation of the Program Components FR Score in the FR algorithm if the program meets specific criteria, which could include the following:

1. Trade allies are an integral component of program delivery, as supported by program logic
2. The trade ally network consists of a limited number of Program Administrator-selected, pre-approved trade allies
3. Only trade allies can implement projects and submit applications on behalf of the customer
4. Trade allies complete signed agreements with the Program Administrator
5. Trade allies complete program-sponsored training

In these cases, the vendor is automatically considered a program factor, and no additional input from the vendor is needed regarding the customer's decision-making process related to the project. The participant's influence rating for the vendor goes directly into the Program Components FR Score algorithm as a program factor (if it is the highest rating given to any program factor).

Scenario #2: Vendors are considered a program factor if the program influenced their recommendation to implement the efficient project

For programs that have a trade ally network, but do not meet the conditions under Scenario #1 above, follow-up interviews with vendors may be used to determine if the vendor should be considered a program factor. To qualify for Scenario #2, a program's trade ally network should meet the following conditions:

1. Trade allies are registered with the program
2. Trade allies typically complete signed agreements with the Program Administrator
3. Trade allies complete program-sponsored training
4. Trade allies drive program participation, as supported by program logic

In these cases, if the size of the project warrants a greater level of effort, a follow-up interview with the vendor may be used to determine if the participant's rating of the vendor's influence should be included as a program factor. A follow-up interview is triggered under the following conditions:

1. The participant rated the influence of the vendor as 8, 9, or 10 (on a scale from 0 to 10)
2. The rating the participant gave to vendor influence is higher than any of the program factor ratings

If completed, the interview should include the following questions:

FR1a On a scale of 0 to 10 where 0 is NOT AT ALL IMPORTANT and 10 is EXTREMELY IMPORTANT, how important was the <PROGRAM>, including incentives as well as program services and information, in influencing your decision to recommend that <CUSTOMER> install the energy efficient <MEASURE> at this time?

FR1b On the same scale, how important was your firm's past participation in an incentive or study-based program sponsored by <PROGRAM ADMINISTRATOR>?

FR2 And using a 0 to 10 likelihood scale where 0 is NOT AT ALL LIKELY and 10 is EXTREMELY LIKELY, if the <PROGRAM>, including incentives as well as program services and information, had not been available, what is the likelihood that you would have recommended this specific <MEASURE> to <CUSTOMER>?

FR3a Approximately, in what percent of projects did you recommend <MEASURE> BEFORE you learned about

[35] Vendors include trade allies, contractors, distributors, suppliers, and other market actors involved in the selection and installation of program-incented equipment on behalf of the participant.

the <PROGRAM>?

FR3b And approximately, in what percent of projects do you recommend <MEASURE> now that you have worked with the <PROGRAM>?

The interview will also include consistency checks, if the vendor provides inconsistent responses to these questions.

The vendor is viewed as a program factor and the rating the participant provided for the vendor goes into the Program Components FR Score algorithm as a program factor if, after consideration of any consistency checks:

1. The response to Q. FR1a or FR1b is 8, 9, or 10

OR

2. The response to Q. FR2 is 0, 1, or 2

OR

3. The difference between the responses to FR3b and FR3a is 80% or greater

If none of these conditions are met, the rating the participant provided for the vendor does not go into the Program Components FR Score algorithm as a program factor.

In the event that an interview is not completed (e.g., the size of the project did not warrant a vendor interview or the vendor could not be reached), the evaluation reports should explain how the rating the participant provided for the vendor was treated. Guidelines for these situations may be added to this document in the future.

Scenario #3: Vendors are considered a non-program factor

For programs that do NOT have a trade ally network that meets the conditions under Scenario #2, vendors are considered a non-program factor. In these cases, the participant's rating of the vendor does not go directly into the Program Components FR Score algorithm as a program factor.

3.2 Core Non-Residential Spillover Protocol

Spillover refers to energy savings associated with energy-efficient equipment installed by consumers who were influenced by an energy efficiency program, but without direct intervention (e.g., financial or technical assistance) from the program.

To place the spillover protocols in context, we begin by defining the NTGR as:

NTGR = (1 – Free Ridership Value + PSO Rate + NPSO Rate)

Where:

PSO Rate = Participant spillover rate

NPSO Rate = Nonparticipant spillover rate

The term (1-Free Ridership) is referred to as the Core NTGR for an efficiency program.

3.2.1 Core Participant Spillover Protocol

The Core Participant Spillover protocol is generally applicable to most commercial, industrial, and public sector programs.

3.2.1.1 Research Methods

Data collection approach. An initial determination of participant spillover may be made based on self-reported findings from surveys of program participants. At a minimum, surveys collecting data pertaining to participant spillover will obtain general information on the specific measures installed and information substantiating their attribution to an energy efficiency program. Research on the specific characteristics of the energy efficient equipment installed and the baseline and operating conditions needed to estimate savings may be done in one of two ways: 1) a detailed battery of measure specific questions may be administered as part of the initial survey; or 2) a separate in-depth follow-up interview may be conducted by the engineer or analyst responsible for the energy savings calculation. In either case, an engineer or analyst will use the collected data to develop an estimate of spillover savings for each project.

Sample Frame. One target for participant spillover research may be the most recent year's program participants who have been sampled for free ridership or process surveys. In the case where a stand-alone spillover study is being conducted, the sample frame may be broader and include those whose participation occurred during the time period of two prior program years.

Because evaluated spillover energy impacts associated with the sample are being extrapolated to the program population, it is important that the sample frame be limited to participating customers for which spillover may potentially be claimed.

Sample frames should be constructed in accordance with the following guidelines:

- Self-directing customers as defined by 220 ILCS 5/8-104(m) should be excluded from the sample frame for natural gas spillover.
- Customers of municipal electric utilities should be excluded from the sample frame for electric spillover.

Timing of Data Collection. Evaluators may either administer the participant spillover module as part of a comprehensive net-to-gross survey, or they may elect to implement it separately. A follow-up in-depth interview may also be conducted by an engineer or analyst to obtain additional details needed to quantify savings. Optimally, the spillover inquiry should be timed in order to allow sufficient time for spillover to occur; at a minimum, three months after the program-incented measure is installed. Projects installed up to two years after program participation occurred may be counted as spillover, provided it can be substantiated.

3.2.1.2 Approach for Identifying and Quantifying Spillover

Attribution Criteria. Program attribution is determined by the responses to the following two survey questions:

1. How important was your experience in the <PROGRAM> in your decision to implement this measure, using a scale of 0 to 10, where 0 is not at all important and 10 is extremely important?
2. If you had not participated in the <PROGRAM>, how likely is it that your organization would still have implemented this measure, using a 0 to 10 scale, where 0 means you definitely WOULD NOT have implemented this measure and 10 means you definitely WOULD have implemented this measure?

The response to the first question cited above is "Measure Attribution Score 1," and the response to the second question cited above is "Measure Attribution Score 2."

There are two methods by which the attribution may be calculated:

1. Program attribution is established if the average of Measure Attribution Score 1 and (10 – Measure Attribution Score 2) exceeds 5.0[36]; either the Measure Attribution Score 1 or (10 – Measure Attribution Score 2) could be below 5.0—as long as the average is greater than 5.0, the threshold is met. If the average is greater than 5.0, 100% of the measure energy savings referenced in the question are considered to be attributable to the program. If the average is not greater than 5.0, none of the measure energy savings are considered to be attributable to the program.

[36] Note that the threshold value for counting spillover has been lowered from 7.0 to 5.0. The rationale for this lower threshold is: (1) the value of >5 is a strong indicator of program influence on the decision to install non-rebated equipment and is currently being used in other states (e.g., California); (2) the previous value of >7 set an unreasonably high standard for demonstrating program influence on the decision to install non-rebated equipment; and (3) past IL evaluation data show that a threshold of >5 will improve spillover estimates as it provides a better approximation of partial spillover (i.e., where a portion of the savings for each measure installed outside the program gets credited as spillover based upon the program influence rating).

2. An attribution rate may be calculated as equal to the sum of Measure Attribution Score 1 and (10 – Measure Attribution Score 2), divided by 20. For instance, if the attribution rate is 0.3, then 30% of the measure energy savings referenced in the question are considered to be attributable to the program.

Program attribution option 2 must be used in cases in which evaluators have performed the data collection and analysis required to attribute energy savings using option 2 identified above.

Calculation of Spillover Measure Energy Savings. Energy savings of spillover measures shall be calculated in one of two ways.

1. Those addressed in the IL-TRM shall be calculated in accordance with the methods and algorithms specified in the IL-TRM, and shall reference the IL-TRM-defined time-of-sale or new construction baseline.
2. For measures not addressed in the IL-TRM, evaluators shall quantify savings using accepted industry-wide savings methods that conform to IPMVP or other industry protocols and documents.

Evaluators will make every effort to ensure that there is no double-counting of participant spillover energy savings across multiple sources of participant and nonparticipant spillover (such as participating customer and trade ally surveys) and will document that effort.

Measure implementation must have occurred within one year of the participant spillover study data collection effort in order to be countable as participant spillover.

For the purposes of accounting for spillover savings attributable to a program, spillover will only be quantified for measures implemented within the Program Administrator's service territory.

3.2.1.3 Key Participant Spillover Survey Questions

The Participant Spillover question module is designed to be a general inquiry that seeks to: (1) assess whether additional energy efficiency improvements were implemented since the rebated project was completed; (2) confirm that these measures either had not received program incentives, or that there were no plans to submit them for program incentives in the future; (3) gather basic information about the additional energy efficiency measures (e.g., their type, size, quantities, and energy efficiency rating); and (4) establish program attribution.

The basic question structure is shown below. The measure-specific questions can be repeated in order to capture multiple measures. Note that there is considerable flexibility to tailor the questions to specific types of applications and programs.

1. Since your participation in the <PROGRAM>, did you implement any ADDITIONAL energy efficiency improvements at this facility or at your other facilities within <PROGRAM ADMINISTRATOR>'s service territory that did NOT receive incentives through <PROGRAM>?
2. What measures did you implement without an incentive?

MEASURE-SPECIFIC QUESTIONS [repeated for each spillover measure][37]

1. How important was your experience in the <PROGRAM> in your decision to implement this <MEASUREX>? Please use a scale of 0 to 10, where 0 is not at all important and 10 is extremely important.
2. Can you explain how your experience with the <PROGRAM> influenced your decision to install this additional high-efficiency measure?
3. If you had not participated in the <PROGRAM>, how likely is it that your organization would still have implemented <MEASURE>? Please use a 0 to 10, scale where 0 means you definitely WOULD NOT have implemented this measure and 10 means you definitely WOULD have implemented this measure.
4. How many of <MEASURE> did you install?3
5. Questions to further define the measure (as applicable):
   a. Type

[37] Example questions to gather engineering information to support the calculation of spillover savings may be accessed here: http://www.ilsag.info/il_ntg_methods.html

   b. Efficiency
   c. Size
   d. Other attributes
6. Can you briefly explain why you decided to install this energy efficiency measure on your own, rather than going through the <PROGRAM>?

3.2.1.4 Reporting of Results

Evaluators will report the following information relating to participant spillover data collection and analysis in annual EM&V reporting: 1) the number of participants surveyed; 2) the number of survey respondents reporting spillover; 3) the number of survey respondents who meet the spillover attribution threshold; 4) the number of respondents for which spillover savings were actually quantified; 5) the spillover savings for each project and overall; and 6) the spillover rate. The term (1-Free Ridership) is referred to as the Core NTGR.

The annual EM&V report should also describe the means by which the participant spillover rate is calculated. Two possible approaches are:

(1) Add the participant spillover rate to each project's Core NTGR. The project-level NTGRs are then weighted by each project's ex ante or ex post (if available) gross savings as a share of the total. This savings-weighted NTGR can then be applied to the ex post gross savings of the participant population. If the sample is stratified, sampling weights must be applied before applying the NTGR to the ex post gross savings of the participant population.

(2) Estimate program spillover effects by summing overall project-level spillover estimates for the sample and dividing this sum by the total ex ante or ex post (if available) gross savings for the sample to produce the participant spillover rate. This participant spillover rate can be added to the Core NTGR for the sample to yield the NTGR. If the sample is stratified, sampling weights must be applied before applying the NTGR to the ex post gross savings of the participant population.

In both cases, the participant spillover rate must be calculated at the project level for Option 1 or at the program level for Option 2, using the following formula.

$$Participant\ Spillover\ Rate = \frac{ISO + OSO\ in\ sample}{Ex\ Post\ Gross\ Impacts\ in\ sample}$$

Where:

ISO = Inside participant spillover

OSO = Outside participant spillover

3.2.2 Core Nonparticipant Spillover Protocol

The evaluation may perform research to measure nonparticipant spillover (NPSO). Evaluators will make efforts to ensure that there is no double-counting of energy savings across multiple sources and will document those efforts.

3.2.2.1 Core Nonparticipant Spillover Protocol – Measured from End Users

NPSO for end users is defined as the energy savings that are achieved when a nonparticipant end user—as a result of the influence of a Program Administrator's programs—implements energy efficiency measures *outside* of the Program Administrator's programs.

One option for the evaluator would be to survey nonparticipating customers and estimate spillover savings for any efficient measures installed that respondents are able to attribute to specific Program Administrator programs. However, in many cases, nonparticipants might find it difficult, if not impossible, to reliably attribute any of their installations to the influence of a specific Program Administrator program. If an evaluator suspects that nonresidential nonparticipants will not be able to reliably attribute spillover savings to any particular Program Administrator program, a second option would be to survey nonparticipants and estimate spillover savings from the installation of efficient measures that respondents are able to attribute to their general knowledge of the Program Administrator incentives and information, regardless of the particular program source. These protocols are written assuming that the NPSO for end users will be estimated using this second option.

Note that this protocol does not address estimating spillover for upstream and midstream programs where the end user is assumed to be completely ignorant of any Program Administrator influence. Of course, when considered feasible, evaluators are free to estimate spillover and spillover rates at the program-specific level with the suggested questions presented in Section 3.2.2.1.2 modified appropriately.

3.2.2.1.1 Research Methods

Data Collection Approach. An initial determination of spillover may be made based on self-reported findings from surveys of nonparticipants. At a minimum, surveys collecting data pertaining to nonparticipant spillover will obtain general information on the specific measures installed and information substantiating the influence of the Program Administrator on the installation decision. Research on the specific characteristics of the energy efficient equipment installed and the baseline and operating conditions needed to estimate savings may be done in one of two ways: (1) a detailed battery of measure specific questions may be administered as part of the initial survey, or (2) a separate in-depth follow-up interview may be conducted by the engineer or analyst responsible for the energy savings calculation.[38] Projects installed within the last two years of the nonparticipant spillover study data collection effort may be counted as spillover, provided program attribution and energy savings can be substantiated. In either case, an engineer or analyst will use the collected data to develop an estimate of spillover savings for each project.

Sample Frame. The sample frame for nonparticipant end user spillover research is composed of customers who have not participated in any programs within the last three years. Because evaluated spillover savings associated with the sample are being extrapolated to the nonparticipant population, it is important that the sample frame be limited to nonparticipants for whom spillover may potentially be claimed.

Sample frames should be constructed in accordance with the following guidelines:

- Self-directing customers as defined by 220 ILCS 5/8-104(m) should be excluded from the sample frame for natural gas spillover.
- Customers of municipal electric utilities should be excluded from the sample frame for electric spillover.
- Entities eligible to participate in the Illinois Department of Commerce and Economic Opportunity programs will not be included in sample frames for the study of nonparticipant spillover attributable to utility-administered programs.
- Entities eligible to participate in the utilities' programs will not be included in sample frames for the study of nonparticipant spillover attributable to programs administered by the Department of Commerce and Economic Opportunity.

Timing of Data Collection. Evaluators might administer the nonparticipant end user spillover study in parallel with the program impact evaluation, potential study or saturation study research, or at a different time.

3.2.2.1.2 Approach for Identifying and Quantifying Spillover

Key Nonparticipant Spillover Survey Questions. The nonparticipant end user spillover question module is designed to be a general inquiry that seeks to: (1) assess whether additional energy efficiency improvements were implemented during the study period; (2) confirm that these measures had not received program incentives and that there were no plans to submit them for program incentives in the future; (3) gather basic information about the additional energy efficiency measure(s), e.g., the type, size, quantities, and energy efficiency rating; and (4) establish the Program Administrator importance ratings. Note that while the example questions can be customized to assess the influence of a specific program in the Program Administrator portfolio, they are currently worded to capture influence of the Program Administrator, regardless of program source.

Below are example questions that might be used in a nonparticipant spillover survey. They are grouped by the following topics:

[38] See http://www.ilsag.info/il_ntg_methods.html for detailed example questions designed to collect information required to estimate spillover savings for a variety of measures.

- **Threshold conditions:** Is there some credible evidence that it was at least possible for the Program Administrator to have influenced the decision to install additional energy efficient measures?
- **Measure description:** Enough information needs to be collected for the measure and its operation to support a credible estimate of savings
- **Attribution:** Is there credible evidence that the Program Administrator had substantial influence on the end user's decision to install the efficient measure outside of any of the programs in the Program Administrator portfolio?

Threshold Conditions. Spillover cases are identified using a threshold approach in which certain minimal conditions must be met for a customer's installation to be considered for spillover. The following are example questions that evaluators may use (individually or in combination) to determine that program administrator influence on the installation is possible:

1. Before installing these measures, did you know that <PROGRAM ADMINISTRATOR> offers energy efficiency programs, incentives, and information to help their business customers make energy efficiency improvements at their facilities?
2. <PROGRAM ADMINISTRATOR> offers incentives for energy efficient equipment upgrades and improvements through its <PORTFOLIO NAME> programs. Before installing these measures, had you heard about the <PORTFOLIO NAME> programs?

If the answer to either question is "yes", then the threshold condition is met.

Measure Description. The interview (either the initial interview or a separate in-depth follow-up interview) can be used to determine the following basic attributes (as applicable) required to support a credible estimate of savings:

1. Type
2. Efficiency
3. Size
4. Other attributes

The named measure(s) must represent equipment that is more energy efficient than either: (1) equipment required by codes or standards; (2) industry-standard practice for certain types of equipment; or (3) for Custom measures, the minimum efficiency equipment available to meet the customer's requirements. For detailed example questions designed to collect engineering information required to estimate spillover savings for a variety of measures, see http://www.ilsag.info/il_ntg_methods.html.

Attribution. The following questions are suggested to assess attribution. These questions should be asked separately for each potential spillover measure:

1. Earlier you mentioned that you knew that <PROGRAM ADMINISTRATOR> offers incentives to customers for installing energy efficient equipment, and also provides information to customers to help them reduce their energy usage. Thinking about all of the reasons you chose to install the energy efficient <MEASURE>, did your knowledge of these incentives and information available through <PROGRAM ADMINISTRATOR> have ANY INFLUENCE on your decision to install <MEASURE>?

<u>ASK IF Q1=YES</u>

2. Using a scale of 0 to 10, where 0 is not at all influential and 10 is extremely influential, how much influence did your knowledge of the incentives and information <PROGRAM ADMINISTRATOR> offers have on your decision to install your energy efficient <MEASURE>?
3. Just to make sure that we understand you correctly, please answer the following hypothetical question. If you had you NOT known about the incentives and information <PROGRAM ADMINISTRATOR> offers, would you still have installed your energy efficient <MEASURE>? Please use a scale of 0 to 10, where 0 means you definitely WOULD NOT have installed your energy efficient <MEASURE> and 10 means you definitely WOULD have done so.

Consistency Checks

Respondents may be asked one or more questions to facilitate understanding and potentially reconcile apparently inconsistent responses. Evaluators should report on the amount of inconsistency encountered and on the resolution to inform future protocol revisions.

ASK IF Q2>7 AND Q3>7 OR Q2<3 AND Q3<3

4. In your own words, can you explain HOW your knowledge of the incentives and information <PROGRAM ADMINISTRATOR> offers influenced your decision to purchase or install your energy efficient <MEASURE>?

The evaluation analyst will assess the response to this open ended question and its consistency with the other questions, and, if warranted based on clear additional information, they will adjust the score based on expert judgment. If an inconsistency exists and the open-ended response does not resolve the inconsistency, the respondent will be removed from the calculation. All instances of this occurring should be documented in the final report. Additional consistency checks, triggered and resolved within the survey with additional questions to participants, remain optional.

Nonparticipant End User Spillover Algorithm. The response to question #2 cited above is "Measure Attribution Score 1," and the response to question #3 cited above is "Measure Attribution Score 2."

There are two methods by which the attribution may be calculated:

1. Provided that the open-ended responses do not contradict influence of the Program Administrator, spillover is considered to be attributable to the Program Administrator if the average of the Measure Attribution Score 1 and (10 – Measure Attribution Score 2) exceeds 5.0[39]; either the Measure Attribution Score 1 or (10 – Measure Attribution Score 2) could be below 5.0—as long as the average is greater than 5.0, the threshold is met. If the average is greater than 5.0, 100% of the measure energy savings referenced in the question are considered to be attributable to the Program Administrator. If the average is not greater than 5.0, none of the measure energy savings are considered to be attributable to the Program Administrator.
2. Provided that the open-ended responses do not contradict influence of the Program Administrator, the attribution rate is calculated as equal to the sum of Measure Attribution Score 1 and (10 – Measure Attribution Score 2), divided by 20. For instance, if the attribution rate is 0.3, then 30% of the measure energy savings referenced in the question are considered to be attributable to the Program Administrator.

Calculation of Spillover Measure Energy Savings. Energy savings of spillover measures shall be calculated in one of two ways.

1. Those addressed in the IL-TRM shall be calculated in accordance with the methods and algorithms specified in the IL-TRM, and shall reference the IL-TRM-defined time-of-sale or new construction baseline.
2. For measures not addressed in the IL-TRM, evaluators shall quantify savings using accepted industry-wide savings methods that conform to IPMVP and other industry protocols and documents.

Evaluators will make every effort to ensure that there is no double-counting of nonparticipant spillover energy savings across multiple sources of nonparticipant spillover reporting (such as nonparticipating customer and trade ally surveys) and will document that effort.

Measure implementation must have occurred within the last two years of the nonparticipant spillover study data collection effort in order to be countable as nonparticipant spillover.

For the purposes of accounting for spillover savings attributable to the Program Administrator, spillover will only be quantified for measures implemented within the Program Administrator's service territory.

3.2.2.1.3 Reporting of Results

Evaluators will report the following information relating to nonparticipant spillover data collection and analysis in annual EM&V reporting: 1) how the sample frame was defined, 2) the number of customers surveyed; 3) the number of survey respondents reporting spillover; 4) the number of survey respondents who meet the spillover attribution

[39] Note that the same 5.0 threshold value is being used for both Participant and Nonparticipant Spillover.

threshold; 5) the number of respondents for which spillover savings were actually quantified; 6) the spillover savings for each project and overall; 7) the nonparticipant spillover rate, and 8) the calculation of the weights used to extrapolate the spillover to the population of nonparticipants from which the sample was drawn.

The EM&V report should also describe the means by which the nonparticipant spillover (NPSO) rate is calculated. For each sampled site, the verified spillover savings should be summed across measures to derive the total end user NPSO for the sampled sites.[40] The estimate of site-level end user NPSO for the entire sample is then extrapolated to the entire nonparticipant population using sampling weights.

There are two options for using the estimated NPSO.

1. Allocate the portfolio-level spillover savings to individual programs in the portfolio based on each program's share of the ex post gross savings. For each program, the spillover rate could then be calculated for each program using the equation below in which the spillover allocated to each program would be the numerator and the ex post program-specific gross savings would be the denominator.

   $$Program - Specific\ NPSO\ Rate = \frac{NPSO_{Program-Specific}}{Ex\ Post\ Gross\ Impacts_{Program-Specific}}$$

   The spillover-adjusted NTGR for each program could then be used to adjust the Core NTGR for each program before calculating the TRC. In calculating the Program-Specific NPSO Rate, the numerator and denominator must be consistent in terms of the time period of measure implementation/potential implementation. While this time period must be within the last two years, it may be for a period of less than two years.
2. The NPSO Rate is calculated at the Sector level. The estimated energy savings associated with program-attributable spillover measures implemented during the study period by the entire nonparticipant population is divided by the ex post gross impacts for all the nonresidential programs in the portfolio occurring during the study period. The C&I Sector NPSO Rate is calculated using the following equation

   $$Portfolio\ NPSO\ Rate = \frac{NPSO_{Portfolio}}{Ex\ Post\ Gross\ Impacts_{Portfolio}}$$

The NPSO rate could then be used to adjust the portfolio core NTGR before calculating the portfolio TRC. Again, in calculating the Portfolio NPSO Rate, the numerator and denominator must be consistent in terms of the time period of measure implementation/potential implementation. While this time period must be within the last two years, it may be for a period of less than two years.

3.3 Small Business Protocol

3.3.1 Free Ridership

The FR algorithm for non-residential small business programs will follow the Core Non-Residential FR Protocol, with the following exceptions:

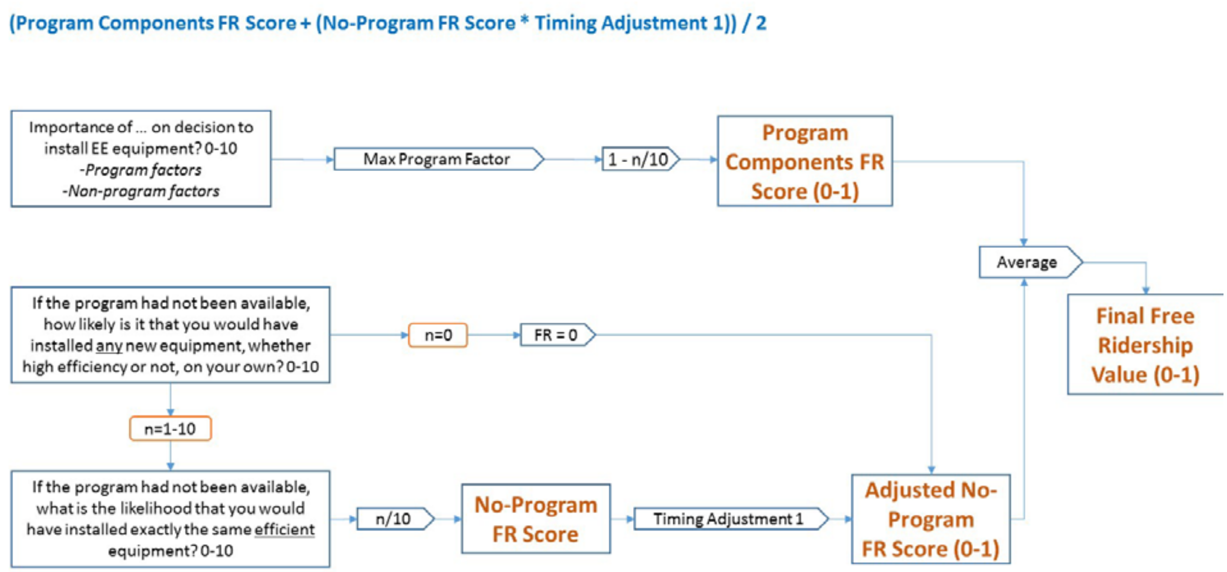
1. To reduce respondent burden, the Program Influence FR Score may be dropped from the Small Business FR algorithm. The influence of nonprogram factors will still be captured in the Program Components FR Score.
2. The counterfactual likelihood question (likelihood the participant would have installed the exact same energy efficiency equipment absent the program) may be preceded with a 0-10 scale question about the likelihood the participant would have installed any new equipment—either standard efficiency or high efficiency—on their own.
   a. If the participant provides a likelihood response of 0, then the No-Program FR Score for that participant is set to 0.
   b. If the participant provides a likelihood response of 1-10, then the participant is asked the same counterfactual questions (including the first timing question) as in the Core Non-Residential FR protocol.

[40] This includes all samples sites including those that reported no spillover savings.

3. To reduce respondent burden, the second question about timing (likelihood the participant would have installed the exact same energy efficiency equipment within 12 months) may be dropped. In this case, the only Deferred Free Ridership specification would be the one applying Timing Adjustment 1.

The diagram below, Figure 3-3, depicts the Small Business FR approach with the above exceptions implemented.

Figure 3-3. Small Business Free Ridership



Evaluators will calculate free ridership values for small business projects as follows:

(1) If Program Influence FR Score is dropped:

FR = AVERAGE ([Program Components FR Score], [No-Program FR Score * Timing Adjustment 1])

(2) If Program Influence FR Score is included:

FR = AVERAGE ([Program Components FR Score], [Program Influence FR Score], [No-Program FR Score * Timing Adjustment 1])

3.4 C&I New Construction Protocol

3.4.1 Free Ridership

The FR algorithm for non-residential new construction programs will follow the Core Non-Residential FR protocol, with the following exception:

- The concept of project timing and deferred free ridership is not applicable to new construction projects.[41] As a result, the various deferred free ridership specifications outlined in Figure 3-1 and Figure 3-2 will not be included in the free ridership estimation for new construction projects.

[41] New Construction programs intervene in the early phases of ongoing construction projects (i.e., after the decision to build has been made). As a result, participation in a New Construction program would not be expected to accelerate the construction of the new building.

Evaluators will calculate free ridership values for new construction projects as follows:

FR = AVERAGE ([Program Components FR Score], [Program Influence FR Score], [No-Program FR Score])

3.5 Study-Based Protocol

3.5.1 Free Ridership

The FR algorithm for non-residential study-based programs (See Figure 3-4) will follow the Core Non-Residential FR protocol, with the following exceptions:

- The counterfactual likelihood question (Q.4 in Figure 3-5 and Figure 3-6, below) will be preceded by five questions.[42]
- Q.1 A 0-10 scale question about the likelihood that the participant would have conducted the study absent the program will be included.

At the measure-group level, the following should be included:

- Q.2a A yes/no question to determine if the participant performs regular maintenance on the equipment treated through the program
- Q.2b If the response to Q.2a is "yes," a yes/no question to determine if the maintenance always includes the treatment provided through the program
- Q.3a A yes/no question to determine if the participant had prior awareness of the performance issues identified through the study
- Q.3b A 0-10 scale question about the participant's level of familiarity with the recommended actions to rectify the performance issue.

The counterfactual likelihood question (Q.4 – likelihood the participant would have taken action absent the program) and the first counterfactual timing question (used to develop Timing Adjustment 1) will be asked at the measure-group level. Measure-group level responses will be aggregated to the project level, using savings-based weights.

There will be two options for developing the No-Program FR Score:

1. The measure-group level Adjusted No-Program FR Score will be developed following Algorithm 1 of the Core Non-Residential FR approach, using responses to the counterfactual likelihood question (Q.4) and Timing Adjustment 1.
2. The measure-group level No-Program FR Scores will be assigned, based on responses to Q.1, Q.2b, Q.3a, and Q.3b, as follows:
    a. If Q.2b = Yes, then No-Program FR Score = 1. This assumes that if the participant performs regular maintenance on the treated equipment and that maintenance always includes the issue addressed through the program, then the participant is a full free rider for that measure group for purposes of calculating the No-Program FR Score.
    b. If Q.3a = No and Q1 = 0 and Q.2b ≠ Yes, then No-Program FR Score = 0. This assumes that if the participant was not aware of the performance issue and had a zero likelihood of performing the study absent the program and their maintenance practices do not always include the issue addressed through the program, then the participant is not a free rider for that measure group for purposes of calculating the No-Program FR Score since they would not have found out about the issue absent the program.
    c. If Q.3b = 0 and Q1 = 0 and Q.2b ≠ Yes, then No-Program FR Score = 0. This assumes that if the participant had no familiarity with how to rectify the performance issue, had a zero likelihood of performing the study absent the program, and their maintenance practices do not always include

[42] It should be noted that the question numbering in Figure 3-5 and Figure 3-6 is for reference purposes only; the additional questions do not have to immediately precede the counterfactual likelihood question.

the issue addressed through the program, then the participant is not a free rider for that measure group for purposes of calculating the No-Program FR Score since they would not have known how to address the issue absent the program.

d. For all other combinations of responses to Q.1, Q.2b, Q.3a, and Q.3b, the measure-group level Adjusted No-Program FR Scores will be developed following Algorithm 1 of the Core FR approach, using responses to the counterfactual likelihood question (Q.4) and Timing Adjustment 1.

Figure 3-4. Study-Based Free Ridership—Overview

(Program Components FR Score + Program Influence FR Score + (No-Program FR Score * Timing Adjustment 1)) / 3

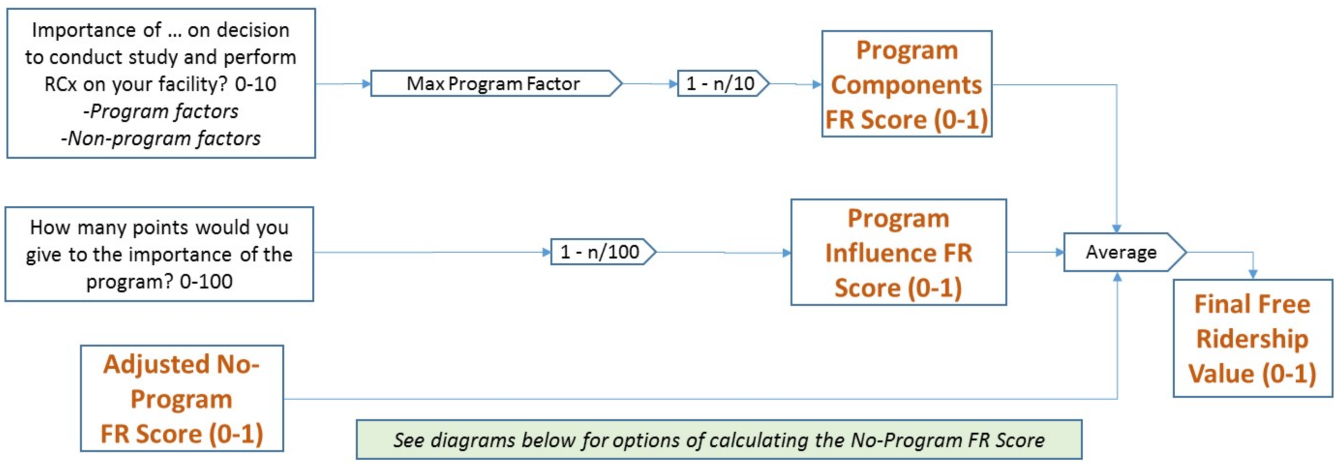


Figure 3-5. Study-Based Free Ridership—No-Program FR Score Option #1

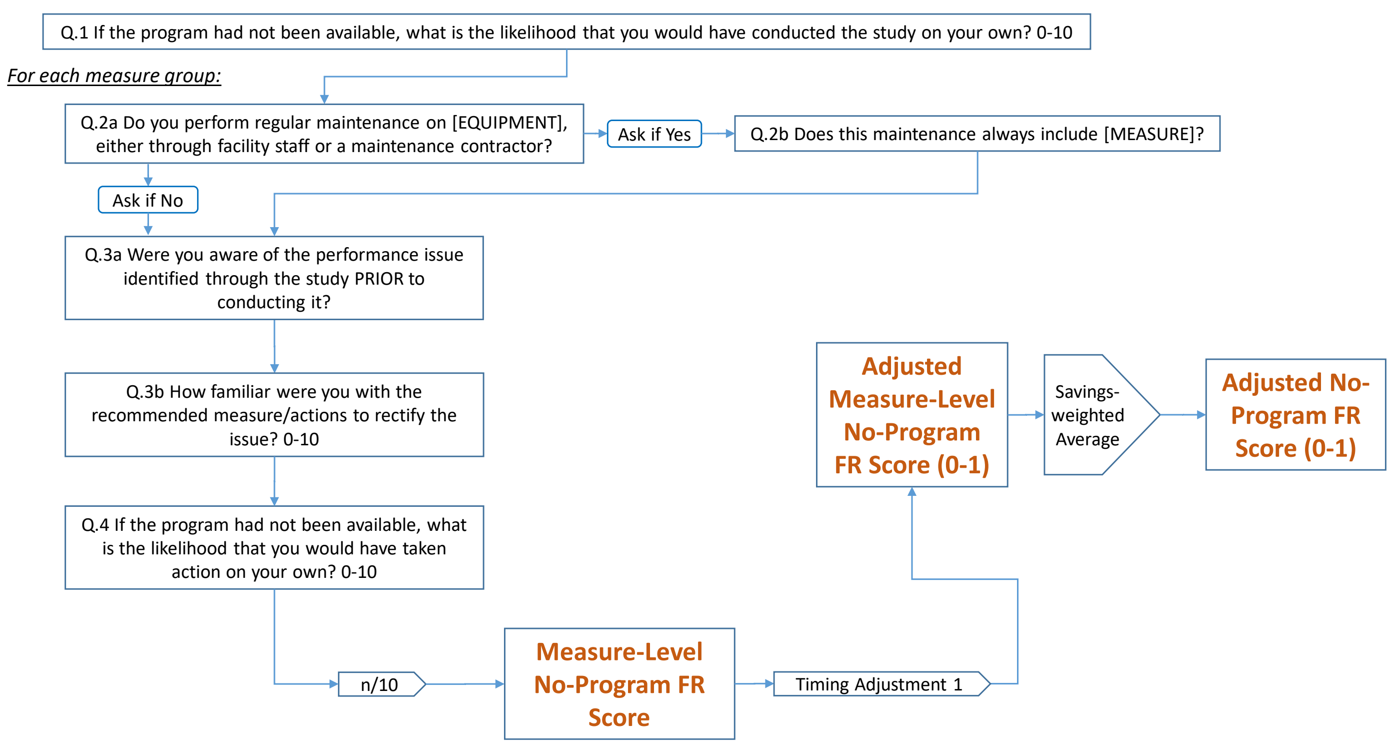
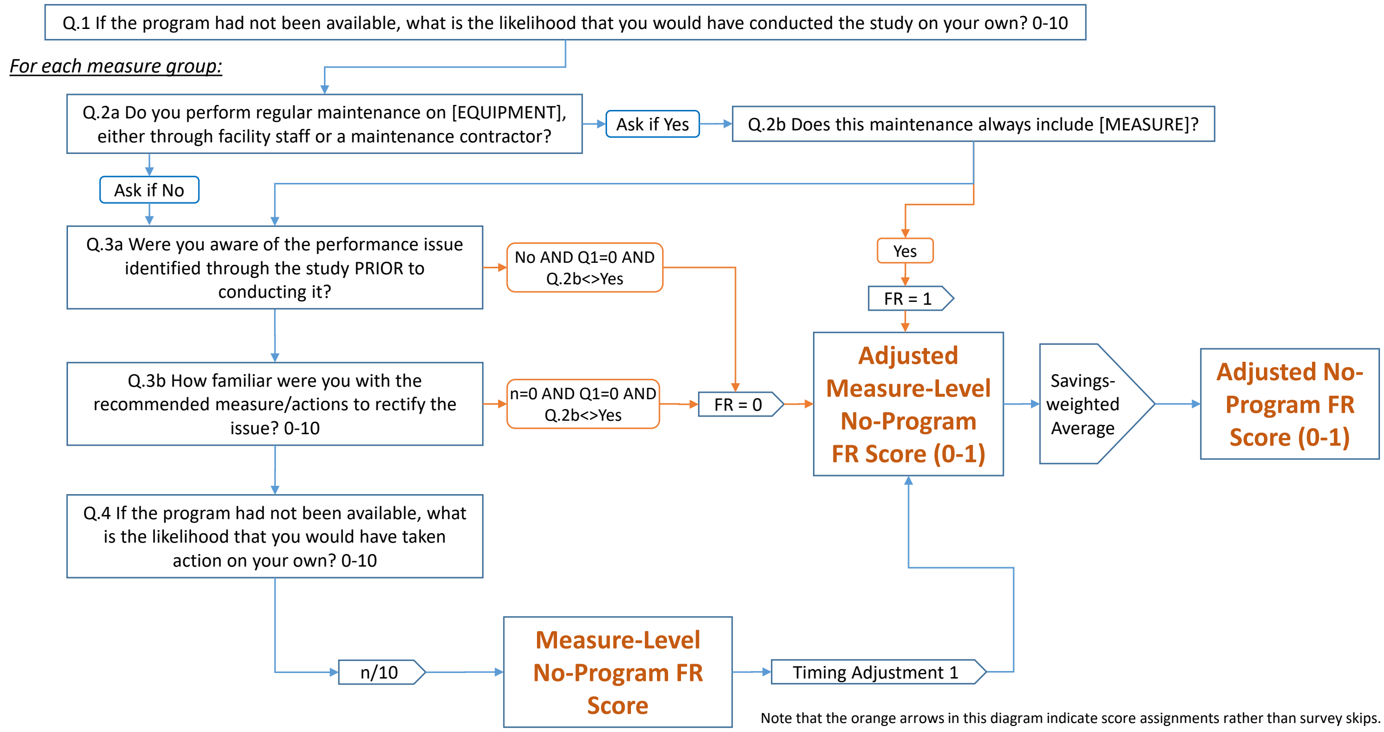


Figure 3-6. Study-Based Free Ridership—No-Program FR Score Option #2



Evaluators will calculate free ridership values for study-based programs as follows:

> FR = AVERAGE ([Program Components FR Score], [Program Influence FR Score], [No-Program FR Score * Timing Adjustment 1])

Evaluators will develop estimates of free ridership based on the two No-Program FR Score options outlined above. Evaluators will select one of these for purposes of calculating the annual incremental energy savings for comparing to the legislated goal. Evaluators will present the results of both estimates of free ridership in EM&V reporting.

3.6 Technical Assistance Protocol

This protocol is applicable to programs that provide technical assistance to encourage the adoption of energy efficiency measures in non-residential facilities, but do not provide financial incentives.

Program-attributable savings from Technical assistance programs are achieved when a program participant—as a result of the program's influence via the training or technical assistance provided—undertakes energy efficiency improvements on their own, without any direct financial assistance from any other Illinois energy efficiency program.

An initial determination of program-attributable savings is made based on self-reported findings from surveys of program participants. At a minimum, surveys collecting data pertaining to participant measure implementation will obtain general information on the specific measures installed and information substantiating their attribution to the program. Research on the specific characteristics of the energy-efficient equipment installed and the baseline and operating conditions needed to estimate savings may be done in one of two ways: 1) a detailed battery of measure specific questions may be administered as part of the initial survey; or 2) a separate in-depth follow-up interview may be conducted by the engineer or analyst responsible for the energy savings calculation. These collected data may be augmented by detailed facility and measure characteristics if provided by program staff.

3.6.1 Free Ridership

- The FR algorithm for Technical Assistance programs is identical to the Core Non-Residential FR protocol, with the following exception:

  - For the Program Components score, the list of program and non-program components differs extensively from conventional programs and therefore, is described in some detail here. As under the Core Protocol, evaluators administer survey questions to obtain participants' rating of the importance of a comprehensive list of program and non-program factors on the decision to implement energy efficiency measures. Examples of Technical Assistance program factors that may be included are: Documentation in a program-provided technical report of the energy saving opportunities from installing the measure.
- Verbal information or guidance provided by a program representative or energy auditor during a training course or an on-site visit.
- A follow-up communication from the utility regarding implementing the recommendations provided through the audit, training or technical assistance.

Examples of Technical Assistance non-program factors that may be included are:

- Information from trade shows, conferences, or other professional gatherings
- Recommendation from an equipment vendor that sold you the measure and/or installed it
- Previous experience with the measure
- A recommendation from a design or consulting engineer
- Standard practice in your business/industry
- Corporate policy or guidelines
- Payback on the investment

4 Residential and Low Income Sector Protocols

The table below lists Illinois residential programs and the NTG protocol applicable to each program.[43] If the design of a given program changes significantly, then it may mean that the NTG protocol listed for that program in this document is no longer appropriate. If that happens, the evaluator should follow the procedures outlined in Section 1.4: Diverging from the IL-NTG Methods.

Table 4-1. Residential and Low Income Programs

Program Administrator	Free Ridership Protocol	Program Name
Ameren Illinois	4.2 Appliance Recycling Protocol	Appliance Recycling Initiative
	4.3 Residential Upstream Lighting Protocol	Retail Products Initiative – Lighting Products
	4.4 Prescriptive Rebate (With No Audit) Protocol	HVAC Initiative Retail Products Initiative – Non-Lighting Products
	4.6 Multifamily Protocol	Multifamily Initiative
	4.7 Energy Saving Kits and Elementary Education Protocol	Direct Distribution of Efficient Products Initiative
	5.3 Consumption Data Analysis Protocol	Behavior Modification
	†	Income Qualified Initiative Public Housing Initiative Affordable Housing New Construction (any remaining DCEO commitments)
ComEd	4.2 Appliance Recycling Protocol	Fridge and Freezer Recycling
	4.3 Residential Upstream Lighting Protocol	Lighting Discounts
	4.4 Prescriptive Rebate (With No Audit) Protocol	Appliance Rebates Heating and Cooling Rebates Weatherization Rebates
	4.5 Single-Family Home Energy Audit Protocol	Home Energy Assessments
	4.6 Multifamily Protocol	Multifamily Assessments
	4.7 Energy Saving Kits and Elementary Education Protocol	NTC Middle School Take Home Kits Elementary Energy Education Kits
	4.8 Residential New Construction Protocol	Residential New Construction
	5.3 Consumption Data Analysis Protocol	Residential Behavior
	†	Income Eligible Lighting Discounts Income Eligible Single Family Retrofit Income Eligible Multi-Family Retrofit Affordable Housing New Construction Food Bank LED Distribution Program Income Eligible Kits Program
Nicor Gas	4.4 Prescriptive Rebate (With No Audit) Protocol	Home Energy Efficiency Rebates (Single Family)
	4.5 Single-Family Home Energy Audit Protocol	Home Energy Savings (Single Family Assessment/ Direct Install)
		Weatherization (Wx) Prescriptive (Air/Duct Sealing and Insulation)
	4.6 Multifamily Protocol	Multi-Family (Assessment/ Direct Install) Multi-Family Prescriptive Rebates

[43] The “Free Ridership Protocol Name” in the second column of the table refers to the numbered sections in this document, e.g., “4.6 Multifamily Protocol.”

Program Administrator	Free Ridership Protocol	Program Name
	4.7 Energy Saving Kits and Elementary Education Protocol	Elementary Education Kits Energy Saving Kits
	4.8 Residential New Construction Protocol	Residential New Construction
	5.3 Consumption Data Analysis Protocol	Behavioral Energy Savings
	5.2 Code Compliance Protocol	Code Compliance
Peoples Gas/ North Shore Gas	4.4 Prescriptive Rebate (With No Audit) Protocol	Home Energy Rebates
	4.5 Single-Family Home Energy Audit Protocol	Home Energy Jumpstart
	4.6 Multifamily Protocol	MF Custom MF Partner Trade Ally MF Prescriptive Multifamily (Direct Install)
	4.7 Energy Saving Kits and Elementary Education Protocol	Elementary Energy Education
	5.3 Consumption Data Analysis Protocol	Home Energy Reports
All	5.4 Code Compliance Protocol	Statewide Codes Collaborative

† There has been general consensus among Illinois stakeholders that the NTG value for Income Eligible programs is not likely to be significantly different from 1.0, particularly where the person making the participation decision is the Income Eligible resident. Until SAG establishes a different policy, the NTG value will be deemed at 1.0. Discussions will be held with SAG members on the value in and methods for performing such research and the timing of the application of such research.

4.1 Residential Cross-Cutting Approaches

The approaches in this section can apply to more than one program type but do not supersede program-specific approaches presented in later sections.

4.1.1 Survey Design Issues

Free ridership questions should be asked near the beginning of a participant survey, before asking satisfaction questions. This should prevent participants from confusing free ridership questions with the satisfaction questions, which could influence free ridership scores.

4.1.2 Participant Spillover

Effective program marketing and outreach generates program participation and increases general energy efficiency awareness among customers. Spillover can be calculated using participant survey questions, which ask participants about energy-savings actions they have taken on their own since participating in the program. Questions should be sufficiently specific to ensure energy savings associated with spillover can be reasonably well-quantified. These may include questions about measure types or measures installed, quantities, and efficiency levels. When program implementers provide recommendations to participants and can provide data on the types of recommendations made to specific participants, evaluations should attempt to determine whether participants took the recommended actions outside of the program at sites within the program administrator's service territory; if so, savings from those recommended actions should be attributed to the program.

To reduce the respondent's burden, the survey should first ask participants about the influence the program had on their taking additional energy-saving actions on their own. In particular, the evaluation team should ask two close-ended questions to determine program influence on spillover actions. The two required questions, preceded by an optional open-ended warm-up question, are:

- OPTIONAL: Did the program influence you in any way to make these additional improvements?

1. How important was your participation in the <PROGRAM ADMINISTRATOR'S> program on your making additional energy efficiency improvements on your own? [Scale from 0-10 where 0 is "not at all important" and 10 is "extremely important"]
2. If you had not participated in the <PROGRAM ADMINISTRATOR'S> program, how likely is it that you would still have implemented this measure, using a 0 to 10, scale where 0 means you definitely WOULD NOT have implemented this measure and 10 means you definitely WOULD have implemented this measure?

The response to the first required question cited above is "Measure Attribution Score 1," and the response to the second required question cited above is "Measure Attribution Score 2." The specific measures referenced in the question are considered to be attributable to the program if the "Spillover Score" is greater than 5.0:

Spillover Score = (Measure Attribution Score 1 + (10 – Measure Attribution Score 2))/2 > 5.0

If these conditions are met, the evaluator determines that the specific measures referenced in the question are attributable to the program; otherwise, the evaluator determines that the specific measures referenced in the question are not attributable to the program. The attribution criterion represents a threshold approach, in which energy impacts associated with measures implemented by program participants outside the program are either 100% program-attributable or 0% program-attributable.

For each measure mentioned, customers will be asked how they know the measure is more efficient than other models. If the respondent can identify the measure as ENERGY STAR or name an efficiency level that the evaluator confirms as being above the minimum federal standard, or if they identify a technology that the evaluator can confirm is above the minimum federal standard, it will count towards Participant Spillover.

Finally, depending on the measure type cited by the customer, follow-up questions should ask customers to provide reasonable information to allow the evaluator to estimate the amount of savings using IL-TRM protocols, such as quantity of appliances or the location and amount of insulation.

To calculate the spillover energy and demand savings for these actions, the appropriate version of the IL-TRM should be used. To develop the spillover rate, the total energy and demand impacts from the sampled participants who installed additional measures due to participation in the program are summed, and then this sum is divided by the total ex post sample energy and demand impacts:

$$Participant\ Spillover\ Rate\ (PSO) = \frac{Sum\ of\ Energy\ or\ Demand\ from\ Additional\ Measures\ Installed}{Sample\ Ex\ Post\ Gross\ Energy\ or\ Demand\ Impacts}$$

The equation used to adjust the Core NTGR based on participant spillover is as follows:

$$NTGR = (1 - FR + PSO)$$

4.1.2.1 Data Collection

Respondents should be drawn from a random sample of current or up to one year of previous program participants. Regardless of the participation year, spillover should be measured within the last 12 months (from the survey date), but after previous participation; the tracking database should supply this information.

4.1.2.2 Data Analysis

The following four steps calculate spillover:

1. Calculate total spillover savings for each participant installing an efficient measure not rebated through the program where the Spillover Score is greater than 5.0:

$$\text{Measure Spillover} = \text{Measure Savings} * \text{Number of Units}$$

2. Total savings associated with each program participant to calculate overall participant spillover savings.
3. $\text{Spillover Percentage Estimate} = \frac{\sum \text{Sample Spillover kWh Savings}}{\text{Sample Evaluated Program kWh Savings}}$

4.1.3 Nonparticipant Spillover Measured from Customers

The evaluation may perform research to measure nonparticipant spillover (NPSO). If so, care should be taken to ensure spillover is not double-counted with a trade-ally approach. The basic method uses a two-step process: (1) conduct a nonparticipant survey to identify potential spillover measures and (2) if needed, conduct a follow-up call or on-site visit by technical staff to confirm attribution and obtain information needed to estimate energy savings.

4.1.3.1 Basic Method

4.1.3.1.1 Sampling

As spillover may be rare in the nonparticipating population, determining spillover will likely require a large sample of customers who have not participated in any energy efficiency programs, including a behavioral program, within the past three years. Customers will be removed from the sample frame if their account numbers can be cross-referenced against a list of program participants from the previous three years. The survey should target household members responsible for paying utility bills. Survey respondents will be asked a screening question (whether they have participated in a program in the past three years) to confirm their household qualifies as a true nonparticipant.

4.1.3.1.2 Measure-Specific Questions

Depending on the spillover measure type reported by the customer, follow-up questions should be included to gather sufficient information to reasonably assess the saving amount by applying the IL-TRM, understanding that assumptions must be made if IL-TRM inputs cannot be easily supplied by the participant. Such assumptions should be conservative, or, if not conservative, reasons for deviating from the conservative application should be documented. Measures that cannot be reasonably quantified within available evaluation budgets should be excluded from spillover calculations.

For measures included in the IL-TRM, savings will be assessed using the IL-TRM algorithms. Baselines for measures not in the IL-TRM will be assessed based on appliance standards and building codes, if applicable, and, if not, through engineering judgements of existing or market conditions. Engineering assumptions and analysis by the evaluator will be applied for measures not included in the IL-TRM. Key assumptions should be documented in the report.

4.1.3.2 Attribution Approach

To receive credit for energy savings, the nonparticipant must fit the following criteria: (1) be familiar with the Program Administrators energy efficiency campaign (e.g., ActOnEnergy for Ameren); and (2) indicate that some aspect of the Program Administrator's energy efficiency programs motivated their purchases. Influence will be measured on a scale of 0 to 10, where 10 is extremely influential and 0 is not at all influential. Savings attribution requires a Spillover Score of greater than 5.0.

Survey respondents will be asked a series of questions following the logic shown in Figure 4-1. First, the customer will indicate whether they know about their Program Administrator's energy efficiency programs and/or marketing messages. If customer is aware, the survey will ask if they or anyone in their household made an energy efficiency improvement within the last year, and if so, what improvements they made. Responses to these questions will generate a list of potential spillover measures (shown at point "[A]" in Figure 4-1). Customers will be asked how they know the measure is more efficient than other models. If the respondent can identify the measure as ENERGY STAR or name an efficiency level that the evaluator confirms as being above the minimum federal standard, or if they identify a technology that the evaluator can confirm is above the minimum federal standard, it will count towards NPSO. At this point in the NPSO process, the customer could be referred for a follow-up call with a technical interviewer.[44]

To assess attribution for each spillover measure mentioned, the customer will be asked questions to be scored in two areas. Spillover may be program-attributable for those measures for which self-report data meet the following threshold condition:

[44] Customers who installed efficient lighting (CFL/LED) will not be eligible for NPSO if those savings are already claimed by an upstream lighting program. A separate NPSO protocol is provided specifically for upstream lighting programs.

Spillover Score = (Attribution Score 1 + (10 – Attribution Score 2))/2 > 5.0

4.1.3.2.1 Attribution Score 1

The first score, "Attribution Score 1," measures the influence level (on a scale of 0 to 10, where 10 is extremely influential and 0 is not at all influential) their Program Administrator had on the purchase of the measure.

Influence can derive from the following:

1. General information about energy efficiency provided by the Program Administrator (e.g. through a bill insert)
2. Information from a contractor or retailer related to the Program Administrator's programs.
3. Word-of-mouth from people installing energy-efficient equipment and receiving a rebate from the Program Administrator.

Attribution Score 1 is the maximum score (or Yes response) assigned to any source of influence from the Program Administrator.

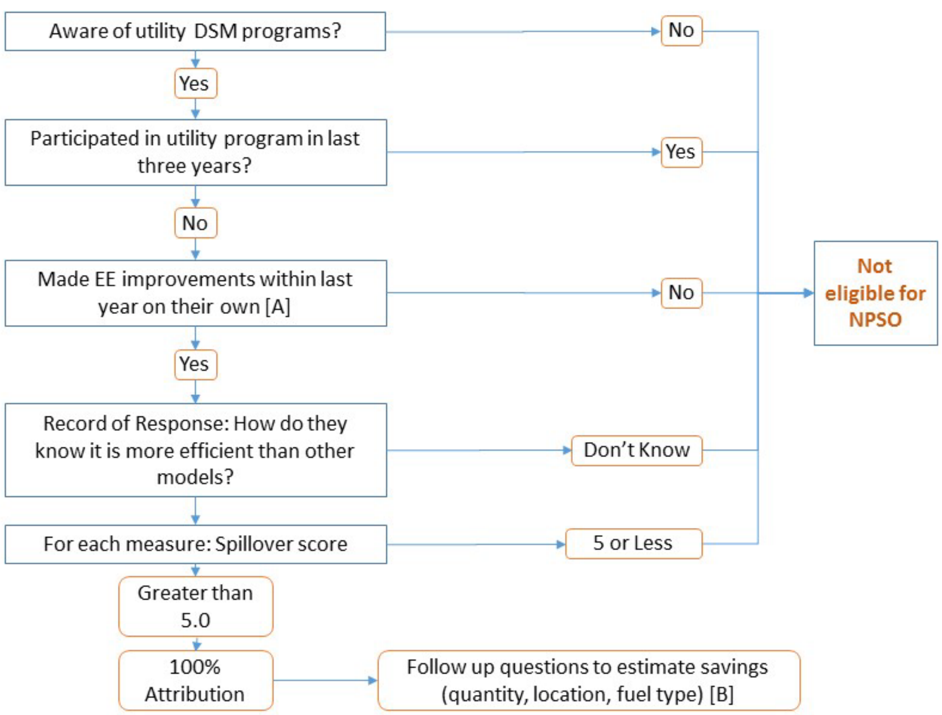
4.1.3.2.2 Attribution Score 2

The second score, "Attribution Score 2," comes from the customer's response to a single question to assess the counterfactual, asking about the likelihood (on a scale of 0 to 10, where 10 is extremely likely and 0 is not at all likely) that the customer would have installed the measure had they not been influenced by the program.

The Spillover Score is then the average of the Attribution Score 1 and (10 – Attribution Score 2). If that Spillover Score is greater than 5.0, 100% of the savings are attributed to the Program Administrator for that measure.

Finally, depending on the measure type cited by the customer, follow-up questions will gather information to enable an estimate of savings (shown in the figure as [B]), such as quantity of appliances or the location of insulation.

Figure 4-1. NPSO Question Logic



4.1.3.3 Scoring

Survey respondents' answers to the NPSO questions will determine total energy and demand savings attributed to the program. Table 4-2 lists NPSO measures under column A, the Spillover Score under column B, the estimated measure savings under column C, the percentage of allocated savings under column D, and the total allocated savings under column E. Column F shows the calculated average energy savings per spillover measure, determined by dividing the total allocated savings (the sum of column E) by the number of surveyed nonparticipating customers. The table shows how kWh NPSO savings would be calculated; calculations of therm or demand savings would be accomplished in the same manner.

Table 4-2. Estimation of Respondents' NPSO Savings

A	B	C	D	E	F
Spillover Measure	Spillover Score	Measure Savings (kWh)	Allocated Savings	Total kWh Savings	Average kWh Per Surveyed Customer
Measure1	Scale of 0 to 10	Savings1	100% if [B] > 5.0 0% if [B] ≤ 5.0	[C] x [D]	N/A
Measure2	Scale of 0 to 10	Savings2		[C] x [D]	
MeasureN	Scale of 0 to 10	SavingsN		[C] x [D]	
				Sum of column E = Total kWh Savings	Total kWh Savings ÷ Number of Completed Surveys

Table 4-3 shows the process for estimating total NPSO generated by the Program Administrator during the program year (for electric savings). The savings attributed from the survey population will be extrapolated to the nonparticipating residential customer population to determine the overall NPSO savings. Then NPSO energy savings will be converted into a percentage using the total evaluated electric savings for the program year. A similar process would apply for calculating therm or demand NPSO.

Table 4-3. Calculation of Total NPSO Generated

Variable	Description	Source/Calculation
F	Average kWh Energy Savings per Surveyed Customer	Survey data and Savings Calculation
J	Total Nonparticipating Residential Population	Customer database
K	NPSO MWh Energy Savings Extrapolated to Nonparticipating Population	[F × J] ÷ 1,000 kWh/MWh
S	Total Evaluated MWh Savings	Residential Portfolio Savings
G	NPSO Spillover Rate	K ÷ S

4.2 Appliance Recycling Protocol

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.

The basic and enhanced methods are described next.

4.2.1 Basic Method

4.2.1.1 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 secondary market impacts, described below.

4.2.1.1.1 Data Collection

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 program. Responses provided by participants generally can be categorized as follows:

1. Kept the appliance.
2. Sold the appliance to a private party (either an acquaintance or through a posted advertisement).
3. Sold or gave the appliance to a used-appliance dealer.
4. Gave the appliance to a private party, such as a friend or neighbor.
5. Gave the appliance to a charity organization, such as Goodwill Industries or a church.
6. Had the appliance removed by the dealer from whom the new or replacement appliance was obtained.
7. Hauled the appliance to a landfill or recycling center.
8. Hired someone else to haul the appliance 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 appliance would have remained active within the participating home ("I would have kept it and continued to use it") or used elsewhere within the Program Administrator'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 by the evaluators.[45]

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. Such would-be acquirers could:

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 appliances 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 appliance had it been readily available, but because the appliance 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 appliances 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 the would-be acquirers would find another unit, while others would not.** This possibility reflects the awareness that some acquirers were in the market for an appliance 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 within a Program Administrator's service territory 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.

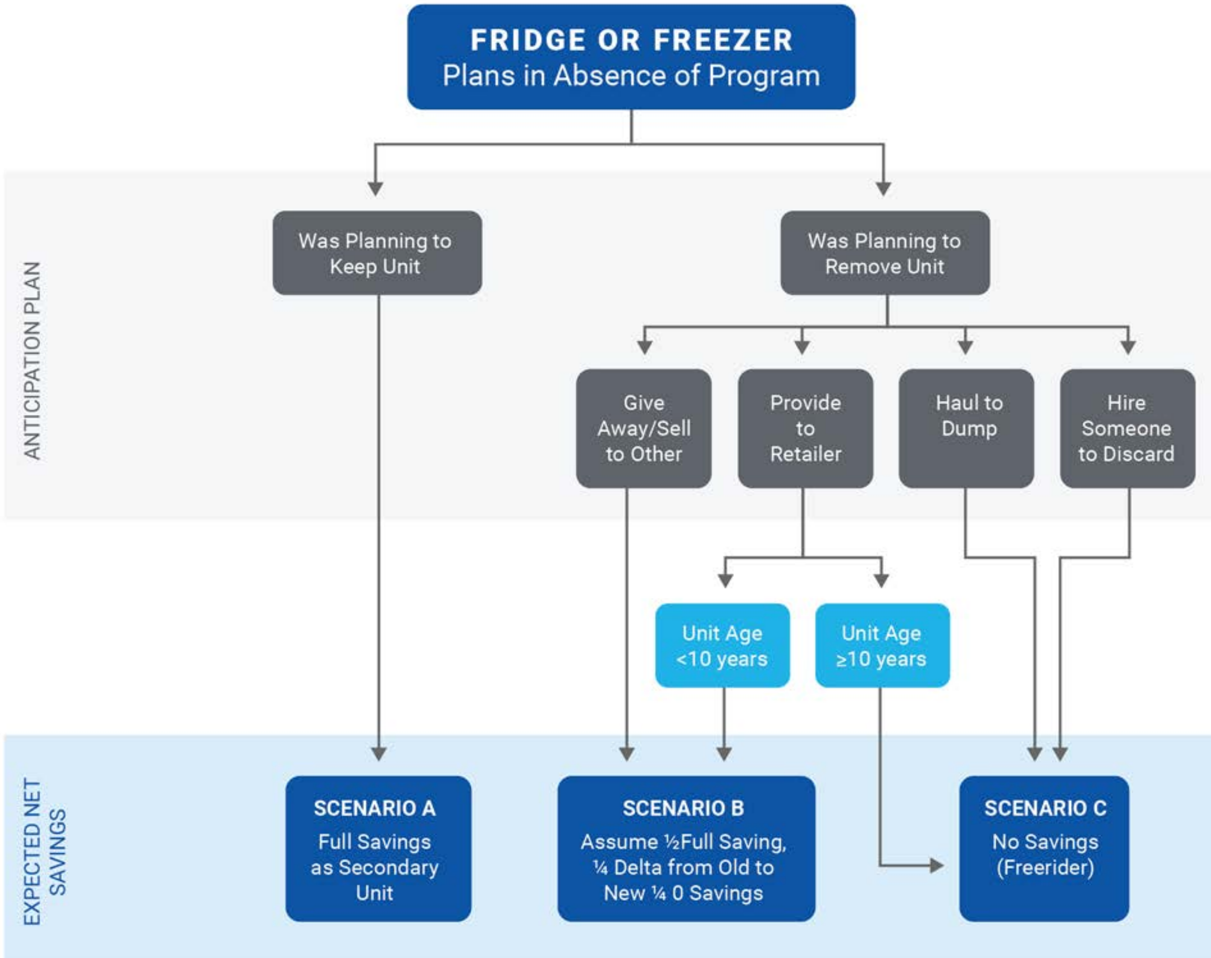
Once the proportion of would-be acquirers who are assumed to find alternate units is determined, the next question is whether the alternate unit was likely to be another used appliance (similar to those recycled through the program) or, with fewer used appliances presumably available in the market due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

4.2.1.2 Integrating Free Ridership and Secondary Market Impacts

The flow chart shown in Figure 4-2 illustrates how net savings will be derived for an ARP. As shown, below, expected savings fall into three different scenarios.

[45] 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.

Figure 4-2. Appliance Retirement Scenarios



Source: Adapted from the Pennsylvania Statewide Evaluator Common Approach for Measuring Net Savings for Appliance Retirement Programs, Guidance Memo-026, March 14, 2014.

4.2.1.3 Scoring 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:

$$FR = (free\ ridership\ and\ secondary\ market\ impacts\ \% - induced\ replacement\ \%)$$

Table 4-4 demonstrates the proportion of a sample population classified into each of the eight potential (Tertiary Classification) categories and the resulting weighted net savings.

Table 4-2. 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 No Induced Replacement	N/A	25%	1,026	0	1,026
Would have removed unit	Scenario B: Transferred No	N/A	30%	1,026	520	506

Primary Classification	Secondary Classification	Tertiary Classification	Population (%)	UEC (kWh) w/out Program	UEC (kWh) w/ Program	kWh Savings
	Induced Replacement					
	Scenario C: Removed from Service	Recycled/ Destroyed	20%	0	0	0
		Retailer would Recycle	13%	0	0	0
Net Savings (kWh)						475

*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 (Unit Energy Consumption) values presented in the table represent example values, factoring in part-use.

4.2.2 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 EPY6 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 appliances). It is difficult to answer this question with certainty, absent Program Administrator-specific information regarding the change in the total number of appliances (overall and used appliances specifically) that were active before and after program implementation. In some cases outside of Illinois, 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 Program Administrator-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 that 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, this 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 it is a reasonable proxy for the proportion of opportunistic acquirers. This proportion would replace the 50% default assumption (scenario C in Figure 4-2) of would-be-acquirers that would or would not find an alternate unit.
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 EPY6 evaluation and used qualitatively to validate participant results.

4.3 Residential Upstream Lighting Protocol

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 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. For programs such as this, in store customer surveys are preferable to the traditional self-report telephone surveys that ask customers to recall their past light bulb purchases. Light bulbs are a small and relatively insignificant purchase for most people, thus the recall bias could be substantial.

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.[46] 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.

4.3.1 Basic Method

4.3.1.1 Free Ridership

Free ridership for this program is calculated as 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 No-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

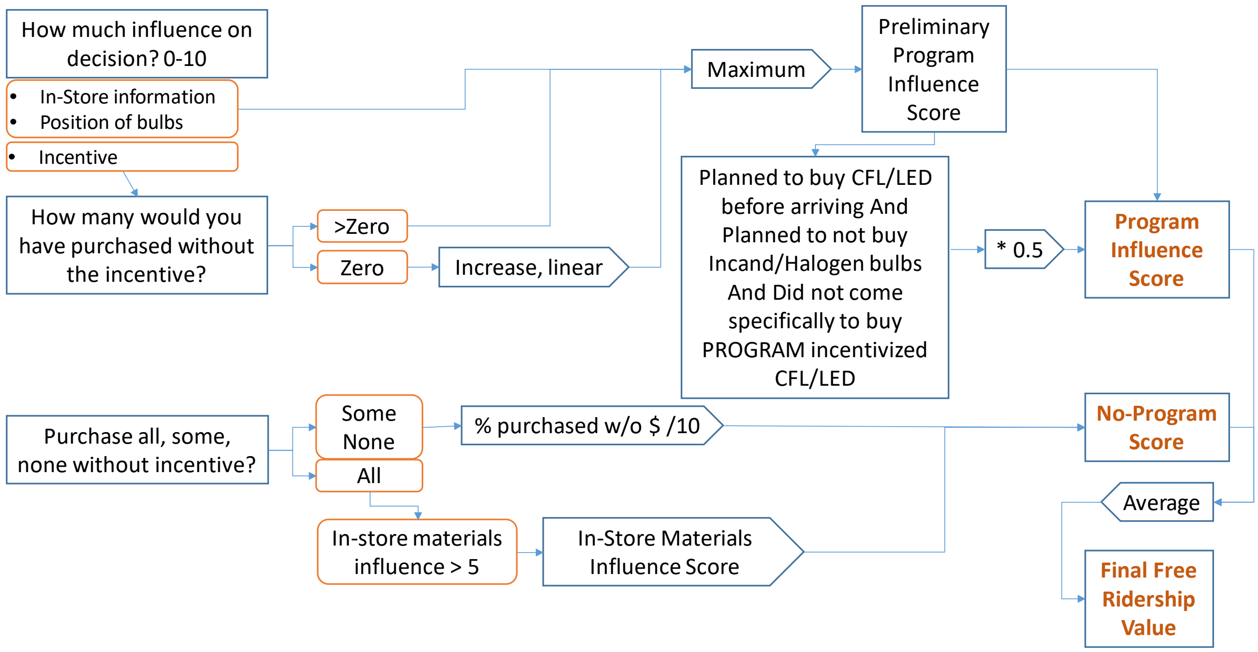
[46] ComEd has used this method since EPY2. Ameren began using it in EPY5.

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

2. The No-Program Score is used to estimate how many program bulbs a survey respondent would have purchased in the absence of the residential lighting program.

Figure 4-3 illustrates the scoring algorithm for Residential Upstream Lighting Free Ridership via In-Store Intercepts.

Figure 4-3. Residential Upstream Lighting Free Ridership via In-Store Intercept



4.3.1.2 Data Collection

To estimate free ridership, the evaluation teams will 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 No-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. Program administrator-incentivized bulbs (Yes/No)
3. Influence of various program factors:
   a. Program incentive
   b. In-store information (printed materials or information from Program Administrator representatives or retail personnel)
   c. Positioning of discounted bulbs within the store

Primary No-Program Score Questions

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

4.3.1.3 Scoring Algorithms

Using the data collected from program participants during the in-store intercept surveys, Program Influence and No-Program Scores are calculated for each survey respondent and then combined to estimate a respondent-specific Free Ridership Score.

4.3.1.3.1 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 should be increased for survey respondents (using a linear decreasing function)[47] who indicated that, absent the incentive, they would not have purchased any of the program bulbs they were purchasing that day.

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 who had not come to the store specifically to buy Program Administrator-incentivized program bulbs, should 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.

4.3.1.3.2 Calculation of the No-Program Score

The No-Program Score is based on whether a respondent states they would have purchased all, some, or none of the program-discounted bulbs in the absence of Program Administrator incentives. Respondents reporting they would have purchased all of the efficient bulbs without the incentive should be considered free riders and receive a No-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 No-Program Score of 10, the maximum. Respondents reporting they would have purchased some of the efficient bulbs without the incentive should be assigned a No-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 Program Administrator had a moderate to high influence on their decision, should have their No-Program Scores adjusted to equal the level of influence they attributed to these program-sponsored informational materials.

4.3.1.4 Calculation of Free Ridership

The Free Ridership rate is calculated as follows:

Free Ridership = 1 – (Program Influence Score + No-Program Score)/20

Using the calculated Program Influence and No-Program Scores, Free Ridership is calculated as one minus the sum of the two scores (Program Influence Score plus No-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 No-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 this free ridership assessment methodology.

[47] The function, adjusted monetary score = (monetary score + 10)/2, increases the monetary score using a decreasing linear function. This function results 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 past Illinois evaluations, this adjustment has typically changed less than 10% of all monetary scores.

4.3.2 Participant Spillover

For this program, 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.

4.3.2.1 Data Collection

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

Primary Program Influence Score Question

1. Influence of the lighting program or in-store information on the customer's decision to purchase non-discounted CFLs or LEDs. (0 to 10 scale where 0 is not at all influential and 10 is extremely influential)

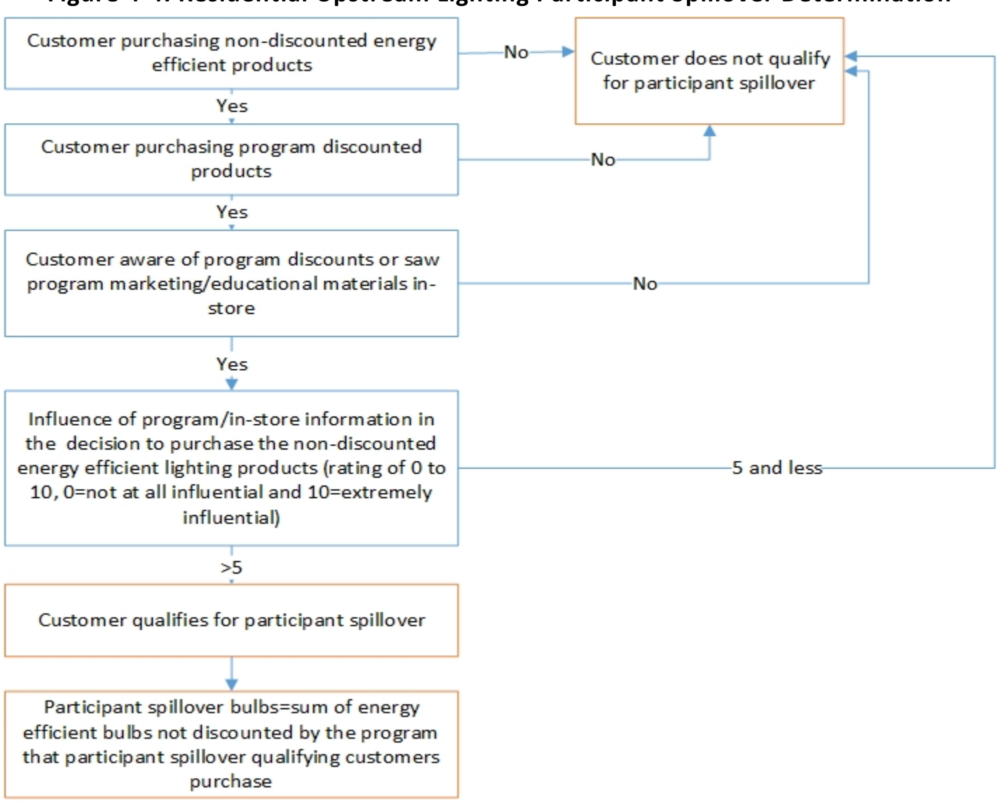
4.3.2.2 Scoring Algorithm

To estimate participant spillover, the number of program-influenced, non-discounted efficient bulbs (CFLs or LEDs) purchased by program participants is divided by the total number of program bulbs purchased by these program participants. This results in the Participant Spillover Rate.

Step 1: Estimate the total number of non-discounted energy efficient bulbs purchased by respondents that had also purchased program-discounted bulbs and were influenced by the program. Respondents who gave a rating of greater than 5 on the program influence question are considered to be influenced by the program.

Figure 4-4 below provides a visual depiction of the process of qualifying non-discounted bulbs as participant spillover bulbs.

Figure 4-4. Residential Upstream Lighting Participant Spillover Determination



Step 2: Calculate the total number of program-discounted bulbs purchased by summing the number discounted bulbs purchased by all respondents.

Program Bulb Purchases = sum(Number of Discounted CFLs or LEDs purchased)

Step 3: Calculate the spillover rate by dividing the total number of spillover bulbs purchased by the total number of program-discounted bulbs purchased.

Spillover Rate = Spillover Purchases/Program Purchases

4.3.3 Nonparticipant Spillover

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.

4.3.3.1 Data Collection

Data collected during in-store intercept surveys with customers purchasing efficient bulbs not discounted by the program should be used to estimate nonparticipant spillover. During these surveys, customers purchasing non-discounted efficient bulbs (CFLs or LEDs) and not purchasing any program-discounted bulbs should be asked questions about awareness of the program discounts and point-of-purchase program marketing and educational materials. These questions are used to determine whether the residential lighting program influenced their purchases of non-discounted efficient bulbs.

Primary Program Influence Score Question

1. Influence of the lighting program or in-store information on the customer's decision to purchase non-discounted CFLs or LEDs. (0 to 10 scale where 0 is not at all influential and 10 is extremely influential)

4.3.3.2 Scoring Algorithm

The nonparticipant spillover scoring algorithm involves estimating the total number of nonparticipants, the incidence of nonparticipants in the sample, the total number of nonparticipant spillover bulbs, and the average number of nonparticipant spillover bulbs per customer in the sample, and then extrapolating the sample estimates to the population of the utility customers. Below are the steps used to calculate the nonparticipant spillover rate.

Step 1. Determine nonparticipant spillover in the sample by following the steps outlined below.

A. Determine the total number of nonparticipating customers in the survey sample:

Nonparticipating customers (survey) = customers who did not purchase any program-discounted energy efficient lighting products. These customers may have purchased non-discounted energy efficient lighting products, less efficient lighting products or both.

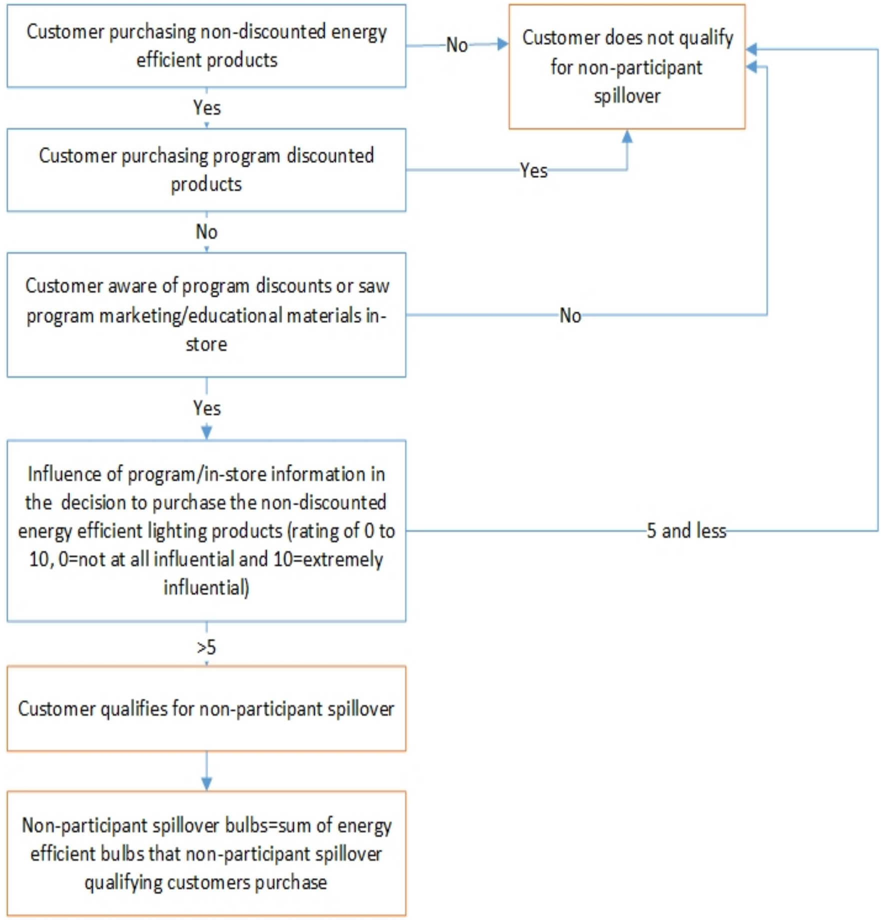
B. Determine the incidence of nonparticipating customers in the survey sample by dividing non-participating customers by total customers in the sample:

Incidence of nonparticipating customers (survey)=Nonparticipating customers (survey)/total customers (survey)

C. Determine total number of nonparticipant spillover bulbs by summing CFLs and LEDs not discounted by the program that were purchased by nonparticipating customers who were aware of the program discounts or marketing promoting energy efficient lighting and were influenced by it. Spillover qualifying bulbs are those purchased by customers who rate the program's influence as greater than 5. The graphic below provides a visual depiction of the process of qualifying non-discounted products as spillover products.

Figure 4-5 below provides a visual depiction of the process of qualifying non-discounted bulbs as nonparticipant spillover bulbs.

Figure 4-5. Residential Upstream Lighting Nonparticipant Spillover Determination



D. Determine the average number of non-participating spillover bulbs per non-participating customer by dividing the total number of non-participating spillover bulbs in the survey by the total number of non-participating customers in the survey.

Average number of nonparticipating spillover bulbs (survey)=total number of nonparticipant spillover bulbs (survey)/nonparticipating customers (survey)

Step 2. Extrapolate nonparticipant spillover to the population

A. Determine the total number of nonparticipating customers in the population by applying the nonparticipant incidence rate from the sample to the population

Total number of nonparticipating customers (population)=Utility residential customer count incidence of nonparticipating customers (survey)*

B. Determine the total number of spillover bulbs by multiplying the average number of spillover bulbs per nonparticipating customer in the survey by the total estimate of nonparticipating customers

*Total number of nonparticipant spillover bulbs=Average number of nonparticipant spillover bulbs (survey)*total number of nonparticipating customers (population)*

Step 3. Calculate nonparticipant spillover rate by dividing the total number of nonparticipant spillover bulbs in the population by the total number of program-discounted bulbs:

Nonparticipant spillover rate=total number of nonparticipant spillover bulbs/total number of program discounted bulbs

4.3.3.3 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.

4.4 Prescriptive Rebate (With No Audit) Protocol

Prescriptive Rebate programs typically offer predetermined rebates to residential customers for purchasing measures such as high-efficiency furnaces, clothes washers, brushless/electronically commutated motors (ECMs), boilers, boiler reset controls, water heaters, air-source heat pumps (ASHPs), ground-source heat pumps (GSHPs), central air conditioners (CACs), programmable thermostats, smart thermostats, insulation, air sealing, duct sealing, and desktop power management software. The program may require installation by a registered program ally, but it does not require a home audit (although purchases may be made in response to an audit).

These programs encourage consumers to undertake the following:

- Purchase higher-efficiency equipment than they otherwise would have, had they shopped for such equipment at the same time (replace on burnout); and
- Replace operating but inefficient equipment with higher-efficiency equipment (early replacement).

The basic method for estimating free ridership and participant spillover (See Section 4.1.2) for these programs uses a participant self-report, based on a standard battery of questions. An enhanced method may utilize trade ally surveys to provide another quantitative assessment, which may be triangulated with the basic method approach. As discussed further in Section 5.2, trade ally surveys may also be used to assess nonparticipant spillover.

4.4.1 Basic Method

4.4.1.1 Free Ridership

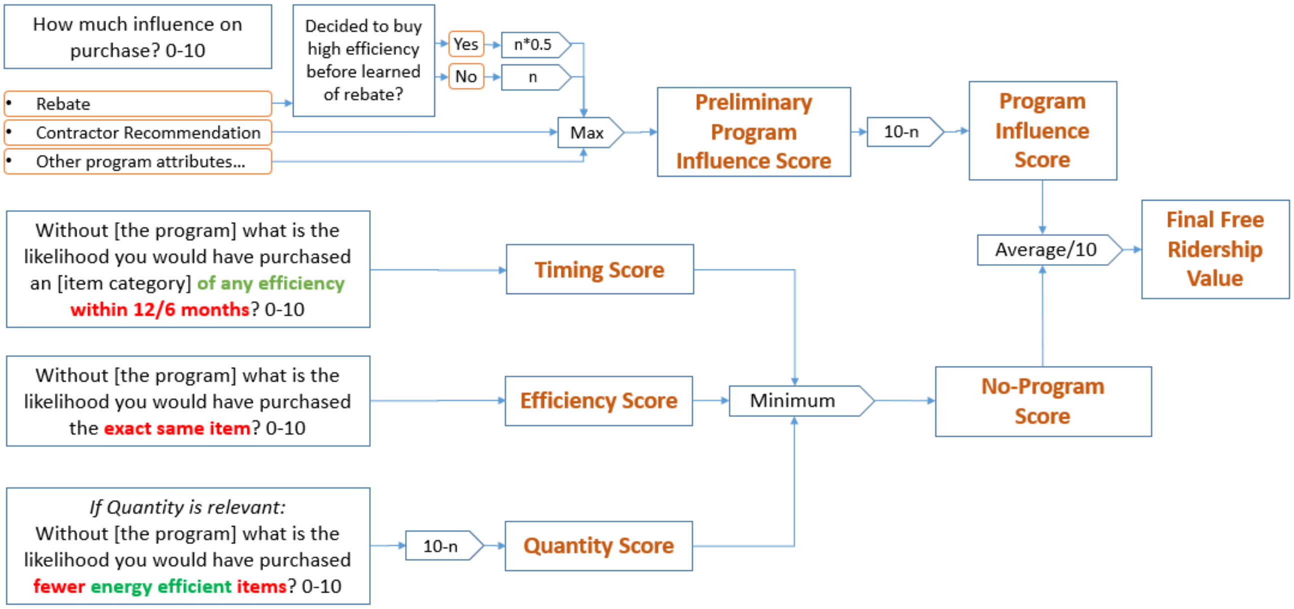
The free ridership assessment battery is brief to avoid applying an undue survey burden, yet it seeks to reduce self-report biases by including two main free ridership components:

- A Program Influence component, based on the participant's perception of the program's influence on carrying out the energy-efficient project; and
- A No-Program component, based on the participant's intention to carry out the energy-efficient project without program funds.

When scored, each component assesses the likelihood of free ridership on a scale of 0 to 10, with the two scores averaged and for a combined total free ridership score. As different and opposing biases potentially affect the two main components, the No-Program component typically indicates higher free ridership than the Program Influence component. Therefore, combining these decreases the biases.

Figure 4-6 illustrates the scoring algorithm.

Figure 4-6. Residential Prescriptive Rebate (With No Audit) Free Ridership



4.4.1.1.1 Calculation of the Program Influence Score

Program influence is assessed by asking respondents, on a scale from 0 (not at all important) to 10 (extremely important), how important they found various program elements were on their undertaking the project the way they did. The number of elements included will vary, depending on the program's design. Logic models, program theory, and staff interviews typically inform the list of elements. Programs typically use the following elements to influence customer behavior: information; incentives or rebates; interaction with program staff (i.e., technical assistance); interaction with program proxies, such as members of a trade ally network; building audits or assessments; and financing.

In addition to asking about specific program influences, surveys ask respondents whether they planned to purchase a high-efficiency version of the product before learning of the rebate program. The respondent's rating of the rebate's influence is adjusted by 0.5 for those answering the question "yes."[48] Evaluators should conduct a sensitivity analysis around the use of this adjustment and present it in the report.

The Preliminary Program Influence Score equals the maximum influence rating for any program element rather than, for example, the mean influence rating. This is based on the rationale that if any given program element had a great influence on the respondent's action, then the program itself had a great influence, even if other elements had less influence.

An inverse relationship occurs between high program influence and free ridership: the greater the program influence, the lower the free ridership. The Program Influence (PI) Score = 10 - Preliminary Program Influence Score.

4.4.1.1.2 Calculation of the No-Program Score

The No-Program (NP) Score is based on three measures of the likelihood of a participant purchasing the exact same item(s) at the same time in the absence of the program. Each of these likelihood measures are assessed on a 0-10 scale in which 0 means not at all likely and 10 means very likely.

First, the participant should be asked their likelihood of purchasing an item of *any efficiency* within 12 or 6 months

[48] The Illinois NTG Working Group discussed using this question to check for consistencies rather than adjusting the score. The NTG working group agreed that it is preferable not to directly ask about conflicting language with residential customers and to utilize an open ended question instead to assess possible reasons for conflicting statements. It is the experience of the NTG working group members that residential customers tend to be more impatient with these types of questions and can typically respond easier to an open-ended question about their motivations.

(12 months for a single or big ticket item and 6 months for less expensive items) for the Timing (T) Score. Participants who were influenced by the program to replace still-functioning equipment will likely give a low score to this question, while participants who needed to replace burned out equipment will give a high score. This measure enables the analysis to use a single algorithm for both early replacement and replace-on-burnout scenarios.

Next, the participant should be asked a key question that asks the respondent to gauge their likelihood of purchasing the *exact same item* (e.g., make, model, efficiency) had the program not existed. This measure forms the Efficiency (E) Score. A respondent stating the likelihood of purchasing the same exact item as a 5 on a scale of 0 to 10 is assigned an Efficiency Score of 5.

If multiple quantities of an item are purchased, the respondent should be asked about the likelihood of purchasing fewer energy-efficient items. The response to this question is subtracted from 10 to compute the Quantity (Q) Score.

The No-Program Score is the minimum of the Timing, Efficiency, and (if applicable) Quantity Scores. Finally, the No-Program Score is averaged with the Program Influence Score to calculate the Final Free Ridership Value.

$$No\ Program\ Score\ (NP) = Min(T, E, Q)$$

$$Free\ Ridership\ (FR) = Mean(PI, NP)$$

4.4.1.1.3 Consistency Checks

To address the possibility of conflicting responses (i.e., low intention score and high influence score), the survey should include consistency checks that, at a minimum, ask participants an open-ended question to address the program's influence. For example:

- In your own words, please tell me the influence the program had on your purchase of the <insert measure name>.

In this case, the evaluation analyst will assess the response to this open ended question and its consistency with the other questions, and, if warranted based on clear additional information, they will adjust the score based on expert judgement. If an inconsistency exists and the open-ended response does not resolve the inconsistency, the respondent will be removed from the calculation. All instances of this occurring should be documented in the final report. Additional consistency checks, triggered and resolved within the survey with additional questions to participants, remain optional.

Missing responses to specific questions should be treated as "missing" for that particular question, but the observation or case will be retained in the analysis. Evaluation reports should note if this affects more than 5% of the responses.

4.5 Single-Family Home Energy Audit Protocol

Single-Family Home Energy Audit programs (or energy assessment programs) seek to secure energy savings for residential customers by providing audits, direct-install measures, and incentives for additional energy efficiency opportunities. The participation process generally begins with an energy audit, performed by a program-affiliated companies or individuals; this involves an auditor assessing the customer's home to identify energy-saving opportunities. At that time, the auditor may install free instant-savings measures, such as CFLs, low-flow showerheads, and faucet aerators. Auditors also may educate customers about incentives available through the audit program (e.g., air sealing, insulation) or other Program Administrator-sponsored energy efficiency programs.

For these programs, free ridership and participant spillover (See Section 4.1.2) estimates rely on participant self-reports, gathered through surveys.

4.5.1 Basic Method

Given the multiple components of some audit programs, net impacts should be estimated using survey batteries tailored to a customer's experience (e.g., receipt of free direct-install measures and discounted or rebated measures). The following sections outline the approach for two program components, one dealing with the direct installation of free low-cost measures and a second dealing with envelope measures, such as air sealing and insulation.

4.5.1.1 No-Cost, Direct Install Measures

For free measures directly installed by program staff due to the audit, free ridership calculations should include the following components: Timing, Efficiency, and Quantity.

This approach provides several important benefits, such as deriving a partial free ridership score based on the likelihood that the participant would take similar actions in the absence of the audit. For example, partial scores can be assigned to customers who planned to install the measure, but the program influenced that installation, particularly in terms of timing (e.g., the program might have accelerated the installation) or quantity (e.g., the program might have led to installation of additional program-qualified measures).

Outlines of components and their associated survey questions follow:

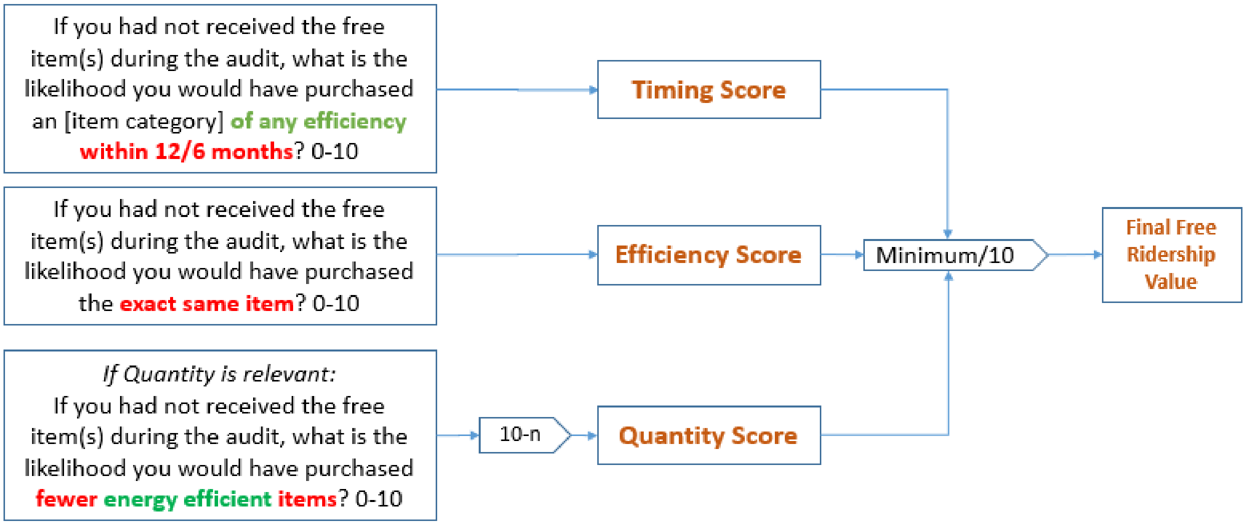
- **Timing (T).** The first question is compute the Timing (T) Score accounts for earlier installation of measures due to the program by asking respondents about their likelihood (0-10 scale) to have installed an item *of any efficiency* within 6 or 12 months, had they not received it through the program (12 months for a single or big ticket item and 6 months for less expensive items).
- **Efficiency (E).** This score reflects the likelihood that customers would have installed the exact same energy-efficient measures, had the program not existed. For free measures, this is based on a question asking respondents to rate the likelihood that they would have installed the exact same measures had they not received them for free through the audit (on a 0 to 10 scale, where 0 is not at all likely and 10 is extremely likely). A higher likelihood value means a higher level of free ridership (i.e., a lower attribution level for the program).
- **Quantity (Q).** The question to compute the Quantity (Q) Score asks respondents about the likelihood that they would have installed fewer measures or performed less weatherization without the program. The response to this question is subtracted from 10 to compute the Quantity Score, as a lower score means a greater likelihood the respondent would have installed the same or a greater number of measures.

Given the low cost of the measures provided through the direct-install component of most audit programs and the number of measures received per participant, efforts have been made to streamline the free ridership battery to reduce the respondent's burden. As such, the overall Final Free Ridership Value per measure can be calculated by taking the minimum of the Timing, Efficiency, and Quantity Scores, as shown in the following equation:

$$Free\ Ridership\ (FR) = Min(T, E, Q)$$

Figure 4-7 illustrates the algorithm for no-cost measures.

Figure 4-7. Single-Family Home Energy Audit Free Ridership—No Cost Measures



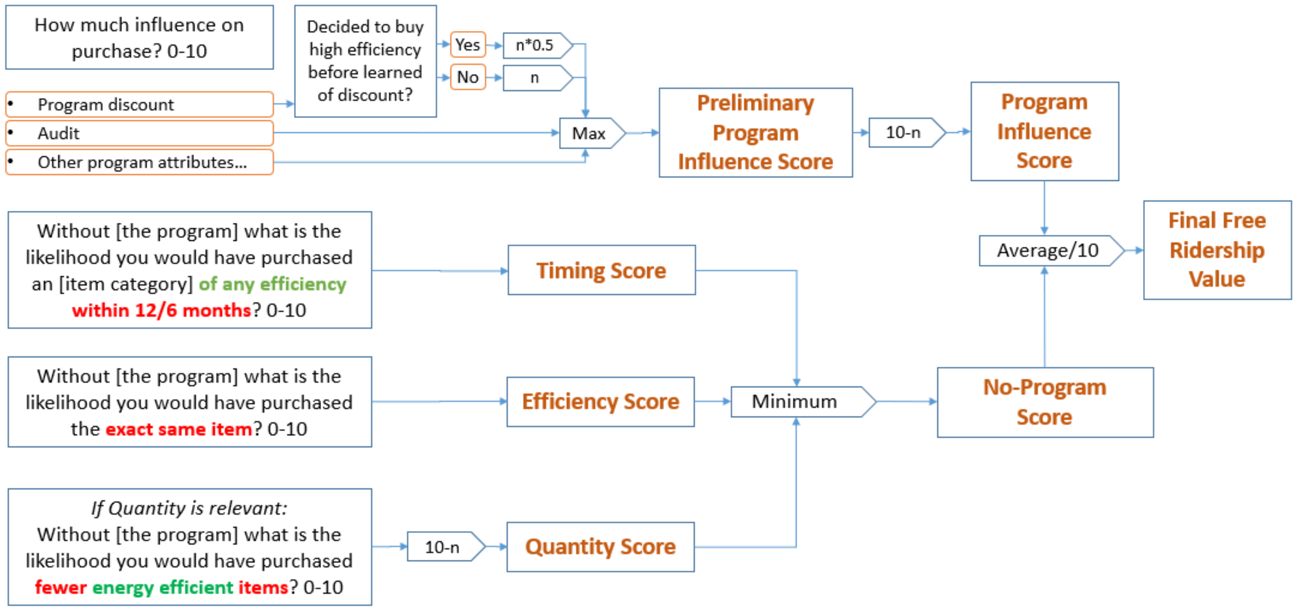
4.5.1.2 Rebated/Discounted Measures

Estimating NTG for rebated measures (typically for building shells) requires a more rigorous process than estimating NTG for free direct-install measures. In particular, the approach integrates an assessment of various program components that may have influenced the participant's installation of the measures. For discounted envelope measures, the basic free ridership factor consists of the following two components:

- A Program Influence component, based on the participant's perception of the influence of various program elements—including the discount and the audit itself—on carrying out the energy-efficient project; and
- A No-Program component, based on the participant's likelihood of purchasing the exact same items at the same time in the absence of the program.

The free ridership method for discounted measures is identical to that used in the Prescriptive Rebate (With No Audit) protocol, with the one exception that the questions about program influence should be sure to include the audit itself as one of the program attributes. Evaluators should refer to Section 4.4.1.1 for details of the method. Figure 4-8 illustrates the algorithm for discounted measures.

Figure 4-8. Single-Family Home Energy Audit Free Ridership—Discounted Measures



4.5.1.3 Consistency Checks

To address the possibility of conflicting responses (e.g., the high likelihood to install the same measure in the program's absence and the high importance of program factors), the survey should include consistency checks that, at a minimum, ask participants an open-ended question to address a program's influence, such as the following:

- In your own words, please tell me the influence the program had on your purchase of the <insert measure name>.

For low or no-cost, direct-install measures, surveys should include two questions to assess a program's influence on the respondent. The first should be asked at the beginning of the NTG battery, and the second should be asked at its conclusion. Questions include the following:

- Prior to the audit, had you purchased any <measures>? Y/N
- IF YES AND LIKELIHOOD TO INSTALL WITHOUT THE PROGRAM IS <7: Given that you had purchased <measures> before receiving the audit, why didn't you purchase additional <measures> on your own without the program? [OPEN END]
- IF NO AND LIKELIHOOD TO INSTALL WITHOUT THE PROGRAM IS >6: Given that you have not purchased <measures> before, why were you likely to purchase <measures> on your own without the program? [OPEN END]

In both cases, the evaluation analyst will assess responses to open ended questions and their consistency with the other questions; if warranted, based on clear additional information, the evaluator will adjust the original question score if required. If inconsistency occurs and the open-ended response does not resolve it, the original question response will be removed from the calculation. Final reports should document all instances of such adjustments. Optionally, additional participant questions can be included to trigger and resolve additional consistency checks.

Missing responses to specific questions (e.g., don't know or refused) should be treated as "missing" for those particular questions, but the analysis retains the observation or case. The evaluation reports should note if this affects more than 5% of responses.

4.6 Multifamily Protocol

Multifamily energy efficiency programs typically offer direct installation of low-cost, energy-efficient measures in multifamily dwelling units, in addition to rebates for common area lighting retrofits, air sealing, insulation, and improvements to HVAC systems and controls. These programs have various target audiences from owners, managers, or developers of market rate multifamily housing to those operating lower income or assisted living housing. Across these groups, properties must generally have a minimum of between three and five units to qualify for the programs.

Most multifamily program savings are typically achieved by encouraging customers to install higher-efficiency equipment than they would have installed on their own. However, programs may also encourage early replacement of still functioning equipment that is less efficient, thus impacting the timing of the installation, so that savings is realized earlier. The incentive may also make it more affordable for customers to install a greater number of high-efficiency measures.

The basic method for estimation of free ridership and participant spillover (See Section 4.1.2) for these types of programs is based on participant self-report gathered through surveys. For common area and building shell components of the program, participants are property managers and owners responsible for building maintenance and renovation. However, depending on the program design for the in-unit component of the program and specifically the installation of efficient lighting, participating in the program (i.e., install program measures) may be driven by either property managers/owners or tenants or, potentially, both. This distinction is due to the fact that in some market-rate apartments, the tenant is responsible for decisions related to the installation of program measures, including light bulbs, while this is not common practice in income-qualified or assisted-living settings. For other in-unit measures, such as faucet aerators and low-flow showerheads, evaluators interview property managers/owners regarding program influence, as these measures are typically direct installed by program staff, and there is a limited likelihood of tenants making changes to these features.

4.6.1 Basic Method

Estimating NTG for rebated measures requires a more rigorous process than estimating NTG for free direct-install measures. In particular, the approach integrates an assessment of various program components that may have influenced the participant's installation of the measures. For discounted measures, the basic free ridership factor consists of the following two components:

- A Program Influence component, based on the participant's perception of the influence of various program elements—including the discount and the audit itself—on carrying out the energy-efficient project; and
- A No-Program component, based on the participant's likelihood of purchasing the exact same items at the same time in the absence of the program.

The free ridership method for discounted measures is identical to that used in the Prescriptive Rebate (With No Audit) protocol, with the one exception that the questions about program influence should be sure to include the audit itself as one of the program attributes. Evaluators should refer to Section 4.4.1.1.1 and 4.4.1.1.2 for details of the method. Figure 4-9 and Figure 4-10 also illustrate the algorithms for CFL/LED and non-CFL/non-LED measures[49].

[49] Evaluators should word the survey questions to reflect whether measures were free or purchased with an incentive.

Figure 4-9. Multifamily Free Ridership—Non-CFL/Non-LED Measures

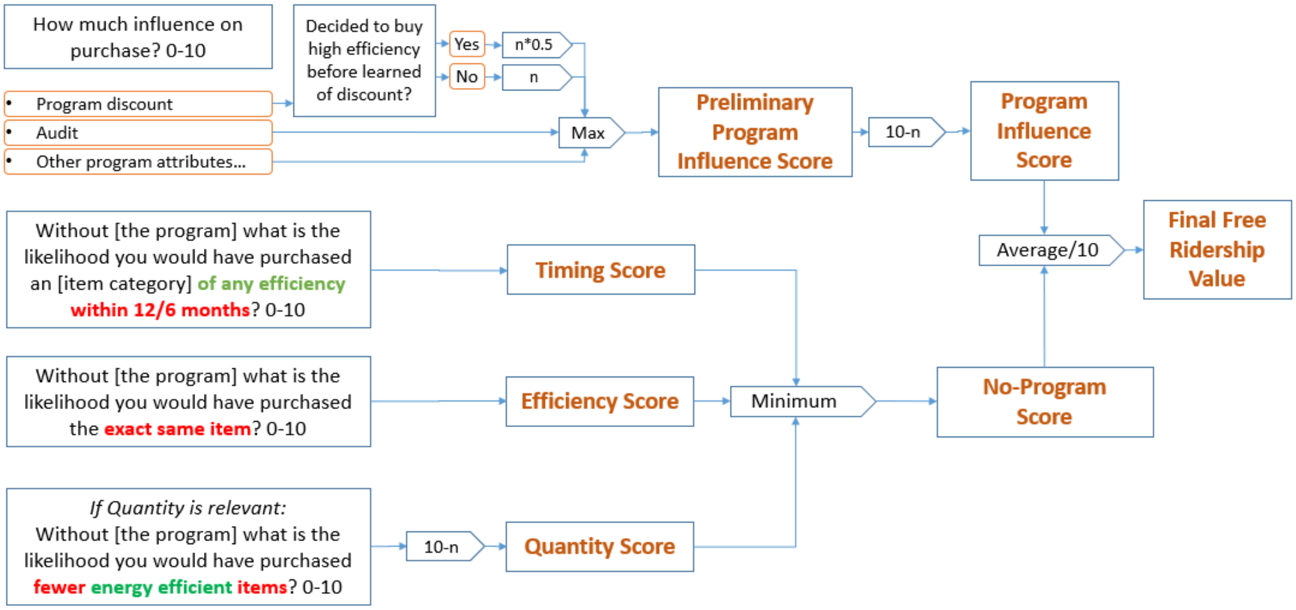
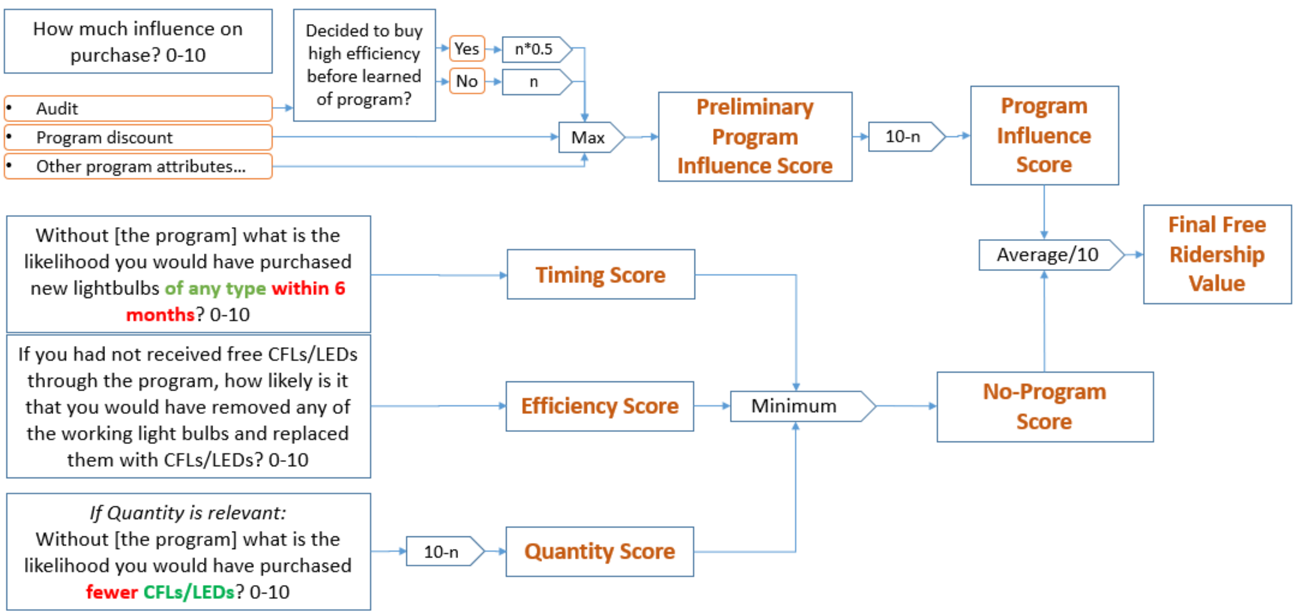


Figure 4-10. Multifamily Free Ridership for Property Managers—CFL/LED Measures



4.6.1.1 Consistency Checks

To address the possibility of conflicting responses (e.g., high likelihood to install the same measure without the program, high importance to program factors), the survey should include consistency checks that, at a minimum, ask participants an open-ended question to address the program's influence. For example[50]:

- In your own words, please tell me the influence the program had on your purchase of the <insert measure name>.

The evaluation analyst will assess the responses to the open ended questions and their consistency with the other survey questions, and, if warranted based on clear additional information, will adjust the original question score. If

[50] Evaluators should word the consistency check questions to reflect whether measures were free or purchased with an incentive.

the open-ended response does not resolve the inconsistency, responses to the original question should be removed from the calculation. The survey may include additional consistency check triggers and resolutions through additional participant questions. The final report should document how often the consistency check rules were triggered, how often adjustments were made to scores, and how often inconsistencies could not be resolved.

Missing responses to specific questions (including don't know or refused) should be treated as missing for that particular question, but the analysis should retain that observation or case. Evaluation reports should note if this affects more than 5% of the responses.

4.6.1.2 Data Collection

A participant survey should be used as the primary source of data collected for estimating free ridership in residential multifamily programs. As discussed, evaluators may field surveys with owners, property managers, or tenants, depending on a program's design and theory. Determining the appropriate audience from which to gather information for estimating free ridership depends on the program's design, and, ultimately, the party responsible for deciding to install specific program measures.

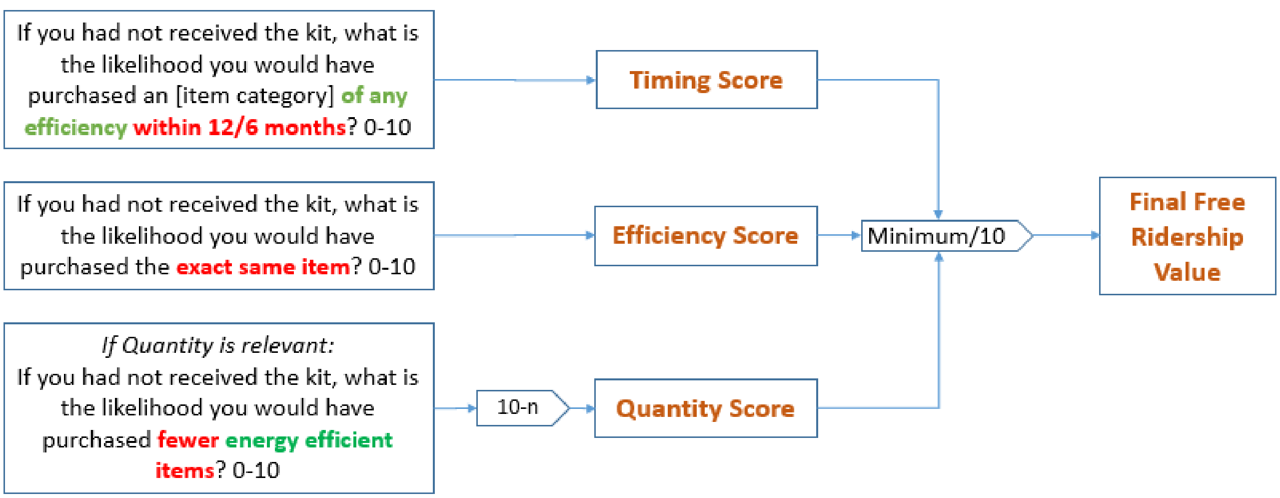
4.7 Energy Saving Kits and Elementary Education Protocol

Energy Saving Kits and Elementary Education Programs aim to secure energy savings through the distribution of kits containing various energy-saving measures, including (but not limited to): high-efficiency lighting (CFLs or LED lamps); bathroom and kitchen faucet aerators; and low-flow showerheads. Energy Saving Kits operate as an opt-in program; customers can request a kit by completing an Internet or phone application. Elementary Education Program participants do not request a kit as kits are distributed to all students in a classroom.

Free ridership and participant spillover (See Section 4.1.2) estimations for both programs rely upon participant self-report information gathered through surveys, despite the differences in distribution models. This methodology can be used for other energy-saving kit programs, including kits with alternative distribution methods (e.g., kits dropped off at a participant's home).

The following section contains a description of the basic NTG method used. Figure 4-11 illustrates the method.

Figure 4-11. Energy Saving Kits and Elementary Education Free Ridership



4.7.1 Basic Method

Free ridership calculations should include the following components: No-Program, Timing, and Quantity.

This approach provides several important benefits, such as the ability to derive a partial free ridership score based on the likelihood that similar actions would have taken place, even if the participant had not received a kit. For instance, partial scores can be assigned to customers with plans to install the measure, but the program at least influenced that installation, particularly in terms of timing (e.g., the program might have accelerated the installation) or quantity (e.g., the program might have led to the installation of additional measures).

Outlines of components and their associated survey questions follow:

- **Timing (T).** The first question is compute the Timing (T) Score accounts for earlier installation of measures due to the program by asking respondents about their likelihood (0-10 scale) to have installed an item *of any efficiency* within 6 or 12 months, had they not received it through the program (12 months for a single or big ticket item and 6 months for less expensive items).
- **Efficiency (E).** This score reflects the likelihood that customers would have installed the exact same energy-efficient measures, had the program not existed. This is based on a question asking respondents to rate the likelihood that they would have installed the exact same measures had they not received them for free through the kit (on a 0 to 10 scale, where 0 is not at all likely and 10 is extremely likely). A higher likelihood value means a higher level of free ridership (i.e., a lower attribution level for the program).
- **Quantity (Q).** The question to compute the Quantity (Q) Score asks respondents about the likelihood that they would have installed fewer measures without the program. The response to this question is subtracted from 10 to compute the Quantity Score, as a lower score means a greater likelihood the respondent would have installed the same or a greater number of measures.

Given the low cost of measures provided in the energy-saving kits as well as the number of measures included in each kit, efforts have been made to streamline the free ridership battery to reduce the respondent's burden. As such, the overall Final Free Ridership Value per measure can be calculated by taking the minimum of the Timing, Efficiency, and Quantity Scores, as shown in the following equation:

$$Free\ Ridership\ (FR) = \ Min(T, E, Q)$$

Missing responses to specific questions (e.g., don't know or refused) should be treated as "missing" for that particular question. Despite missing responses, the case will be retained in the analysis (pairwise deletion). The evaluation reports should present the percent missing for each of the three questions.

4.7.1.1 Data Collection

Evaluators should use a participant survey as the primary data collection source for estimating free ridership in Energy Saving Kits and Elementary Education Programs. As a general rule, a free ridership rate should be calculated for each separate kit component, and then be weighted by savings to determine the program-level results.

4.8 Residential New Construction Protocol

Residential New Construction programs typically offer builder training, technical information, marketing materials, and incentives to builders for the construction of eligible homes. Eligible homes must meet specific standards, designed to achieve energy efficiency levels above local building codes. Programs may use different tiers of standards to meet correspondingly different incentives.

The basic method for estimating free ridership and participant spillover for these programs is based on builder participant self-reporting, gathered through surveys.

The following section describes the basic method used.

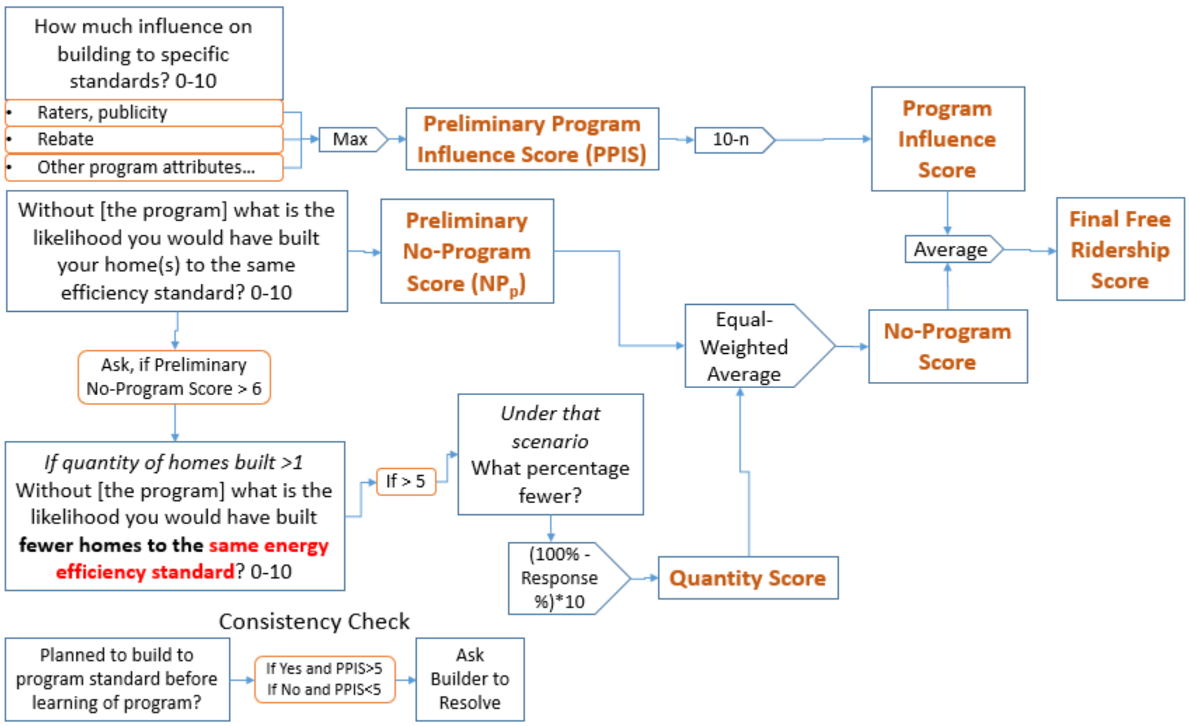
4.8.1 Basic Method

For this program, a free rider is a builder who would have constructed a home at the program's efficiency level in the program's absence. Given the multiple methods available to achieve desired home energy efficiency levels, survey questions consider the builder's likelihood of meeting the same energy efficiency standard, rather than whether or not the builder would have installed certain energy efficiency measures. Figure 4-12 (below) illustrates the method in more detail.

Evaluators assess Program Influence by asking respondents, on a scale from 0 (not at all important) to 10 (extremely important), how important they found various program elements in deciding to build to specific energy efficiency standards. The number of elements included vary, depending on the program's design. Logic models, program theory, and staff interviews typically inform the list of program elements included. Programs typically use the following elements to influence builder actions: marketing materials; incentives or rebates; contacts with HERS Raters; and technical assistance.

In addition to asking about specific program influences, surveys should ask builders whether they planned to build homes to the same standard before learning of the program.

Figure 4-12. Residential New Construction Free Ridership



4.8.1.1.1 Calculation of the Program Influence Score

The Program Influence Score (PI) equals 10 minus the maximum influence rating for any program element rather than, for example, the mean influence rating. This is based on the rationale that if any given program element had a great influence on the respondent's action, the program itself had a great influence, even if other elements had less influence.

4.8.1.1.2 Calculation of the No-Program Score

Evaluators calculate the No-Program score using a set of questions that ask respondents to gauge their likelihood of building homes to the same standards and in the same quantities had the program not existed. Three separate responses are considered in calculating the No-Program Score:

- The likelihood, on a scale of 0 to 10, that the builder would have built their homes to the same efficiency standard (Preliminary No-Program Score (NP_p))
- If that likelihood is greater than 6, the likelihood of fewer homes being built to the same efficiency standard.
- If that likelihood is greater than 6, the response to the question "for that scenario, what percentage of fewer homes would be built to the standard?" (Quantity Score = (100% - % answer) * 10, which will be a number between 0 and 10)

The resulting No-Program (NP) Score is calculated as follows:

$$NP = Mean(NP_p, Q)$$

The overall Free Ridership Value derives from the average of the PI and NP scores, as shown in the following formula:

$$FR = Mean(PI, NP)$$

4.8.1.2 Consistency Checks

To address the possibility of conflicting responses (e.g., the high likelihood to build to the same efficiency standards without the program, the high importance of program factors), the survey should include, at a minimum, consistency checks that ask participants an open-ended question to address the program's influence. For example:

- In your own words, please tell me the influence the program had on your building practices.

If a high (>6) Preliminary Program Influence Score (PPIS) results, yet the builder planned to meet the same efficiency standard prior to learning of the program; or if the Preliminary Program Influence Score is lower (<7), and the builder did not plan to build to the standards prior to learning of the program, the survey should include a question to determine why this occurred, using wording that gets at the following inconsistencies:

- IF Preliminary Program Influence Score is >6 and Builder planned to meet the same efficiency standard prior to learning OF THE PROGRAM: Given that you had plans to meet the standard prior to learning about the program, why do you think the <program elements> were influential in your meeting the standard? [OPEN END]
- IF Preliminary Program Influence Score is <7 and Builder had no plans to meet the same efficiency standard prior to learning of the program: Given that you had no plans to meet the standard prior to learning about the program, why do you think the <program elements> were not more influential in your meeting the standard? [OPEN END]

The evaluation analyst will assess the responses to the open ended questions and their consistency with the other survey questions, and, if warranted based on clear additional information, will adjust the original question score. If the open-ended response does not resolve the inconsistency, responses to the original question should be removed from the calculation. The survey may include additional consistency check triggers and resolutions through additional participant questions. The final report should document how often the consistency check rules were triggered, how often adjustments were made to scores, and how often inconsistencies could not be resolved.

Missing responses to specific questions (including don't know or refused) should be treated as missing for that particular question, but the analysis should retain that observation or case. Evaluation reports should note if this affects more than 5% of the responses.

4.8.2 Participant Spillover

Participant spillover occurs when, due to program participation, a builder increases the energy efficiency of homes built outside the program (but inside a utility's service territory) by adopting certain building practices used in participating homes. Participant spillover can be calculated based on participant builder survey questions that ask builders about homes built within the utility service territory but outside the program. Survey questions ask whether the builder increased the energy efficiency standards of non-program homes after participating in the program, and the number of homes they applied these increased standards to, within the utility's service territory. Depending on the program characteristics, spillover should be measured as changes in specific building practices or as installation of specific measures. The text below assumes the program has been targeted at modifying building practices.

Spillover may be recorded depending on responses to the following questions:

1. How important was your experience in the <PROGRAM ADMINISTRATOR'S> program in your incorporating this building practice your other homes, using a scale of 0 to 10, where 0 is not at all important and 10 is extremely important?
2. If you had not participated in the <PROGRAM ADMINISTRATOR'S> program, how likely is it that you would still have incorporated this building practice using a 0 to 10, scale where 0 means you definitely WOULD NOT have implemented this practice and 10 means you definitely WOULD have implemented this practice?

Responses to the first question establish the Practice Attribution Score 1, and responses to the second question establish the Practice Attribution Score 2. Spillover may be program-attributable for building practices with self-report data meeting the following condition:

Spillover Score = (Practice Attribution Score 1 + (10 – Practice Attribution Score 2))/2 > 5.0

For responses meeting these conditions, an evaluator determines that specific building practices referenced in the question are attributable to the program; otherwise, the evaluator determines that specific building practices referenced in the question are not attributable to the program. The attribution criteria represent a threshold approach, in which energy impacts associated with building practices program participants implement outside the program are either 100% program-attributable or 0% program-attributable.

For each building practice discussed, builders will be asked how they know the building practice is more efficient than other options. If the respondent can identify the building practice as ENERGY STAR or name an efficiency level that the evaluator confirms as above the minimum federal standard, or if they identify a technology that the evaluator can confirm is above the minimum federal standard, this counts towards participant spillover.

Finally, depending on the building practice cited by the builder, follow-up questions should ask customers to provide reasonable information to allow the evaluator to estimate the amount of savings using IL-TRM protocols, such as quantity of appliances or the location and amount of insulation.

To calculate the spillover energy and demand savings for these actions, further questions should be asked to assess the gross savings of the building practice, through the appropriate version of the IL-TRM, if available, and the number of homes to which it applied. To develop the Spillover Rate, the total energy and demand impacts from the sampled participants who implemented efficient building practices in other homes due to participation in the program is summed, and then this sum is divided by the total ex post sample energy and demand impacts:

$$Participant\ Spillover\ Rate\ (PSO) = \frac{Sum\ of\ Energy\ or\ Demand\ from\ Additional\ EE\ Practices}{Sample\ Ex\ Post\ Gross\ Energy\ or\ Demand\ Impacts}$$

The equation used to adjust the Core NTGR based on participant spillover is as follows:

$$NTGR = (1 - FR + PSO)$$

4.8.2.1 Sample

The sample for a spillover survey should be a random sample of current and up to one year previous program participants. Regardless of the year of participation, spillover should be measured within the set of homes that were completed within 12 months of the survey date.

4.8.3 Builder Nonparticipant Spillover

In addition to participant free ridership and spillover, new construction programs may create NPSO through builders exposed to the program but not actually participating. Rather, they implement some or all of the efficiency measures incorporated through the program in order to compete with builders that are participating.[51] NPSO caused by builders can be determined by surveying two groups of builders:

- "Drop out" builders, who participated in the program previously but have not participated in the past 12 months.
- True nonparticipating builders that report they were aware of the program or that other builders were taking steps to improve new home efficiency, but had never participated.

Surveys ask nonparticipating builders if their knowledge of other builders' increased focus on energy efficiency influenced their building practices and in what manner, to quantify the program's impact on nonparticipating homes. The survey questions will first identify specific building practices that go beyond the implemented energy code for the specific jurisdiction in which the builder is active. Table 4-6 lists the latest building energy code in place for most

[51] NPSO also can arise from nonparticipating customers as a direct result of general energy efficiency education and promotion efforts. A separate protocol addresses such NPSO. Care should be taken to ensure the different approaches do not double-count NPSO.

areas of Illinois. Evaluators should make efforts to ensure the building code under enforcement for each jurisdiction is used as the baseline when evaluating spillover savings.

Table 4-8. IECC 2015 Building Energy Code

Component	IECC 2015
Thermostat	Heating 72F; Cooling 75F Programmable Thermostat
Ceiling	U-0.026
Walls	U-0.060
Floors	U-0.033
Slab	R-10, 2ft
Windows	U-0.32
Infiltration	5ACH50
Duct Leakage	4CFM/100CFA
Duct Insulation	R-8 Attic Supply, R-6 Otherwise
Heat Pump	8.2 HSPF
Furnace	80 AFUE
Component	**IECC 2015**
Boiler	82 AFUE
AC	13 SEER
Lighting	75% CFL
Appliances	RESNET Default
Gas Water Heat*	0.58 EF
Electric Water Heat*	0.92 EF

*EF varies based on water heater storage volume and draw pattern; values in table for 40 gallon water heater with medium draw pattern.

For each component that is more efficient than code, the following additional questions are asked:

1. How many homes did you sell in <period> that incorporated this upgrade?
2. Of these homes, how many would have incorporated this upgrade, had the <program> not existed?

Evaluators should ensure that nonparticipant builders receive sufficient time to collect specific data and not rely on "guesses" to respond. Responses should also clarify whether sales counts are specific to the utility service territory in question.

The following steps calculate the program's nonparticipant builder spillover percentage:

1. Compute the difference between the total reported number of efficiency upgrades sold and the total that would have been sold in the program's absence to obtain the total number of upgrades by type of upgrade for that builder.
2. Multiply the total net number of upgrades of each type sold by each surveyed builder by the average gross unit savings for each upgrade type.
3. Sum the result for each builder from the previous step, and weight the results by the ratio of the population of non-active builders to the sample to compute the total spillover energy over the program period.
4. Divide the spillover energy savings by program gross savings.

Should a general population survey be implemented for nonparticipant spillover, care should be taken to ensure spillover is not double-counted.

5 Cross-Sector Protocols

The following sections include protocols that may be applicable to programs in the residential as well as in the commercial, industrial, and public sectors. Table 3-1 Commercial, Industrial, and Public Sector Programs and Table 4-1 Residential and Low Income Programs present information regarding the applicability of these protocols to specific programs.

5.1 Combining Participant and Trade Ally Free Ridership Scores

For a program where trade allies play a prominent role in delivering the energy efficiency measure and promoting the program, an estimate of free ridership from trade allies can be combined with one from participants to form a combined free ridership value. Elsewhere, the NTG Protocol (see Section 3.1.1.3) discusses using trade ally surveys to adjust **project-level** free ridership scores. This section discusses combining a **program-level** free ridership score from trade allies with a program-level free ridership score from participants.

If an evaluation uses this approach, the evaluator's NTG report should present the conditions that support the argument that the combined value is more likely to be reflective of reality. That argument should consider the following topics:

1. **Trade Ally Role.** What role do the trade allies play in the program? How were participating trade allies chosen? How might they differ from nonparticipating trade allies? Why does that support the proposition that their view on free ridership is accurate and reasonably unbiased?
2. **Participant Role**. What role do the participants play in deciding which measures are installed and why does that support the proposition that their view on free ridership is accurate and reasonably unbiased?
    a. For example, the participant's role in the decision **may** be significantly less in some types of programs like new construction or multifamily direct install programs. (The participant free ridership data collection method may already account for this by, for example, treating the building owner as the participant rather than the tenants.)
3. **Market Conditions**. What conditions exist in the market that support the proposition that either the trade allies' view or the participants' view on market behavior may be more accurate?
    a. For example, if the market was in its infancy before the program began and as a result participants' ability to take the energy efficiency action was limited, the trade allies may have a more accurate view on the counterfactual than the participants.
4. **Bias.** What are the hypothesized biases of the participants and trade allies? Where do they stem from? What evidence is there that they exist? How well has the data collection approach sought to mitigate that bias?
5. **Offsetting Bias.** Do the hypothesized biases of participants and trade allies offset each other or do they move the free ridership value in the same direction?

5.1.1 Trade Ally Free Ridership Calculation

The NTG protocols do not yet contain a standardized approach for measuring free ridership from trade allies. That approach should be developed for future versions of the TRM. In the meantime, if an evaluation team decides to estimate trade ally free ridership, they should collaborate with other Illinois evaluators on the survey design and calculation algorithm.

5.1.2 Triangulation

Where appropriate, evaluators should combine participant and trade ally free ridership values by weighting each value in the final result. The weighting of each value should be based on considerations of the likely bias, accuracy, and representativeness of the results. The following presents one approach for determining weights. This is an example only. The evaluator should create an approach appropriate for the program.

Example. Combined participant and trade ally free ridership results by rating the analysis methodology and data collected using responses (rated on a scale of 0 to 10) to the following three questions:

1. All things being equal, on a scale of 0 to 10, with 0 being not at all likely and 10 being extremely likely, how likely is the approach to provide a more accurate estimate of free ridership?
2. Similarly, on a scale of 0 to 10, with 0 being not at all valid and 10 being extremely valid, how valid and reliable is the data collected and the analysis performed (i.e., consider non-response bias, missing data (e.g., whether data collected was based on recollection or record keeping?)
3. On a scale of 0 to 10, with 0 being not at all representative and 10 being extremely representative, how representative is the sample (accounting for sampling error {confidence and precision}, and non-response bias, and any sample frame bias)?

The weight for each free ridership estimate is the average score for that estimate divided by the sum of the average scores for both estimates.

Table 4-5 provides an example scoring illustrating the calculated weights.

Table 5-1. Example Triangulation Weighting Approach

NTG Triangulation Data and Analysis	Participants	Trade Allies
1. How likely is this approach to provide an accurate estimate of free ridership?	6	8
2. How valid is the data collected/analysis?	3	5
3. How representative is the sample?	8	10
Average Score	5.7	9
Sum of Averages	14.7	14.7
Weight	39%	61%

5.2 Spillover Measured Through Trade Allies

Many energy efficiency programs rely on trade allies to help spread program awareness and promote energy efficiency among their customers. Some programs establish lists of participating trade allies and provide trade allies with training, education, and/or marketing materials. Spillover might occur when a trade ally's business practices are influenced by a program but at least some of their energy efficient installations do not receive a program incentive.

For the purposes of measuring trade ally spillover, we define trade allies as (1) retailers, contractors or other market actors who work with end-user customers on the selection and installation of energy-using equipment; and (2) distributors who supply equipment to stores and other market actors, rather than to end-user customers. For the purposes of this section, manufacturers are not included in the definition of trade allies.[52] In addition, we differentiate between the following types of trade allies:

1. Active Trade Allies
    a. Trade allies who were active in the program during the evaluation period and appear in program tracking databases. The tracking data contains information on the quantity of incented measures associated with these trade allies and their savings;
2. Inactive Trade Allies
    a. Trade allies who are on the utility's trade ally list (and have received at least some utility training or education) but who were not active during the evaluation period and do not appear in program savings tracking databases for the evaluation period;
    b. Trade allies who were previously active in the program (and may have been on the utility's trade ally list) but have dropped out; and/or

[52] The exclusion of manufacturers from the definition of trade ally does *not* suggest that manufacturers cannot create spillover. Rather, manufacturers are excluded because the methodologies outlined in this section do not apply to them.

c. Trade allies who have never been active in the program and were never on the utility's trade ally list.

When deciding whether to conduct trade ally spillover research, the evaluator should consider the following:

- **Likelihood of trade ally spillover:** When limited evaluation resources are available, the evaluator should weigh the likelihood of trade ally spillover against the cost of the analysis when prioritizing evaluation efforts. E.g., programs that provide incentives but no training or education are less likely to generate spillover than programs that do provide training or education. Similarly, spillover from active trade allies is generally more likely than spillover from inactive trade allies, and spillover from inactive trade allies who have previously been active in the program is generally more likely than spillover from inactive trade allies who have never been active in the program.
- **Potential double-counting of spillover reported by end-use customers and trade allies:** Spillover from active trade allies and spillover from inactive trade allies are mutually exclusive, i.e., as long as the populations and samples are correctly defined, there is no danger of double-counting spillover from these two groups (see also discussion in Section 2.2). However, if the evaluator measures spillover through trade allies and end-use customers for the same evaluation period, care needs to be taken to avoid double-counting. Evaluators should clearly document potential double-counting of spillover and the steps taken to avoid it.

The following subsections provide suggested approaches for measuring spillover from active and inactive trade allies. Different approaches are outlined for these two groups because of the different types of data available for each of them. For active trade allies, program tracking data contains information on their program activity (the quantity of incented measures associated with each active trade ally and their savings). This data allows for a more rigorous spillover methodology than can be used for inactive trade allies, for whom this information does not exist.

5.2.1 Spillover from Active Trade Allies

Trade allies that are active in an energy efficiency program are more likely to create spillover than inactive trade allies, as their exposure to any program messaging and training/education is likely to be current and therefore more influential on their business practices. Active trade allies may create spillover if their program participation changes their business practices and leads to the completion of non-incented energy efficient projects that would otherwise not have happened. For example, as a result of program training, a trade ally might feel more comfortable talking about the benefits of energy efficiency and recommend energy efficient solutions more often. If these recommendations result in energy efficient projects, but no incentive is claimed, spillover from inactive trade allies may be present.

For active trade allies, the spillover methodology varies slightly for downstream programs and midstream programs. Approaches for both types of program are discussed below.

5.2.1.1 Downstream Programs

Surveys can be used to ask active trade allies if the program influenced their sales of high-efficiency equipment to participating or nonparticipating customers and to quantify the program's impact on their high-efficiency sales. To assess if a sampled trade ally created spillover, the following screening criteria are recommended (the order of these may be adjusted by the evaluator):

1. The percentage of the trade ally's installations/sales that are high efficiency and/or the total volume of high efficiency installations/sales increased since the trade ally became exposed to the program.
2. The trade ally rated the program as important to at least one of these (as described above) high efficiency installation increases.
3. The trade ally installed/sold at least some high efficiency equipment or products during the evaluation period that did not receive an incentive.
4. The trade ally's recommendation was influential in the customers' choice of high efficiency equipment/product over standard efficiency equipment/product in instances where the equipment did not receive a program incentive.

5. The open-ended response about why customers with eligible projects do not receive an incentive supported that the non-incented high efficiency installations can be considered spillover.

Sampled trade allies who do not pass one of the above screening criteria do not qualify for spillover and may be skipped out of the rest of the spillover module.

To quantify spillover for each sampled trade ally, the survey collects information on the percentage of the trade ally's total equipment installations/sales (in terms of projects or measures) that was (1) standard efficiency, (2) high efficiency that DID receive a program incentive, and (3) high efficiency that DID NOT receive a program incentive. Based on these responses, the share of a trade ally's high efficiency installations/sales that received an incentive can be calculated as follows:

$$\textit{\% of TA's High Efficiency Equipment that Received Incentive} = \frac{\textit{\% High efficiency that DID receive a program incentive}}{\textit{\% High efficiency that DID receive a program incentive} + \textit{\% High efficiency that did NOT receive a program incentive}}$$

With this data, and the trade ally's savings from the program tracking database, the following equation is used to calculate the savings of high efficiency equipment that did not receive an incentive:

$$\textit{Savings of Non-Incented High Efficiency Equipment} = \frac{\textit{Savings from Program Database}}{\textit{\% of TA's High Efficiency Equipment that Received Incentive}} - \textit{Savings from Program Database} * \textit{Size Adjustment}$$

The last term in the above equation is a size adjustment that accounts for the possibility that savings from non-incented projects/measures might be different from incented ones. Information on the relative size of incented versus non-incented projects/measures is also collected in the survey.

Using this approach, spillover savings are considered to be equal to the savings of non-incented, high efficiency equipment/products, as calculated in the equation above. To compute the program spillover percentage for active trade allies, the following steps are used:

1. **Develop the spillover ratio for sampled trade allies** by summing their spillover savings and dividing this total by the program-tracked savings associated with the sampled trade allies.
2. **Develop spillover savings for the population of active trade allies** by applying the spillover ratio from Step 1 to all program savings associated with a trade ally (whether a survey respondent or not).
3. **Develop the overall spillover ratio for active trade allies** by dividing the trade ally spillover estimate from Step 2 by total program savings (whether associated with a trade ally or not).

5.2.1.2 Midstream Programs

Similar to downstream programs, surveys can be used to ask active trade allies in midstream programs if the program influenced their sales of high-efficiency equipment to participating or nonparticipating customers and to quantify the program's impact on their high-efficiency sales. To assess if a sampled midstream trade ally created spillover, the following screening criteria are recommended (the order of these may be adjusted by the evaluator):

1. The percentage of the trade ally's sales that are high efficiency and/or the total volume of high efficiency sales increased since the trade ally became exposed to the program.
2. The trade ally sold at least some high efficiency equipment or products during the evaluation period that did not receive an incentive.
3. The trade ally's recommendation, marketing, or equipment/product stocking or placement was influential in the customers' choice of high efficiency equipment/product over standard efficiency equipment/product in instances where the equipment did not receive a program incentive.

Sampled trade allies who do not pass one of the above screening criteria do not qualify for spillover and may be skipped out of the rest of the spillover module.

To quantify spillover for each sampled midstream trade ally, the survey collects information on the percentage of the trade ally's total equipment sales (in terms of projects or measures) that was (1) standard efficiency, (2) high efficiency that DID receive a program incentive, and (3) high efficiency that DID NOT receive a program incentive. Based on these responses, the share of a trade ally's high efficiency sales that received an incentive can be calculated as follows:

$$\textit{\% of TA's High Efficiency Sales that Received Incentive} = \frac{\textit{\% High efficiency that DID receive a program incentive}}{\textit{\% High efficiency that DID receive a program incentive} + \textit{\% High efficiency that did NOT receive a program incentive}}$$

Through additional survey questions,[53] the evaluator should develop an attribution percentage, i.e., the proportion of non-incented high efficiency projects or measures that are attributable to the program. With this data, and the trade ally's savings from the program tracking database, the following equation is used to calculate the trade ally's spillover savings:

$$\textit{Spillover Savings} = \frac{\textit{Savings from Program Database}}{\textit{\% of TA's High Efficiency Sales that Received Incentive}} * \textit{Savings from Program Database} * \textit{Attribution \%} * \textit{Size Adjustment (if applicable)}$$

The last term in the above equation is a size adjustment that accounts for the possibility that savings from non-incented projects/measures might be different from incented ones. Information on the relative size of average energy savings of incented versus non-incented projects/measures is also collected in the survey if the evaluator expects a potential difference in relative size.

To compute the program spillover percentage for active midstream trade allies, the following steps are used:

1. **Develop the spillover ratio for sampled trade allies** by summing their spillover savings and dividing this total by the program-tracked savings associated with the sampled trade allies.
2. **Develop spillover savings for the population of active trade allies** by applying the spillover ratio from Step 1 to all program savings associated with a trade ally (whether a survey respondent or not).
3. **Develop the overall spillover ratio for active trade allies** by dividing the trade ally spillover estimate from Step 2 by total program savings (whether associated with a trade ally or not).

5.2.2 Spillover from Inactive Trade Allies

Inactive trade allies may create spillover if they are exposed to the program but do not directly facilitate program participation, i.e., they did not complete any projects through the program during the evaluation period. Rather, they promote and stock higher-efficiency equipment due to the influence of the program on the market.

Surveys can be used to ask inactive trade allies if the program influenced their sales of high-efficiency equipment to participating or nonparticipating customers and to quantify the program's impact on their high-efficiency sales. The general questions take the following form:

- Q.1: How many <measures> did you sell in <utility>'s service territory in <period>?
- Q.2: How many of them were <efficiency level> or higher?
- Q.3: Had the <program> not existed, how many <measures> of <efficiency level> or higher do you think you would have sold in <utility>'s service territory?

Evaluators should attempt to allow trade allies sufficient time to collect specific data (e.g., by sending information ahead of the interview or conducting additional follow-up; this might require providing incentives as inactive trade

[53] As some trade allies may find it difficult to directly quantify the program's attribution effect on non-program sales, the evaluator may need to use a series of questions to guide the trade ally to provide an estimate of the overall attribution. Questions may include asking about what factors influence sales of non-program efficient equipment/products and how the program influences individual factors to provide context for an overall attribution estimate.

allies tend to be hard-to-reach) and not rely on "guesses" to respond. Additional questions should be included to document how the program influenced sales of additional energy efficient measures and why these measures did not receive an incentive.

For programs that offer a number of different measures, the evaluator should select and ask about a small number of measures or measure groups that are most likely to generate spillover, e.g., the program's highest impact measures. The selection of trade allies to include in this research will depend on the measures selected, e.g., if the highest impact measures are lighting measures, the population of trade allies from which to sample should be lighting contractors.

The following steps are used to calculate the spillover percentage for inactive trade allies:

1. **Develop the total number of spillover units for each trade ally** by computing the difference between the total reported number of high-efficiency units sold and the number that would have been sold in the program's absence, for each measure type.
2. **Develop the total spillover savings for each trade ally** by multiplying the trade ally's total number of spillover units (from Step 1) by the average gross unit savings, for each measure type.
3. **Compute the total spillover savings for the program period** by summing the spillover savings from all sampled trade allies (from Step 2) and multiplying this sum by the ratio of the population of inactive trade allies to the sample, for each end-use.
4. **Compute the program spillover percentage** by summing the spillover savings for all end-uses (from Step 3) and dividing this sum by program gross savings.

It should be noted that the methodology for inactive trade allies requires the evaluator to quantify the number of trade allies in the population. Depending on which types of inactive trade allies are targeted by the research, determining the size of the population may be challenging and may lead to uncertainty in the results. When targeting trade allies that are on the utility's trade ally list (but are not active) or those who have been active in the past but have dropped out, program records allow for accurate estimation of the population size. However, when targeting trade allies that have never been active in the program and were never on the utility's trade ally list, secondary market data is required to develop estimates of population size. The evaluator should carefully document the target population for any inactive trade ally research, data sources used to quantify the population size, and any uncertainty associated with their estimates.

5.3 Consumption Data Analysis Protocol

This protocol refers to impact analyses that use consumption data from customer's monthly bills (commonly referred to as billing analysis) or AMI meter reads[54] to estimate program energy savings. This protocol discusses different consumption data methods and where they fall on the NTG spectrum with respect to participant spillover, nonparticipant spillover, and free ridership; this has implications for whether a NTGR needs to be applied after the consumption data analysis estimate is obtained in order to achieve an estimate of net savings. Decisions of whether to apply a NTGR after conducting a consumption data analysis should be made by the evaluator on a case-by-case basis taking into account the guidelines of this protocol for when these methods are net, gross, or somewhere in between.[55] The remainder of this section discusses NTG for various consumption data analysis methods and then goes through some details of the various analysis methods.

In general, consumption data analysis methods split into two approaches. One approach is to use a comparison group in a randomized control treatment (RCT) design, a random encouragement design (RED) or a quasi-experimental design. These comparison group approaches can, under the right circumstances, be used to directly

[54] Benefits of using AMI data can include: having more observations per customer, which may improve model precision; obviating concerns over billing periods with differing numbers of days; and, for hourly models, providing the ability to observe intraday load shifting in addition to energy savings.

[55] For example, it is generally accepted that programs for income qualified customers have little to no free ridership as these customers are unlikely to install the measures without the incentive of the program. For specific guidance on income qualified programs see Section 4.

estimate net savings eliminating the need for a NTGR adjustment. A second approach is to estimate savings without a comparison group (for example, using a pre/post regression model for program participants). Approaches without a comparison group produce gross savings and must be adjusted by a NTGR to achieve net savings.

In consumption data analysis, energy consumption of the treatment and control groups can be appropriately compared through a regression analysis, using time-series observations on the usage of individual customers in the treatment and comparison groups during the pre- and post-treatment periods. Due to the combined time-series/cross-section structure of such data sets, panel regression techniques can be used.[56]

In general, consumption data analysis methods are best suited to the following situations:

1. When the expected net savings per participant (i.e., the effect size) are large or when large participant/nonparticipant sample sizes are possible.
2. When the program can be designed using a randomized controlled trial (see Section 5.3.2).
3. Programs where nonparticipant spillover is expected to be trivial within the comparison group.
4. Cases where self-selection bias can be effectively controlled for.

5.3.1 Consumption Data Analysis and NTG

Different consumption data analysis methods produce different savings estimates in terms of the NTG spectrum, as summarized in Table 5 – 3. These methods will always yield gross savings with respect to nonparticipant spillover and net savings with respect to participant spillover. However, the savings estimates may be net, gross, or somewhere in between with respect to free ridership, depending on the evaluation technique.

Table 5-3. NTG Summary for Consumption Data Analysis

Consumption Data Analysis Method	Free Ridership	Participant Spillover*	Nonparticipant Spillover**
Randomized Controlled Trial (RCT)	✓	✓	§
Random Encouragement Design			
No Instrumental Variable (IV)	†	✓	§
IV	†	✓	§
IV w/ Inverse Mills Ratio (IMR)	†***	✓	§
Quasi-Experimental Design (QED) ****			
Matching			
To Nonparticipants	†*****	✓	§
To Prior or Future Participants	§	✓	§
Regression Discontinuity (RD)	✓	✓	§
Variation-in-Adoption (VIA)	§	✓	§
Without a Comparison Group	§	✓	§
§ Indicates not accounted for (gross) ✓ Indicates fully accounted for (net)			

[56] "Panel" refers to the data set consisting of time-series observations on energy consumption of a cross-section of treatment and control customers. Panel estimation techniques refer to the model's inclusion of terms that control for individual customer heterogeneity (e.g., customer fixed effects or a lagged dependent variable), and cluster-robust standard errors, which can accommodate differing error variances across customers and an intracustomer correlation of errors.

† Indicates partially accounted for (between net and gross)
* Participant spillover within the analysis timeframe in the same building and fuel type is captured. Other sources of participant spillover may not be captured. See the subsection on participant spillover below for details.
** Nonparticipant spillover is not captured as a positive in consumption data analysis and may actually reduce the estimate of savings if it occurs within the comparison group. See the subsection on nonparticipant spillover below for details.
***This method has been tested in simulation but needs further use in practice.
**** Note that this is a non-exhaustive list of QED evaluation techniques.
***** As noted in first few paragraphs of Section 5.3, these comparison group approaches can, under the right circumstances, be used to produce an of estimate net savings, eliminating the need for a NTGR adjustment (see Goldberg et al., 2017).

When consumption data analysis methods are being used to update the TRM, the update should explicitly state how a NTGR should be applied to the given measure or program in the future. The language used should consider different program delivery mechanisms (which often have different NTG values) and how stable the NTG value is likely to be over time (thus allowing for consideration of how frequently it should be updated).

5.3.1 Nonparticipant Spillover

Nonparticipant spillover is never captured by consumption data analysis, making these savings estimates gross with respect to nonparticipant spillover (i.e., nonparticipant spillover is not accounted for by the estimate directly from the consumption data analysis without further adjustment). To the extent that nonparticipant spillover occurs in the comparison group being used for evaluation, the effect of the program may be underestimated as the difference between the participant group and the comparison group is decreased by the amount of nonparticipant spillover. If nonparticipant spillover is expected to be large (based on the best research available or given the program's logic model) and occur within the evaluation comparison group, that may be a reason to use other methods for evaluating savings. If a billing analysis is done in these cases, a traditional nonparticipant spillover analysis (using techniques like nonparticipant surveys or interviews) should be used to help quantify this effect (these analyses are discussed in various subsections of Chapter 4 of this protocol). Within the comparison group, it can also be difficult to distinguish the effects of nonparticipant spillover, free ridership, and market transformation as all of these effects increase uptake of a measure without going through the program among the nonparticipant group.

In cases where nonparticipant spillover is not expected to occur in the comparison group but may occur in the broader population (for example, if we go from a pilot evaluation where measures were restricted among the comparison group to a full program deployment), adjustments for nonparticipant spillover (or justification for why there is no nonparticipant spillover) should be made as appropriate on a program-by-program basis.

5.3.2 Participant Spillover

Participant spillover is captured by consumption data analysis, making these savings estimates net with respect to participant spillover (i.e., participant spillover is accounted for by the estimate directly from the consumption data analysis without further adjustment). This occurs because consumption data analysis measures all changes in participant usage (captured by the utility billing system or AMI meter reads) regardless of whether the changes are related to the program. A few caveats apply:

1. Consumption data analysis does not capture participant spillover that occurs outside the home or business being analyzed. For example, spillover at a participant's vacation home or spillover at other facilities owned by the same firm.
2. Consumption data analysis does not capture participant spillover that occurs in a different fuel type. For example, if the analysis is done on electric data but there is participant spillover into natural gas.
3. Consumption data analysis does not capture participant spillover that occurs outside the analysis period (typically a one-year period).

If these sources of participant spillover that are not captured are expected to be large (based on the best research available or given the program's logic model), adjustments or additional analysis to capture these types of participant spillover may be required.

5.3.3 Free Ridership

With respect to free ridership, consumption data analysis can produce savings estimates that are net, gross, or somewhere in between (i.e., free ridership can be fully, not at all, or partially accounted for by the estimate directly from the consumption data analysis without further adjustment). Where they fall depends on whether the comparison group accounts for (or nets outs) free ridership in the estimation. For a summary of where each method falls see Table 5-1, above.

Methods that yield gross savings estimates with respect to free ridership have no comparison group or have a comparison group that is made up of other (prior or future) participants. In these cases, a free ridership adjustment (or justification of why there is no free ridership) is necessary. These methods include:

- Matching to older or newer participants[57]
- Variation-in-adoption (VIA)[58]
- Any method without a comparison group

Methods that yield net savings estimates with respect to free ridership have a nonparticipant comparison group that has the same level of free ridership as the participants. In these cases, the comparison group is engaging in energy efficiency activities at the same rate as the participant group would have without the program. This nets out the free ridership and means no free ridership adjustment is necessary. These methods include:

- Randomized controlled trial (RCT)
- Regression discontinuity (RD)
- Random encouragement design (RED) under at least one of the following conditions:
  - Analysis is done using instrumental variables with an inverse mills ratio[59]
  - Designs where only the encouraged group can join the program (and as such the participants who join the program include only compliers and not always takers[60])
  - There is no relationship between how much energy a customer will save by participating and their inclination to participate

Methods where there is a nonparticipant comparison group that is expected to have a different level of naturally occurring adoption than the participant group can result in savings estimates that fall somewhere between net and gross with respect to free ridership. For example, a group of participants would be expected to be comprised of more natural adopters than a group of nonparticipants who never joined the program. These methods include:

- RED (in situations not covered by the previous list showing when RED is net)
- Matching to nonparticipants

In these cases, it is up to the evaluator to decide whether an estimate is most appropriately considered net or gross on an analysis-by-analysis basis. Some guidelines include:

- Measures where instant upstream rebates exist for a large portion of the market are likely gross as there should be very few customers who got the measure in the nonparticipant group
- Measures for income qualified customers are typically considered net as these customers are unlikely to

[57] Except in the case of income qualified programs where the use of future participants can produce an estimate of net savings. For specific guidance on income qualified programs see Section 4.

[58] See Harding and Hsiaw (2013). This is a distinct method from the UMP Chapter 8 (Agnew and Goldberg, 2017) pooled fixed effects approach which can be estimated with multiple years of participants. VIA hinges on rolling enrollment and in essence uses each participant as a control and a treatment customer through time. The Chapter 8 pooled fixed effects approach uses participants from an earlier time period as a comparison group for participants from a later time period.

[59] For details see: Goldberg, M.; Agnew, K.; Train, K.; Fowlie, M. (2017). *Mitigating Self-Selection Bias in Billing Analysis for Impact Evaluation.* Pacific Gas and Electric Company. CALMAC Study ID PGE0401.01. <http://www.calmac.org/publications/Mitigating_Self_Selection_Bias_in_Bill_Analysis_8.4.17.pdf>

[60] See Section 5.3.2.

install the measures without the incentive of the program

In some cases, evaluators may be able to implement techniques when using a nonparticipant comparison group such that the savings are sufficiently close to net and do not require further net to gross adjustment. One example of these techniques is the IV-IMR method proposed in Goldberg et. al. (2017). The UMP Chapter 21 (Violette and Rathbun, 2017) also has some discussion of getting net savings estimates using these approaches, although UMP Chapter 8 (Agnew and Goldberg 2017) should be reviewed in conjunction as it is more specific to consumption data methods. However, these techniques often require customer characteristic data that is not readily available to evaluators and some of them needed to be further tested beyond theoretical simulations.

5.3.4 Consumption Data Analysis Designs with a Comparison Group

This section discusses descriptions of and considerations for estimating savings via consumption data analysis designs with a comparison group. Although the ideas of net and gross savings are touched upon, the full discussion on whether each of these methods produce net or gross savings and under what circumstances is in Section 5.3.1.

5.3.5 Randomized Controlled Trials

In a randomized controlled trial (RCT) design, evaluators (and sometimes implementation contractors) randomly assign sampled members of a population of interest to a treatment group or a control group. Among the benefits offered by an RCT—when properly applied—is that it produces net savings estimates by netting out free ridership.[61] The evaluation of a program must be designed and implemented this way from the outset; it is not possible for an evaluation team to apply RCT evaluation techniques after the program has been implemented if random assignment to treatment and control groups was not done before program launch. While such designs are rarely possible outside of Home Energy Report programs, one should not overlook the possibility of such designs in evaluating new pilot programs.

For some programs, evaluators must take a second step to ensure savings are not being double-counted, either counting savings being claimed by other programs or savings already credited to earlier program efforts (often called “legacy uplift”). Only net increases in participation in other programs should be considered in this uplift adjustment; changes to total savings do not need to be made based on decreases in participation in other programs.

5.3.6 Random Encouragement Designs

In a random encouragement design (RED), eligible customers are randomly assigned between an encouraged group (who receives incremental encouragement to join the program[62]) and a non-encouraged, or control, group (who does not receive the encouragement). Members of either group can join the program, but the encouraged group is expected to do so at a higher rate.[63] If the encouragement is not effective at driving the encouraged group into the program at a higher rate than the non-encouraged group then the evaluation design breaks down and other (likely quasi-experimental) methods will be needed to estimate program savings.

In an RED, both the encouraged and non-encouraged group are made up of the following:

1. Always takers – customers who will join the program with or without the encouragement
2. Compliers – customers who only join the program if they receive the encouragement
3. Never takers – customers who will never join the program, regardless of whether they receive the

[61] RCTs eliminate free rider bias because the random assignment of customers to treatment and control groups equally distributes such participants between the two. Due to differential attrition and random chance, small differences may occur between the distributions of free riders in the two groups for any given sample. Their expected values, however, will be identical, and in any case the size of any such discrepancies shrinks as sample size increases. Thus, this is only a potential concern for programs with unusually small numbers of participants.) Upon comparing the two groups’ energy consumption, free riders’ energy savings in the control group cancel out those in the treatment group, eliminating free rider bias.

[62] The encouragement could take many forms including targeted marketing or direct monetary incentives.

[63] This design does not preclude mass marketing of the program to all customers but relies on the encouragement being effective at driving the encouraged customers into the program at a higher rate than the non-encouraged customers.

encouragement

In the non-encouraged group, the always takers can be distinguished from the compliers and never takers (they're the portion of the non-encouraged group who joins the program), but the compliers and never takers cannot be distinguished from one another (they're both observed not to join the program). In the encouraged group, the never takers can be distinguished from the always takers and compliers (they're the portion of the encouraged group who does not join the program), but the always takers and compliers cannot be distinguished from one another (they're both observed to join the program).

Like RCTs, REDs are a form of experimental design. An RED is known to give an unbiased estimate of net savings (with respect to free ridership) for the compliers. Applying this savings to the always takers group requires some explanation of why it is likely to be accurate. Additionally, the RED design provides the average net savings per participant for those who participate because of the encouragement but otherwise would not (compliers). This is not necessarily the same as the net savings for the original program without extra encouragement. In particular, we would expect free-ridership to be lower among those who need extra encouragement. Thus, the RED might be expected to overstate net savings for the original program if free-ridership is present but would still provide useful information.

There are several methods for evaluating REDs using panel data including methods using instrumental variables (IVs) and the inverse mills ratio (IMR).[64]

5.3.7 Quasi-Experimental Designs

Where randomized assignments prove infeasible, quasi-experimental design (QED) evaluation methods can be substituted (although experimental designs are typically preferable when possible). Depending on the exact QED implemented, the savings may be net, gross, or somewhere in between with respect to the different pieces of a NTG adjustment (participant spillover, nonparticipant spillover, and free ridership). The specifics of net versus gross estimation are covered in Section 5.3.1, this subsection does not rehash this issue but rather describes estimation for a subset of QED methods.

Three quasi-experimental approaches are commonly used to evaluate behavior-based energy efficiency programs that cannot be constructed as experiments:[65]

- Regression discontinuity (RD)
- Variation-in-adoption (VIA)[66]
- Matched controls (MC)

All three rely on a nonrandom comparison group.

Regression Discontinuity. RD requires basing a program's eligibility on a continuous variable (e.g., customers'

[64] See, for example:

Goldberg, M.; Agnew, K.; Train, K.; Fowlie, M. (2017). *Mitigating Self-Selection Bias in Billing Analysis for Impact Evaluation.* Pacific Gas and Electric Company. CALMAC Study ID PGE0401.01. <http://www.calmac.org/publications/Mitigating_Self_Selection_Bias_in_Bill_Analysis_8.4.17.pdf>

Fowlie, M.; Greenstone, M.; Wolfram, C. (2015). *Are the Non-Monetary Costs of Energy Efficiency Investments Large? Understanding Low Take-up of a Free Energy Efficiency Program*. American Economic Review: Papers and Proceedings 105(5): 201-204. < https://www.povertyactionlab.org/sites/default/files/publications/389_500%20Weatherization%20AER.pdf>

[65] There are many other types of QEDs that may be appropriate for evaluation but these are some of the most commonly used for evaluation in IL.

[66] See Harding and Hsiaw (2013). This is a distinct method from the UMP Chapter 8 (Agnew and Goldberg, 2017) pooled fixed effects approach which can be estimated with multiple years of participants. VIA hinges on rolling enrollment and in essence uses each participant as a control and a treatment customer through time. The Chapter 8 pooled fixed effects approach using participants from an earlier time period as a comparison group for participants from a later time period. The Chapter 8 pooled fixed effects method is discussed in Section 5.3.3.

adjusted gross income falling below a cutoff value for them to qualify for the program). When this is true, the RD method assumes customers just beyond the cutoff likely will be very similar, on average, to those just inside of it. The method compares changes in energy usage for a group just outside of the eligible range to that of a group of participants just on the other side of the eligibility cutoff. The RD approach, however, is susceptible to an important weakness: misspecification of the regression functional form.[67]

Variation-in-Adoption. The VIA model applies only to program participants.[68] For this method, customers must sign up for the program on a rolling basis. VIA takes advantage of its enrollees' differential timing to compare energy usage of customers opting in to that of customers not yet opting in (but doing so later). The method relies on an assumption that, in any given month, customers that soon opt in have similar characteristics to those who have enrolled, both in observable and unobservable characteristics. For this assumption to prove valid, customers must decide to opt into the program at different times for essentially random reasons (e.g., influenced only by marketing exposure and program awareness).[69] In particular, the decision to opt in should not relate to observable or unobservable household characteristics.[70]

Matched Controls. MC creates a control group by matching each treatment customer to the most similar nonparticipant customer available on the basis of exogenous covariates from the pre-enrollment period known to highly correlate with post-enrollment usage.[71] The covariate most likely to correlate with post-enrollment energy usage in a given time period is customer energy usage during the same period of the preceding year, but other observable factors may be used when available. Implementing MC requires customer usage data for the year preceding all opt-in customers' decisions to participate in the program, along with a large group of nonparticipants who can be assumed to be similar to opt-in customers, aside from their program participation status. Whenever possible, the pool of potential matches should be drawn from the same geography, customer class, and rate category as the participants.

Another option is to pull the nonparticipants from a group of prior or future participants in the program (sometimes referred to as the cohort design[72]). These groups are similar to current participants since we know that they also join the program at an earlier or future date, significantly mitigating the issue of self-selection bias (wherein, customers who join the program are different from those who do not in unobservable ways).[73] However, using this design can significantly decrease the number of participants for analysis and the size of the potential matching group. It can also require the evaluator to delay the analysis if more recent participants are being used as the comparison group.[74]

---

[67] The most common misspecifications are: mistaking a nonlinear relationship for a discontinuity; and failing to recognize potential interactions between assignments and the treatment studied. See W.R. Shadish, T.D. Cook and D.T. Campbell, *Experimental and Quasi-Experimental Designs for Generalized Causal Inference*, Wadsworth 2002, pp. 229-238.

[68] Harding, M. and Hsiaw, A. 2013. *Goal Setting and Energy Conservation* Available at: http://people.duke.edu/~mch55/resources/Harding_Goals.pdf.

[69] This differs from an RCT with a recruit-and-delay design, in which customers do not choose when to opt in, but instead are randomly assigned different times to opt in, and from an RCT with a recruit-and-deny design, where customers are randomly denied access to the program.

[70] As the validity of the VIA method depends on this assumption, it should be empirically tested to the extent possible. If program marketing is punctuated and dates of marketing exposure are known, it is possible to test whether household enrollment in any particular month is driven by marketing activity, as opposed to observed household characteristics or unobserved heterogeneity. A test of whether the energy usage of households before they opt in differs from households that opt in during any particular month as opposed to another month is built into the VIA regression model's functional form. See Harding and Hsiaw, op. cit., for details.

[71] See Daniel E. Ho, Kosuke Imai, Gary King, and Elizabeth Stuart, 2007, "Matching as Nonparametric Preprocessing for Reducing Model Dependence in Parametric Causal Inference." *Political Analysis* 15(3): 199-236.

[72] See W.R. Shadish, T.D. Cook and D.T. Campbell. (202). *Experimental and Quasi-Experimental Designs for Generalized Causal Inference*. New York: Houghton Mifflin Company, pp. 148-153

[73] Though there could still be a selection issue based on when customers choose to join the program. As with VIA, the assumption is that the timing of participation is basically random.

[74] The cohort design has also been used, under certain conditions, to control for exogenous factors when estimating gross

The MC method involves identifying a nonparticipant customer whose energy usage closely matches that of a program participant in the months preceding the participant's enrollment in the program. The logic inherent in this approach is: if the analyst finds a set of nonparticipants who, on average, are the same as participants regarding energy consumption before program enrollment, these matches will provide a good counterfactual estimate of how much energy participants would have used in the program's absence.

The MC approach does present a main weakness: it can only identify matches based on observable customer characteristics, which leaves open the exclusion of the possible influence of relevant unobservable variables. While factors other than pre-enrollment energy usage plausibly could be used (e.g., household income, demographics, geographic location) in the matching process to address relevant unobservable characteristics (e.g., attitudes toward energy conservation and environmental concerns), this assumption cannot be directly tested.[75]

There is a special case of MC called propensity-score matching. This develops a binary choice (logistic regression) model to predict the probability that a customer will opt into the program, and then, for a comparison group. The logistic regression reduces each household's set of covariates to a single propensity score. Nonparticipants are then matched to participants based on their propensity scores. This functions well if observable variables used to calculate the propensity score sufficiently correlate with relevant unobservable variables to explain differences between treatment and control customers that cannot be explained by matching on observable variables. With most evaluations of energy efficiency programs, however, little (if any) data are available on nonparticipating customers other than their energy usage. In some cases, the demographic data necessary to estimate these models can be obtained from providers such as Experian and assigned to each participant and nonparticipant.

Self-Selection Bias and QED. Self-selection bias due to observable and unobservable variables is always a possibility with QEDs. One can collect as much information as possible on both participants and members of the comparison group and include them as covariates in the regression model, but there may still be self-selection bias related to unobservable variables. Several techniques have been developed to help mitigate it. Efforts to address the biasing effects of *unobserved* differences using Inverse Mills Ratios began at least as early as the late 1980s. Since then, Train (1993) and Goldberg and Train (1995), using simulated datasets, demonstrated that failing to correct for self-selection can overestimate net savings, but that there are effective strategies to reduce this bias substantially.

One approach is to calculate and enter the propensity score, based on observable variables, as an additional covariate into the regression model. Of course, the most difficult issue to address is the differences between participants and nonparticipants that are unobserved and unobservable. To mitigate both overt and hidden bias, a variety of approaches that attempt to take advantage of recent developments in statistics and econometrics are available:

- Sample selection models (e.g., Heckman's two-step estimator (1978, 1979); treatment effect model (Green, 2003); instrumental variables estimator (Wooldridge, 2002)
- The propensity score matching model (Rosenbaum and Rubin, 1983, 1985; Hansen and Klopfer, 2006; Guo and Fraser, 2014)[76]
- Matching estimators and synthetic controls (Abadie and Imbens, 2002, 2006)
- Instrumental variables approach with the predicted probability of participation serving as the instrumental variable and the inclusion of an Inverse Mills Ratio (IMR) (Goldberg et al., 2017)

Another issue that should be considered is that, when using a comparison group in a QED, the composition of the

savings. See Agnew, K. and M. Goldberg. (2017). Whole Building Retrofit with Consumption Data Analysis Evaluation Protocol: Chapter 8 of the Uniform Methods Project, National Renewable Energy Laboratory.

[75] Such secondary, observable characteristics are rarely available to evaluators of energy efficiency programs, except for geographic location (e.g., postal zone of customer premise).

[76] Note that propensity scores cannot remove hidden biases except to the extent that unmeasured variables are correlated with the measured covariates used to compute the propensity score

comparison group needs to be carefully considered.[77,78] For example, simply selecting a random sample of nonparticipants from the general nonparticipant population could result in an estimate of savings that is somewhere between net and gross, thus overestimating net savings. For a single-measure residential program like an air conditioner (AC) replacement program, the eligible population is the population of customers who have purchased a new air conditioner. That is, part of the eligible population appropriate to a net effects comparison group would be those who purchased and installed some air conditioner, whether efficient or not. Simply selecting from the general residential population would include households with no air conditioner, those with older ACs of varying vintages, those with new standard efficient ACs and those with new program-qualified ACs. The results would be virtually uninterpretable. Of course, for more complex multi-measure programs, finding the appropriate comparison group is far more challenging.

5.3.8 Consumption Data Analysis Designs without a Comparison Group

Although less common, consumption data methods can also be used to estimate savings without the use of a comparison group. These methods typically estimate gross savings, and net savings are found by multiplying gross savings by a separately estimated NTGR. There are basically two types of pre/post models to estimate gross savings:

- the pooled participant-only linear fixed-effects approach
- site-specific regression models

In both modeling approaches, exogenous factors must be controlled for.[79]

Pooled Approach. The pooled approach addresses exogenous change without the inclusion of a separate comparison group. In this model, participants who received a measure installation during a certain time interval serve as a steady-state comparison for other participants in each other time interval. Almost all observation points include premises that are still in their pre-installation period and premises that are in their post-installation period, so the effect of post- versus pre- is estimated to control for exogenous trends. Note that if changes at the site that affect energy use are not or cannot be explicitly modelled the estimated gross savings will be biased. This method is typically used in analysis of residential and small (and occasionally for large) commercial programs.

Site Specific Regression Models. This approach involves the estimation of site-specific regression models to estimate savings. This method is often used for large commercial and industrial customers or in other situations where it is difficult to identify an adequate comparison group (for example, in evaluation of Strategic Energy Management programs). In these cases, single customer regressions are typically run as a time series without a cross-section of customers.

Note that both the pooled approach and the site-specific approach and the conditions that must be met before using them are discussed in Agnew and Goldberg (2017).

5.3.9 Program Implementation and Consumption Data Analysis

The approach the evaluation can use to estimate net savings is greatly dependent on the design of the program and the size of the expected savings (i.e., the signal-to-noise ratio).

RCT and RED: These designs must be integral to a program's implementation. Without the ability to randomly assign customers to the control and treatment groups (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. However, if they not involved at the outset, they cannot carefully review choices made by the implementation team. RCT and RED designs are difficult to perform well within the commercial and industrial sectors due to a low signal-to-noise ratio. One solution for these two sectors is to increase the sample size but this is not always feasible.

[77] See Agnew at al., Section 8.1.3 (The Importance of Measures Applicability)

[78] Katherine Randazzo, Richard Ridge and Seth Wayland. Evaluating Whole-Building Programs: It is harder and easier than you think! Presented at the International Energy Program Evaluation Conference in August 2017.

[79] Exogenous factors include non-program-related effects due to the economy and other factors affecting energy consumption.

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 and other variables, if available. QED is also difficult to perform well within the commercial and industrial sectors due to a low signal-to-noise ratio. One solution for these two sectors is to increase the sample size but this is not always feasible.[80]

Methods without a Comparison Group: These methods can also be implemented by the evaluator after the program has been designed. They are most appropriate in situations where it is difficult to construct an appropriate comparison group.

For any kind of evaluation design, evaluators may also analyze the data to help understand the savings within specific segments if sufficient information and data points are available.

References

- Abadie, Imbens, 2002, 2006
- Agnew, Goldberg, 2017
- Donaldson, 2009
- Fowlie, Greenstone, Wolfram, 2015
- Goldberg, Agnew, Train, Fowlie, 2017
- Goldberg, Train, 1995
- Green, 2003
- Guo, Fraser, 2014
- Hansen, Klopfer, 2006
- Harding, Hsiaw, 2013
- Heckman, 1978, 1979
- Ho, Imai, King, Stuart, 2007
- Mohr, 1995
- Shadish, Cook, Campbell, 2002
- Scriven, 2008
- Randazzo, Ridge, Wayland, 2017
- Rosenbaum, Rubin, 1983, 1985
- Train, 1993
- Wooldridge, 2002

5.4 Code Compliance Protocol

The protocol represents a basic framework for estimating the NTGR that may be refined based on impact evaluation results. The NTGR is used to convert an estimate of gross savings into an estimate of net savings. Two general methods can be used to estimate gross energy impacts: (1) utility billing data analysis; and (2) building energy modeling.[81] The specific method used depends on the availability of necessary data.

5.4.1 Data Collection

5.4.1.1 Program Documentation

To inform the NTGR estimate, the evaluator documents program delivery. Information collected includes the

[80] A power analysis can be undertaken before the actual analysis to determine whether the sample size available is likely to be large enough to produce statistically significant savings at the desired confidence level.

[81] The modeled energy savings approach is similar to the approach described by Department of Commerce in Exhibits 6.1 and 6.2 from excerpts of Docket 13-0499 through estimation of potential energy savings.

following: the number, location, and dates of training workshops; the topics covered; materials disseminated; the number of trainees in each workshop and the type of trainee; and the hours of instruction.

5.4.1.2 Stakeholder Interviews

To inform the NTGR estimate, the evaluator conducts interviews with key stakeholders involved in the program. Interviews should include training program managers, instructors, and trainees. Trainees typically include contractors, builders, consultants, code officials, and others involved in building design and construction. The interviews seek to gather information on how training affected building design, construction, new code compliance, and enforcement.

5.4.2 Attribution Assessment

The NTGR estimation method stays the same, regardless of the method used to estimate gross energy savings.

A Delphi panel[82] produces an NTGR estimate that reflects the share of gross energy savings resulting from increased code compliance attributable to the program. Formed by selecting four to six knowledgeable professionals not associated with the program in any way,[83] the panel receives estimates of gross energy savings, building construction data, and evidence of attribution—including the results of stakeholder interviews and program documentation. Panel members individually review the information and provide feedback regarding their NTGR estimates and rationales. Responses are compiled, with combined, anonymous responses circulated to all panel members. Panelists review this information, revise their initial estimates and rationales, as they deem appropriate, and provide new estimates and rationales. Evaluators review the second set of estimates and rationales to develop a final attribution estimate, accompanied with a summary of supporting rationales. This NTGR estimate, used in combination with the gross energy savings estimate and building construction data, produces a final estimate of net energy savings attributable to the program.

[82] The Delphi panel should be conducted according to best practices. For example, see: Day J and Bobeva M (2005) "A Generic Toolkit for the Successful Management of Delphi Studies" *The Electronic Journal of Business Research Methodology* Volume 3 Issue 2, pp. 103-116, available online at www.ejbrm.com.

[83] Delphi panelists should have no biases that would affect their assessment of the program's effectiveness. Selected individuals should be knowledgeable about building codes and all factors that could conceivably affect code compliance.

6 Appendix A: Overview of NTG Methods

The evaluation teams present 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 from a single source—the national Uniform Methods Project, Chapter 23: 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 teams recommend 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.

6.3 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 online.[84] 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.

6.3.1 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

---

[84] Historically, evaluators in Illinois have collected the majority of primary data via telephone surveys. As evaluations increasingly leverage online surveys to collect information relevant to attribution, careful attention should be paid to mode effects that are due to interviewer-administered versus self-administered surveys (e.g., scale direction effects). It is recommended that evaluators, where possible, assess the differences between telephone and online survey methods for the purposes of future updates to these protocols.

6.3.2 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.

6.4 NTG with Consumption Data Analysis

As mentioned in Section 5.3, evaluators use randomized control trials (RCTs), random encouragement designs (REDs), and quasi-experimental designs (QEDs) using consumption data (like monthly bills or AMI meter reads) to estimate savings for a variety of programs. RCTs estimate net savings by design but other consumption data analysis methods may be net, gross, or somewhere in between. In some cases, evaluators may be able to use methods that produce estimates that are acceptably close to net without further adjustment, while in other cases a NTGR may need to be developed outside the consumption data analysis and then multiplied by the estimate to produce net savings. Therefore, the NTG adjustment method will differ and needs to be justified by the evaluator on a case by case basis.

6.5 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.

Deemed or stipulated NTG ratios may be based on previous primary data collection, a review of secondary data, or agreed to among stakeholders. In Illinois, deemed or stipulated NTG ratios should reflect best estimates of likely future actual NTG ratios for the relevant program year, taking into consideration stakeholder input, the evaluator’s expertise, and the best and most up-to-date information.

6.6 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 it is used elsewhere in the country, such as the Pacific Northwest and Delaware.

6.7 Market Analyses

Market analyses can be done in several ways. 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) reductions 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 program interventions over the past eight years.

Market analyses can be backward looking through historical tracing, but it is best used when the logic of an intervention is described and specific market metrics are tracked over time.

6.8 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 are collected from two or more rounds of data collection (which can occur via e-mail, 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.

References

- Mosenthal, et al., 2000
- Powell, 2002

6.9 Program Theory-Driven Approach

This approach is not included in the Violette and Rathbun (2014) document as a high-level method, but it 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.[85]

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 20-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 to collect data 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 under Market Analyses.

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.

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

References

- Weiss, 1997
- Chen, 2000
- Coryn, 2011

6.10 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 the replication of findings.

References

- Yin, 2003
- Stake, 2006

7 Appendix B: References

Abadie, A. and G. W. Imbens. (2002). Simple and bias-corrected matching estimators [Technical Report]. Department of Economics, University of California, Berkeley. Retrieved March 6, 2016, from http://ideas.repec.org/p/nbr/nberte/0283.html.

Abadie, A. and G. W. Imbens. (2006). Large sample properties of matching estimators for average treatment effects. Econometrica, 74, pp. 235-267.

Agnew, K. and M. Goldberg. (2017). Whole Building Retrofit with Consumption Data Analysis Evaluation Protocol: Chapter 8 of the Uniform Methods Project, National Renewable Energy Laboratory.

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.

Fowlie, M.; Greenstone, M.; Wolfram, C. (2015). Are the Non-Monetary Costs of Energy Efficiency Investments Large? Understanding Low Take-up of a Free Energy Efficiency Program. American Economic Review: Papers and Proceedings 105(5): 201-204. < https://www.povertyactionlab.org/sites/default/files/publications/389_500%20Weatherization%20AER.pdf>

Greene, W. H. (2003) Econometric Analysis (5th Ed.). Upper Saddle River, N.J.: Prentice Hall.

Goldberg, M. L. And K. Train, (1995) Net Savings Estimation: An Analysis of Regression and Discrete Choice Approaches. Submitted by Xenergy Inc. To CADMAC Subcommittee on Base Efficiency.

Goldberg, M.; Agnew, K.; Train, K.; Fowlie, M. (2017). Mitigating Self-Selection Bias in Billing Analysis for Impact Evaluation. Pacific Gas and Electric Company. CALMAC Study ID PGE0401.01. <http://www.calmac.org/publications/Mitigating_Self_Selection_Bias_in_Bill_Analysis_8.4.17.pdf>

Gou, Shenyang and Mark W. Fraser, (2015). Propensity Score Analysis: Statistical Methods and Applications. Los Angeles, CA: SAGE Publications.

Hansen, B. B. and S. O. Klopfer. (2006). Optimal Full Matching and Related Designs via Network Flows. Journal of Computational and Graphical Statistics, 15, pp. 1-19.

Harding, M. and Hsiaw, A. 2013. Goal Setting and Energy Conservation Available at: http://people.duke.edu/~mch55/resources/Harding_Goals.pdf.

Heckman, J. J. (1979). Sample Selection Bias as a Specification Error. Econometrica, 47(1), pp. 153-162.

Heckman, J. J. (1978). Dummy Endogenous Variables in a Simultaneous equations system. Econometrica, 46(1), pp. 931-960.

Ho, D.E., Imai, K., King, G. and Stuart, E., 2007. "Matching as Nonparametric Preprocessing for Reducing Model Dependence in Parametric Causal Inference." Political Analysis 15(3): 199-236.

Mohr, Lawrence B. 1995. Impact Analysis for Program Evaluation, Second Edition. Thousand Oaks, CA: Sage Publications, Inc.

Mosenthal, Philip, Prahl, 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.

Randazzo, K., Ridge, R. and Wayland, S. Evaluating Whole-Building Programs: It is harder and easier than you think! Presented at the International Energy Program Evaluation Conference in August, 2017.

Rosenbaum, P. R. and D. B. Rubin. (1983). The central role of the propensity score in observational studies for causal effects. Biometrika, 70, pp. 41-55. 21

Rosenbaum, P. R. and D. B. Rubin. (1985). Constructing a control group using multivariate matched sampling methods that incorporate the propensity score. American Statistician, 39, pp. 33-38.

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.

Train, Kenneth. (1993). Estimation of Net Impact from Energy Conservation Programs. Submitted to the Southern California Edison Company.

Violette, Daniel M.; Rathbun, Pamela. (2017). Chapter 21: Estimating Net Savings – Common Practices: Methods for Determining Energy-Efficiency Savings for Specific Measures. Golden, CO; National Renewable Energy Laboratory. NREL/SR-7A40-68578. http://www.nrel.gov/docs/fy17osti/68578.pdf

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

Wooldridge, J. M. (2002). Econometric Analysis of Cross Section and Panel Data. Cambridge: MIT Press.

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

Attachment B: Effective Useful Life for Custom Measure Guidelines

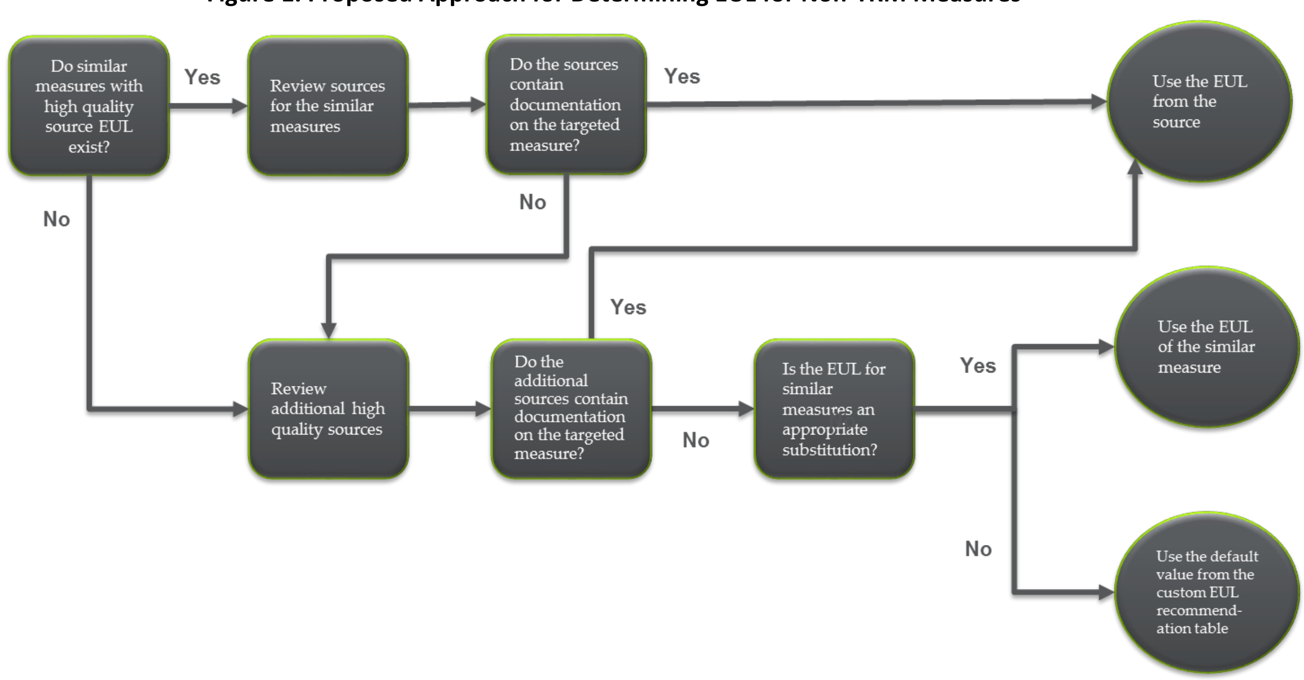
This section provides guidelines on the EUL values to use for the custom measures and programs. The approach for assigning EUL values to non-TRM measures is different from the approach used to assign EUL values to prescriptive measures because the non-TRM EUL (1) may be dependent on a mix of measures, or (2) may not be supported by previous primary and secondary research.

Similar to evaluating custom program savings on a retrospective basis, if there is a defined EUL for a measure or project[86] that does not use a TRM value or the correct TRM value, the evaluator will revise the value accordingly and apply the results in the verified lifetime savings and CPAS. As a result, the implementation team should be consistent and comprehensive in its documentation of the identified EUL.

The complexities of the various approaches for custom-like programs require a program-by-program perspective. The following process should be used to determine the EUL value for custom measures. Figure 1 provides guidance as to what the evaluation team will review and address in providing evaluated CPAS savings. Similar to first year energy savings calculations, appropriate documentation should be provided to support the EUL value which may include references, approach, and reasons.

1. Identify the non-TRM measure and consider if there are similar measures with high quality EUL values already in the TRM. This initial step provides a benchmark for the EUL value.
2. Review the sources used to determine the EUL values for those similar measures. See Table 2.
3. If the sources do not have EUL documentation for the non-TRM measure, research additional sources. The level of research effort should be commensurate with the savings potential for the non-TRM measure.
4. If EUL documentation for the non-TRM measure is insufficient (such as a low-quality source from Table 2), assess if EUL values for similar measures are appropriate substitutes.
5. If none of the above meets the source reference quality criteria, use the recommended default EUL value provided in Table 1.

Figure 1. Proposed Approach for Determining EUL for Non-TRM Measures[87]



[86] A measure is considered one isolated technology that can be defined for energy savings and EUL. A project is made up of a system of technologies such as an HVAC system retrofit where specific measure savings cannot be individually analyzed.
[87] Custom EUL recommendation table is Table 1.

The recommended values in Table 1 are a result of initial research into EUL values for non-TRM measures and may be considered as deemed. The recommended values can be used by program implementers when the steps presented in Figure 1 do not result in sufficient information to determine the appropriate EUL value for a non-TRM measure.

Table 1. Recommended Custom Measure End-Use Categories, Subcategories and Effective Useful Life Values

Program/End-Use Category	End-Use Subcategory	Sample Mapped Measures	EUL (years)	Notes
Combined Heat and Power	Combined Heat and Power	CHP	Capped at 25	Project specific
Compressed Air	Custom Compressed Air – Equipment	Compressed Air Pressure Reduction	15	Default value Future research may show that EULs for compressed air measures vary significantly between equipment and controls.
		Low-Pressure Blower System (replacing compressed air)		
	Custom Compressed Air – Controls	Compressed Air Flow Controller		
	Compressed Air Leak Repair	Compressed Air Leak Repair	1 - 5	A range of possible lifetime values is provided. Therefore, the implementers of this measure must justify the reason for selecting an appropriate measure life for each project and the decision will be subject to evaluation with the risk of adjustments.[88]
Data Centers	Custom Data Centers - Equipment	Data Center	15	Default values Future research may show that EULs for data center measures vary significantly between equipment and controls.
	Custom Data Centers – Controls		15	
Energy Management System	Energy Management System	Energy Management System	15	Default values
HVAC	Custom Electric HVAC – Equipment	Custom Electric HVAC	13	Default values
		VAV Fume Hood		
	Custom Electric HVAC - Controls	Chilled Water Reset	15	
		Fume Hood Occupancy Controls		
		Electric HVAC Controls		

[88] Note during IL TRM v7.0 updates, this assumption was discussed at length with the realization that there is a lack of a strong source for defaulting the lifetime and different applications may vary significantly. It is hoped that future research will help to inform an appropriate assumption(s) to update this assumption for v8.0.

Program/End-Use Category	End-Use Subcategory	Sample Mapped Measures	EUL (years)	Notes
		Low-Flow High Performance Hood -		
		Reduce/Optimize Air Change per Hour (ACH) Rate - Chiller		
		Sash Stops		
Lighting	New Construction/ Custom Lighting	Ceramic MH Lamp	15	Section 4.5.8 of the TRM covers 'Miscellaneous Commercial/Industrial Lighting'. It applies to "energy efficient lighting upgrades that are not captured in other measures within the TRM". The measure applies to retrofits and appears to be applicable to any non-prescriptive lighting measures, which would imply a 15-year measure life for custom lighting measures. It does not cover new construction or controls, thus the recommendation to include these subcategories.
		New Construction Lighting		
	Custom Lighting - Controls	Advanced Lighting Control Systems	8	
Non-Res New Construction	Non-Res New Construction	New Construction – Electric Measures	17.4	Based on research of measure level breakdown of typical projects in a program year.
		New Construction – Gas Measures	20.6	
Refrigeration	Custom Refrigeration	Efficient Refrigeration Condenser	15	Default value Research may show that EULs for refrigeration measures vary significantly between equipment and controls. If that is not the case, the recommended end-use subcategory will continue working well. If that is the case, at that time the end-use subcategories should be updated to the following: Custom Refrigeration – Equipment Custom Refrigeration – Controls
		Floating Head Pressure Controls		
		Refrigerated Cases		
		Refrigeration Compressor		
		Refrigeration Controls		
Res New Construction	Res New Construction	New Construction Electric Measures	18	
	Affordable Housing New Construction	Affordable Housing New Construction	-	Varies by project based on implemented measures

Program/End-Use Category	End-Use Subcategory	Sample Mapped Measures	EUL (years)	Notes
Retro commissioning	Retro commissioning	Electric RCx Measures	7.5	Research may show that EULs for RCx measures vary significantly between RCx categories or RCx delivery methods. If that is not the case, the recommended program level value will continue working well.
Strategic Energy Management	Strategic Energy Management	SEM	5	Only applicable to behavior or operational measures.
Custom – Other	Custom Other	Barrel Wraps for Injection Molders and Extruders Blowers Building Envelope Controls Cooling Tower/Heat Exchanger Filter Injection Molding Machine Low Pressure Drop High Efficiency (Non-HEPA) Air Filters Piping/Duct Modification Pump/Fan Replacement Vacuum System	Custom	This category is intended to capture unique, one-off projects/measures that do not fall under the other recommended end-use categories. Each project/measure should have a custom EUL. To achieve this, the implementer will provide an ex ante EUL for the project/measure and the evaluator will assess it for reasonableness and revise as necessary. As a last resort where there is no basis for a custom EUL a default of 13 years is provided and is deemed appropriate for electric measures.

Source quality will be determined using hierarchy to describe the strength of the identified source as shown in Table 2 below. In cases where a range of values are provided by a source versus an absolute EUL, the median value should be used. In other cases, if more than one high quality source is available with conflicting values, the one with primary research data with strong confidence in the findings should prevail, otherwise, the average EUL should be calculated.

Table 2. Source Strength Type and Examples

	Source Name	Description
TYPE 1: Sources identified as highest strength: Primary research conducted or vetted by third-party entities such as trade organizations, national labs, or government organizations		
1.1	U.S. Department of Energy Federal Energy Conservation Standards	The U.S. Department of Energy (DOE) produces Technical Support Documents (TSD) detailing the analysis behind the federal conservation standards established for each product it regulates. Each TSD contains a chapter, often titled "Life Cycle Cost and Payback Period Analysis", that offers DOE's EUL estimate for the product and explains how this value was derived. Although the method depends on the data available for a given product, DOE's analysis generally relies on some combination of primary research, secondary research, modeling, and/or input from industry experts. The TSDs are linked from DOE's rulemaking page for each product, https://energy.gov/eere/buildings/standards-and-test-procedures. The TSD measure life values are based on shipment data, secondary literature research and primary research which include discussions with industry experts. Navigant considers as high quality because of the stakeholder review process and due diligence required to create these documents. Only the best available sources are used to support the EUL values used in life-cycle cost analysis for DOE federal equipment standards.
1.2	LED lighting reports prepared by Navigant	Navigant has performed extensive market research on the state of LED lighting for the US. DOE Solid State Lighting Program most recently published in 2016. It includes typical lifetime operating hours for each lamp type by sector. https://energy.gov/sites/prod/files/2016/09/f33/energysavingsforecast16_2.pdf
1.3	Appliance Magazine	Appliance Magazine publishes an annual report on the market value, life expectancy, and expected unit replacements for a range of consumer appliances. The appliances listed in this report change from year to year, so older versions of the report may be referenced for products no longer listed. As noted in the report, these EUL estimates represent the expert judgment of magazine staff based on input obtained from many sources. Portrait of the U.S. Appliance Industry (2001-2009). U.S. Appliance Industry: Market Share, Life Expectancy and & Replacement Market, and Saturation Levels (2010). U.S. Appliance Industry: Market Value, Life Expectancy and Replacement Picture (2011-2014).
1.4	C&I Measure Life and Persistence Project	In 2011, Northeast Energy Efficiency Partnership sponsored this study of EUL of commercial and industrial lighting. The primary objective of this study was to conduct primary and secondary research and analysis for estimates of measure lifetimes that included on-site verification of CFL bulbs and fixtures, LED exit signs, HID fixtures, and T8 fixtures. Installations occurred from 1999-2009. http://www.neep.org/sites/default/files/resources/NEEP_CI_Persistence_Report-FINAL.pdf
TYPE 2: Sources identified as medium-high strength: Meta-analyses conducted by third-party organizations, that show some level of evaluating the studies that comprise the dataset		
2.1	California DEER	The most recent and comprehensive DEER documentation of EUL sources was from 2008 and 2014. The 2008 version identifies all the sources reviewed and justification for selected measure life. The 2014 measure list identifies the source used for the measure life. Many of the original references are from 2005, http://deeresources.com/files/deer2005/downloads/DEER2005UpdateFinalReport_ItronVersion.pdf, p. 11-1.

	Source Name	Description
2.2	Regional Technical Forum (RTF) reference workbook	Ongoing revisions as measures undergo review. Similar to the 2008 DEER, the RTF identifies all the sources reviewed and justification for selected measure life.
2.3	GDS Reports	GDS Measure Life Report Residential and Commercial/Industrial Lighting and HVAC Measures – 2007. This study used various data sources such as DEER, state TRMs, and evaluation studies with a working group to review and decide on each value.
2.4	Focus on Energy Report	Focus on Energy Evaluation Business Programs: Measure Life Study Final Report: August 25, 2009 – this is a critical review of studies, workpapers and technical guides including a review of the underlying sources or supporting research.
2.5	ASHRAE	Original source is from Akalin, M.T. 1978. Equipment life and maintenance cost survey (RP-186). ASHRAE Transactions 84(2):94-106; Recent work is ASHRAE system life database (research project 1237-TRP) - which is a crowd-sourced approach to collecting actual system data. https://xp20.ashrae.org/publicdatabase/system_service_life.asp?selected_system_type=7
TYPE 3: Sources identified as medium strength: Compilations conducted by third-party organizations. Original sources should be cited, and locatable where applicable		
3.1	State TRMs	Many state TRMs reference each other and other sources of varying strength. Due diligence on reference documentation is not always present for the measure life. Many TRMs are reviewed via a stakeholder process.
3.2	ENERGY STAR calculators prepared by U.S. EPA and DOE (depending on the references used)	EPA's Energy Star offers calculators to help consumers and businesses estimate the energy and cost savings that could be realized by choosing to buy Energy Star certified products. Within these calculators, Energy Star offers a typical EUL and cites the source. Energy Star generally cites a single high-quality source (e.g., DOE, Appliance Magazine) for each EUL value and offers no analysis or discussion of the selected value. Energy Star's calculators can be accessed at www.energystar.gov. For example, their appliance calculator is available at www.energystar.gov/sites/default/files/asset/document/appliance_calculator.xlsx.
TYPE 4: Sources identified as medium-low strength: Primary research conducted by interested parties such as manufacturers, distributors, retailers or installers		
4.1	Interview with interested parties (with no statistical rigor or analysis)	Manufacturer, distributor, installer, etc. have a vested interest and may overstate the benefit.
TYPE 5: Sources identified as low strength: Source where the basis of measure life is anecdotal, based on design specs, warranty period, etc.		
5.1	Industry blogs, Implementer or Navigant experience	Typically based on professional judgment and not rooted in any data.