Residential Self-report Net-to-Gross Methodology Update and Sensitivity Analysis (MA21X22-B-RSRNTGUP)

May 24, 2022

SUBMITTED TO:
The Massachusetts Electric and Gas Program Administrators

SUBMITTED BY:
NMR Group, Inc.
Tetra Tech, Inc.

Residential Self-report NTG Methodology Update & Sensitivity Analysis Study

Study Objectives

Assess whether the measurement methodology laid out in the 2020 report *Consistent Methodology for Self-Reported Residential Net-to-Gross Measurement* (MA19X03-B-RSRNTG) worked as intended, identify lessons learned from the PY2020 NTG studies that used the methodology, assess the effectiveness of the survey questions and algorithms used to calculate residential NTG, and update documentation to improve future NTG studies and help ensure the methodology is implemented consistently.

Methodology

Review feedback from Conduct sensitivity analysis to assess effectiveness of survey questions and algorithms used to calculate NTG and identify any results highly sensitive to changes in inputs and assumptions.

1. Assess prevalence of "don't know" and "refused" responses
2. Assess impact of using mean, minimum, or maximum scores
3. Assess frequency of inconsistent responses between intent and influence
4. Examine rate of missing components of intent score & frequency of contractor reconciliation

Key Findings

- **Responses of "don't know" or "refused" were relatively infrequent and had minimal impact on overall NTG ratios**
- **Market event questions were asked in each study; however, the skip results of these questions were not used**
- **Inconsistent responses between intent and influence ranged from 3% to 26% of the responses**
- **Using the mean of the intent and influence score is the most reasonable calculation approach**
- **Contractor spillover appears to be a conservative estimate**

Recommendations

1. Adopt the revised self-reported residential NTG methodology.
2. Document any deviations from the methodology in evaluation reports.
3. Continue to review the impact of timing, quantity, and efficiency scores on free ridership and the extent of missing values.
4. Include consistency questions in situations in which intent and influence responses are inconsistent.

Table of Contents

EXECUTIVE SUMMARY **1**
METHODOLOGY 1
CONCLUSIONS 2
RECOMMENDATIONS 3
SECTION 1 INTRODUCTION **5**
1.1 OBJECTIVES 5
SECTION 2 METHODOLOGY **6**
SECTION 3 SENSITIVITY ANALYSIS **8**
3.1 FREE-RIDERSHIP 8
3.1.1 Market Event Type 8
3.1.2 Overall Participant Free-ridership Score 14
3.2 PARTICIPANT SPILLOVER SCORING METHODOLOGY ANALYSIS 30
3.3 INCONSISTENT RESPONSE METHODOLOGY ANALYSIS 37
3.4 CONTRACTOR FREE-RIDERSHIP METHODOLOGY ANALYSIS 39
3.4.1 Contractor Free-ridership Score 40
3.4.2 Application of Contractor Free-ridership 41
3.4.3 Contractor Spillover Score 46
SECTION 4 REVISED RESIDENTIAL SELF-REPORTED NTG SURVEY QUESTIONS AND ALGORITHMS 51
4.1 BACKGROUND 51
4.2 PARTICIPANT MARKET EVENT TYPE 53
4.3 PARTICIPANT FREE-RIDERSHIP 54
4.3.1 Intent Score 54
4.3.2 Influence Score 56
4.3.3 Participant Free-ridership Score 58
4.3.4 Consistency Questions 58
4.4 PARTICIPANT SPILLOVER 59
4.5 CONTRACTOR FREE-RIDERSHIP 62
4.6 PARTICIPATING CONTRACTOR SPILLOVER 64
APPENDIX A SCORING ALGORITHMS **67**
A.1 EVENT TYPE ALGORITHM 67
A.2 PARTICIPANT FREE-RIDERSHIP ALGORITHM 68

A.3 PARTICIPANT SPILLOVER ALGORITHM 69
A.4 CONTRACTOR FREE-RIDERSHIP ALGORITHM 70
A.5 PARTICIPANT AND CONTRACTOR FREE-RIDERSHIP RECONCILIATION 71
A.6 PARTICIPATING CONTRACTOR SPILLOVER ALGORITHM 72
APPENDIX B APPLIANCE RECYCLING FREE-RIDERSHIP ANALYSIS 73
B.1 FREE-RIDERSHIP SCORING METHODOLOGY REVIEW 74
B.1.1 Market Event Type 75

Figures

FIGURE 1: RCD AND SELECT PRODUCTS NTG RESEARCH STUDY MARKET EVENT CATEGORIZATION (N=387) 10
FIGURE 2: MARKET EVENT CATEGORIZATION, EXCLUDING THE NEW CONSTRUCTION/MAJOR RENOVATION DESIGNATION (N=70) 12
FIGURE 3: MARKET EVENT CATEGORIZATION EXCLUDING NEW CONSTRUCTION/MAJOR RENOVATION DESIGNATION (N=159) 13
FIGURE 4: FREE-RIDERSHIP CALCULATION 14
FIGURE 5: SPILLOVER CALCULATION 30
FIGURE 6: CONTRACTOR FREE-RIDERSHIP CALCULATION 40
FIGURE 7: CONTRACTOR SPILLOVER CALCULATION 46
FIGURE 9: APPLIANCE RECYCLING PROGRAM FREE-RIDERSHIP SCORING ALGORITHM 74

Tables

TABLE 1: ALGORITHM COMPONENTS REVIEWED BY STUDY 7
TABLE 2: MARKET EVENT-TYPE QUESTIONS 8
TABLE 3: FREQUENCY OF "DON'T KNOW" AND "REFUSED" RESPONSES IN RCD AND SELECT PRODUCTS NTG RESEARCH STUDY MARKET EVENT QUESTIONS 10
TABLE 4: FREQUENCY OF DON'T KNOW RESPONSES IN RETAIL AND RCD INITIATIVES NTG STUDY MARKET EVENT QUESTIONS 11
TABLE 5: FREQUENCY OF DON'T KNOW RESPONSES IN THE R&A PROGRAM NTG STUDY MARKET EVENT QUESTIONS 12
TABLE 6: OVERALL FREE-RIDERSHIP COMPARISON OF USING THE MEAN OF INTENT AND INFLUENCE VS. THE MIN AND MAX 14
TABLE 7: FREQUENCY OF MISSING INTENT AND INFLUENCE SCORES 15
TABLE 8: FREQUENCY OF MISSING INTENT COMPONENT SCORES 17
TABLE 9: FREQUENCY OF INTENT COMPONENT SCORES USED FOR CALCULATING INTENT SCORE 17
TABLE 10: INTENT SCORING COMPARING THE USE OF THE MINIMUM OF TIMING, QUANTITY, AND EFFICIENCY VS. THE MEAN AND MAX 17
TABLE 11: TIMING QUESTIONS 18
TABLE 12: OVERALL FREE-RIDERSHIP OF MARKET EVENT NEW CONSTRUCTION/MAJOR RENOVATION AND ROF DESIGNATIONS WITH TIMING REMOVED/ DEEMED NON-FREE RIDER 19

TABLE 13: TIMING FREE-RIDERSHIP WITH "DON'T KNOW" AND "REFUSED" EXCLUDED AND DEEMED AS NON-FREE RIDERS 20
TABLE 14: TIMING FREE-RIDERSHIP WITH "DON'T KNOW" AND "REFUSED" EXCLUDED AND DEEMED AS NON-FREE RIDERS BY MEASURE 20
TABLE 15: TIMING FREE-RIDERSHIP WITH "DON'T KNOW" AND "REFUSED" EXCLUDED AND DEEMED AS NON-FREE RIDERS 21
TABLE 16: QUANTITY QUESTIONS 22
TABLE 17: QUANTITY FREE-RIDERSHIP WITH "DON'T KNOW" AND "REFUSED" EXCLUDED AND DEEMED AS NON-FREE RIDERS* 23
TABLE 18: QUANTITY FREE-RIDERSHIP WITH "DON'T KNOW" AND "REFUSED" EXCLUDED AND DEEMED AS NON-FREE RIDERS 24
TABLE 19: EFFICIENCY SCORING QUESTION 26
TABLE 20. EFFICIENCY FREE-RIDERSHIP WITH "DON'T KNOW" AND "REFUSED" EXCLUDED AND DEEMED AS NON-FREE RIDERS 26
TABLE 21: EFFICIENCY FREE-RIDERSHIP WITH "DON'T KNOW" AND "REFUSED" EXCLUDED AND DEEMED AS NON-FREE RIDERS 27
TABLE 22: EFFICIENCY FREE-RIDERSHIP WITH "DON'T KNOW" AND "REFUSED" EXCLUDED AND DEEMED AS NON-FREE RIDERS* 27
TABLE 23: INFLUENCE QUESTIONS 28
TABLE 24: FREE-RIDERSHIP USING MAXIMUM VS. MEAN AND MINIMUM INFLUENCE SCORE 29
TABLE 25: PARTICIPANT SPILLOVER QUESTIONS 30
TABLE 26: DON'T KNOW/REFUSED RESPONSE COUNT FOR PARTICIPANT SPILLOVER QUESTIONS 32
TABLE 27: DON'T KNOW/REFUSED RESPONSE COUNT FOR PARTICIPANT SPILLOVER QUESTIONS FOR DEHUMIDIFIERS AND ADVANCED POWER STRIPS 33
TABLE 28: DON'T KNOW/REFUSED RESPONSE COUNT FOR PARTICIPANT SPILLOVER QUESTIONS FOR THE R&A NTG STUDY 35
TABLE 29: CONSISTENCY QUESTIONS 38
TABLE 30: FREQUENCY OF INCONSISTENT RESPONSES BY PROGRAM 39
TABLE 31: FREQUENCY OF CONSISTENCY CHECK QUESTIONS 39
TABLE 32: CONTRACTOR FREE-RIDERSHIP QUESTIONS 41
TABLE 33: PARTICIPANT AND CONTRACTOR FREE-RIDERSHIP SCORING ASSIGNMENT 42
TABLE 34: CONTRACTOR FREE-RIDERSHIP WITH "DON'T KNOW" AND "REFUSED" EXCLUDED AND DEEMED AS NON-FREE RIDERS 44
TABLE 35: CONTRACTOR FREE-RIDERSHIP WITH "DON'T KNOW" AND "REFUSED" EXCLUDED AND DEEMED AS NON-FREE RIDERS 45
TABLE 36: PARTICIPATING CONTRACTOR SPILLOVER QUESTIONS 46
TABLE 37: DON'T KNOW/REFUSED RESPONSE COUNT FOR CONTRACTOR SPILLOVER QUESTIONS 48
TABLE 40: DON'T KNOW/REFUSED RESPONSE COUNT FOR CONTRACTOR SPILLOVER QUESTIONS 48
TABLE 38: DON'T KNOW/REFUSED RESPONSE COUNT FOR CONTRACTOR SPILLOVER QUESTIONS 49
TABLE 1: ALGORITHM COMPONENTS 52

TABLE 42. MARKET EVENT TYPE QUESTIONS 53
TABLE 43. TIMING, QUANTITY, AND EFFICIENCY (INTENT) QUESTIONS 55
TABLE 44. INFLUENCE QUESTIONS 57
TABLE 45. CALCULATION OF THE INFLUENCE SCORE 58
TABLE 46. CONSISTENCY QUESTIONS 59
TABLE 47. PARTICIPANT SPILLOVER QUESTIONS 60
TABLE 48. CONTRACTOR FREE-RIDERSHIP QUESTIONS 63
TABLE 49. PARTICIPATING CONTRACTOR SPILLOVER QUESTIONS 65
TABLE 50: UMP FREE-RIDERSHIP SCORING 73
TABLE 51: NMR'S FREE-RIDERSHIP ANALYSIS FOR "DON'T KNOW" OR "REFUSED" RESPONSES 75
TABLE 52: ARP – CONDITION OF WORKING EQUIPMENT 75

Acronyms

Acronym	Definition
AC	Air conditioning
APS	Advanced power strip
ARP	Appliance Recycling Programs
C&I	Commercial & industrial
CSO	Contractor spillover
DK	Don't know
EEAC	Energy Efficiency Advisory Council
ER	Early replacement
FR	Free-ridership
HVAC	Heating, ventilation, and air conditioning
IL	Illinois
NC	New construction
NTG	Net-to-gross
NTGR	Net-to-gross ratio
PA	Program Administrators
PSO	Participant spillover
R&A	Renovation and additions
RCD	Residential coordinated delivery
ROF	Replace on Failure
SO	Spillover
SR	Self-report
TRM	Technical Reference Manual
UMP	Uniform Methods Project

ES

Executive Summary

In 2020, the NMR cross-cutting evaluation team developed a methodology to facilitate consistent measurement of self-reported residential net-to-gross (NTG), presented in the study *Consistent Methodology for Self-Reported Residential Net-to-Gross Measurement* (MA19X03-B-RSRNTG).[1] This methodology consisted of questions that could be adapted to measure self-reported NTG for various downstream residential initiatives.

NTG evaluations conducted in PY2020 were the first to implement the methodology. As evaluation teams started fielding the survey questions and analyzing the resulting responses, they asked about adapting the batteries to scenarios not explicitly addressed in the methodology. They suggested modifications to avoid misinterpretation and improve the quality of the results.

The objectives of this study were to assess whether the methodology, including the event-type questions, was working as intended and identifying lessons learned from the NTG studies that used the methodology. Another objective was to update the documentation as needed to improve future NTG studies and help ensure that the methodology is implemented consistently.

This document assesses the methodology, including the event type questions, identifies lessons learned from the NTG studies that used the guidance, and provides updates to the methodology document to improve future NTG studies. The document includes an updated methodology reflecting the clarifications and modifications identified during the 2020 NTG evaluations. It also discusses the sensitivity analysis results that assessed the effectiveness of the survey questions and the algorithms used to calculate residential NTG.

METHODOLOGY

The evaluation team obtained feedback from and performed sensitivity analysis on the data from the following three NTG studies: *Massachusetts Residential Programs Net-to-Gross Research of RCD and Select Products Measures* (MA20R28-B-NTGRCDP), *Residential Products Net-to-Gross Study* (MA20X04-E-PRODNTG), and *Renovations and Additions Net-to-Gross Study* (MA20X06-B-RENOADDNTG). The Appliance Recycling Net Savings Update (MA20X03-E-ARNTG) study was initially included in the work scope; however, it follows an approach adapted from the Uniform Methods Project (UMP) to guide the estimation of retrospective net energy savings and NTGRs.[2] Because of this, there are no direct comparisons to the other studies.

The sensitivity analysis assessed the effectiveness of the survey questions and the algorithms used to calculate residential NTG. To identify any results that are highly sensitive to changes in

[1] Consistent Methodology for Self-Reported Residential Net-to-Gross Measurement (MA19X03-B-RSRNTG), May 28, 2020; https://ma-eeac.org/wp-content/uploads/MA19X03-B-RSRNTG_Residential-SR-NTG-Report_FINAL_2020.5.28.pdf

[2] Keeling, J.; Bruchs, D. 2017. "Chapter 7: Refrigerator Recycling Evaluation Protocol." *The Uniform Methods Project: Methods for Determining Energy-Efficiency Savings for Specific Measures.* Golden, CO; National Renewable Energy Laboratory. NREL/SR-7A40-68563. http://www.nrel.gov/docs/fy17osti/68563.pdf.

inputs and assumptions, the team assessed (1) the prevalence of "don't know" and "refused" responses in the intent and influence components, (2) the impact of using mean, minimum, or maximum scores when calculating free-ridership and spillover, and (3) the frequency of inconsistent responses between intent and influence, and how the inconsistency was addressed, if at all. The sensitivity analysis also examined how many components of the intent score were missing and the frequency of contractor reconciliation, including differences between participant and contractor results and how contractor responses were used.

CONCLUSIONS

Based on lessons learned from the 2020 NTG evaluation teams, input from the PA Working Group convened for this study, and the results of the sensitivity analysis, the evaluation team updated the self-report NTG methodology initially presented in the MA19X03-B-RSRNTG study, *Consistent Methodology for Self-Reported Residential Net-to-Gross Measurement* ("Consistent Methodology"). The updated methodology and flowcharts, presented in Section 4 and Appendix A, supersede those presented in MA19X03-B-RSRNTG. All revisions were minor and mostly consisted of clarification around question-wording.

Results from the sensitivity analysis showed that responses of "don't know" or "refused" were relatively infrequent and had minimal to no impact on overall NTG ratios using "Consistent Methodology." The evaluation team also found that responses to specific questions varied due to program implementation differences. These variations do not make them "wrong" or "ineffective," but rather indicate additional context should be applied during reporting.

Market event questions were asked in each study. However, the skip results of these questions were not used in the free-ridership questions for any of the studies. As a result, any respondents who indicated the equipment was part of a new construction or major renovation or replaced on failure were asked the timing questions. This is contrary to "Consistent Methodology," in which timing is not a factor if the equipment fails or needs major repairs. Our analysis shows that free-ridership decreases when the timing scores that should have been excluded from the calculation of free-ridership are removed; however, the change in free-ridership was minimal (+/-1%), resulting in slight changes to overall NTG. This supports the approach laid out in "Consistent Methodology" of excluding timing from free-ridership calculations for failing equipment. By contrast, excluding such results from the dehumidifiers in the Products NTG study slightly increased free-ridership. This difference is likely because a failing dehumidifier is not as critical to a respondent as a failing furnace.

A review of the free-ridership questions found inconsistent responses between intent and influence, ranging from 3% and 26% of the responses. These are situations in which a respondent indicated no intent but high program influence (most common for the R&A program) or high intent but low program influence (most common in the RCD and select products NTG study). The Consistency Methodology document recommended including two questions to probe inconsistent responses between intent and influence, but only the RCD and select products NTG study included them. After reviewing the data, we recommend keeping the two questions as part of the methodology.

A review of calculating free-ridership using the mean of the intent and influence score showed this is the most reasonable approach. The difference between using the minimum, mean, and maximum of the two varies greatly, indicating high sensitivity to the scoring approach.

Contractor spillover appears to be a conservative estimate. When the evaluation team looked at "don't know" and "refused" responses, 40% of the RCD and select products NTG study respondents could not estimate their total measure sales to allow the program to capture any spillover. Any respondents who cannot answer the spillover questions, per the methodology, are removed from the calculation.

RECOMMENDATIONS

1. Adopt the revised self-reported residential NTG methodology presented in Section 4 and Appendix A. The information presented here replaces the methodology presented in Consistent Methodology for Self-Reported Residential Net-to-Gross Measurement (MA19X03-B-RSRNTG).
2. Document any deviations from the self-reported residential NTG methodology in future evaluation reports.
3. Continue to review the impact of timing, quantity, and efficiency scores on free-ridership and the extent of missing values.
4. Include consistency questions, as outlined, in situations where intent and influence responses are inconsistent and use expert judgement from the review of open ends to determine if any cases need to be dropped.

Below we summarize the updates to the NTG questions and calculations.

Update 1. Added text to clarify that the market event questions should be asked for all standalone downstream measures in the market event battery. Weatherization, insulation, and direct install measures completed as part of an audit should be skipped out of this battery.

Update 2. Updated the market event questions to treat new construction and major renovations separately. Major renovation respondents can continue through the market event questions (and timing if applicable) while new construction respondents will continue to skip out.

Update 3. Updated the intent questions to clarify that for incentive-based measures questions, the wording of questions for direct install measures received free as part of an audit should include the phrase *"If you had not received the free <measure(s)> during the audit, …".*

Update 4. Modified the methodology efficiency section to guide how questions can be adjusted for smart thermostats. Provided the efficiency question wording and scoring guidance: *"Without the rebate or assistance from <Sponsor> for your smart thermostat, what type of a thermostat would you have likely installed?*

Update 5. Added text for participant spillover about calculating spillover across fuel types. *"To calculate spillover, at a minimum, the survey demographics should collect information on water heater fuel, clothes dryer fuel, primary heating fuel, and type of cooling equipment to determine what prescriptive savings to apply."* To calculate program level spillover, sum the deemed savings

for all spillover measures, converting to MMBtu, and divide the spillover savings by program savings

Update 6. The participant spillover methodology section included response categories as optional guidance for web surveys when asked how respondents knew the equipment was energy efficient.

Update 7. Added clarification to the participant spillover section about how spillover should be calculated. *"A spillover score is calculated for each measure as follows. If the score is 5 or lower, spillover is set to 0. If the score is higher than 5, the spillover savings is the appropriate TRM (or secondary sources) savings."*

Also added is the text: *"The spillover rate is equal to the sum of the savings associated with the additional TRM (or secondary sources) spillover measure divided by the sum of the program savings. To allow for spillover for both gas and electric measures, savings should be converted to MMBtus.*"

Update 8. Corrected the spillover calculation to read "*Participant Spillover score = (S4 + S5)) / 2.*"

Update 9. Updated the methodology in the contractor free-ridership section to clarify how the results of the contractor surveys should be used. *"Responses to these questions will be used to calculate a contractor free-ridership score that will replace the participant's free-ridership score or averaged with when participants rated the contractor highly influential in doing the project and used the contractor's equipment recommendations. Specifically, for all customer respondents who said their contractor made the decision or influenced it heavily, replace the customer's free-ridership score with the weighted free-ridership of all contractors for the given measure."*

Update 10. Added clarification to the contractor survey that contractors should be asked about the most active sector. This will reduce respondent burden and improve contractors' focus in multiple sectors.

Update 11. Removed NS3 and NS4 from the contractor spillover battery because they are duplicate questions or can be calculated from CF0 and CF0a. Clarified instructions to read *"NS3 can be calculated from NS2-CF0 and NS4 can be calculated from (NS2-CF0)/NS2."*

Update 12. Added an introductory sentence to NS1 in the contractor spillover section. "Moving on from the <CF0 or CF0a> units of high efficiency <measure>that received a Mass Save incentive to talk about **all** your sales of residential <measure>, including standard-efficiency and high-efficiency and rebated and non-rebated units. About how many did your company sell **in total** in <year>?"

Update 13. Added clarification to the contractor spillover section that this battery is asked of participating contractors. There is no consistent methodology for using nonparticipant contractor research in calculating spillover.

Section 1 Introduction

In 2020, the NMR cross-cutting evaluation team developed a methodology to facilitate consistent measurement of self-reported residential net-to-gross (NTG), presented in the study *Consistent Methodology for Self-Reported Residential Net-to-Gross Measurement* (MA19X03-B-RSRNTG). This methodology consisted of questions that could be adapted to measure self-reported NTG for various downstream residential initiatives.

NTG evaluations conducted in PY2020 were the first to implement the methodology. As evaluation teams started fielding the survey questions and analyzing the resulting responses, they posed important questions about adapting the batteries to scenarios not explicitly addressed in the methodology. The evaluation teams suggested modifications to avoid misinterpretation and improve the quality of the results. This document summarizes the evaluation teams' feedback and presents the resulting changes to the questions and algorithms. It also shows the sensitivity analysis results that assessed the effectiveness of the survey questions and algorithms used to calculate NTG for the residential NTG studies completed in 2021.

The methodology document consisted of event type (identification of replacing on failure, early replacement, or new equipment), intent, and influence questions that could be adapted to measure self-reported NTG from various downstream residential initiatives. While the questions used were not formally pretested, most had been used in other jurisdictions and were adapted to fit the Massachusetts programs and framework.

1.1 Objectives

The objectives of this study were to assess whether the methodology, including the event-type questions, was working as intended and identifying lessons learned from the NTG studies that used the methodology. An additional objective was to update the documentation as needed to improve future NTG studies and help ensure that the methodology is implemented consistently.

Section 2 Methodology

The evaluation team conducted various levels of sensitivity analysis to assess the effectiveness of the survey questions and the algorithms used to calculate residential NTG and identify any highly sensitive results to changes in inputs and assumptions. This included assessing (1) the prevalence of "don't know" and "refused" responses in the intent and influence components, (2) the impact of using mean, minimum, or maximum scores when calculating free-ridership and spillover, and (3) the frequency of inconsistent responses between intent and influence, and how the inconsistency was addressed, if at all. The sensitivity analysis also examined how many components of the intent score were missing and the frequency of contractor reconciliation. This included differences between participant and contractor results and how contractor responses were used. Frequencies of results are presented throughout this report, and where possible, data were manipulated to visualize the effects of response or calculation modifications. The evaluation team also examined how many components of the intent score were missing and the frequency of contractor reconciliation. This analysis included differences between participant and contractor results and how contractor responses were used.

The evaluation team obtained feedback from and performed sensitivity analysis on the data from the following three NTG studies:

- ***Massachusetts Residential Programs Net-to-Gross Research of RCD and Select Products Measures* (MA20R28-B-NTGRCDP). For convenience, we refer to this study as the "RCD and select products NTG study."** This study measured NTG for selected equipment supported through the Retail and RCD Initiatives. The eligible sample was 2019 program participants who installed one or more of the following rebated measures: direct install measures (duct sealing/insulation, water-saving devices, and pipe wrap), weatherization measures (air sealing and insulation), non-electric heating equipment (boilers and furnaces), electric HVAC equipment (central ACs and heat pumps), water heaters, and thermostats (wi-fi and programmable). The study fielded three surveys with NTG questions: a single-family RCD and retail participant survey, a multifamily RCD participant survey, and a participating contractor survey. The participant survey yielded 554 responses representing 623 measures completed out of a sample frame of 6,332. For multifamily, 23 surveys were completed from a sample size of 1,400. A total of 60 surveys were targeted, and 69 were completed representing 164 measures completed.
- **The Residential Products Net-to-Gross Study (MA20X04-E-PRODNTG). For convenience, we refer to this as the "products NTG study."** This study fielded a web based NTG and process survey of 2019 Retail and RCD Initiatives participants. The completed sample included 44 leave-behind advanced power strip (APS) recipients, 72 participants who purchased one or more APSs with a rebate, and 70 participants who purchased one or more dehumidifiers.

- **The Renovations and Additions Net-to-Gross Study (MA20X06-B-RENOADDNTG). For convenience, we refer to this as the "R&A NTG study."** This study measured self-reported NTG for the PA's Renovations and Additions program via a homeowner participant survey and a participating contractor survey, addressing water heating, heating and cooling, insulation, air infiltration, ducts, and windows. Of the 664 program participants, 120 surveys were completed representing 468 measures. A total of 31 contractor surveys were completed representing 263 measures.

The Appliance Recycling Net Savings Update (MA20X03-E-ARNTG) study was initially included in the work scope. However, it follows an approach adapted from the Uniform Methods Project (UMP) to guide the estimation of retrospective net energy savings and NTGRs.[3] Because of this, there are no direct comparisons to the other studies, but some respondents answered "don't know"/"refused"; market event analysis can be found in Appendix B.

Not all components of the self-report methodology document apply to all three evaluations. Table 1 below shows the components that were relevant to each evaluation.

Table 1: Algorithm Components Reviewed by Study

Algorithm component	RCD and Select Products NTG (MA20R28-B-NTGRCDP)	Retail and RCD Initiatives NTG (MA20X04-E-PRODNTG)	R&A Program NTG (MA20X06-B-RENOADDNTG)
Market event	Y	Y - Dehumidifiers N - APS	Y - partial
Free-ridership - Intent	Y	Y - Dehumidifiers N - APS excludes efficiency FR	Y - excludes ducts/air infiltration measures
Free-ridership - Influence	Y	Y	Y
Participant spillover	Y	Y	Y
Contractor free-ridership	Y	N	Y
Contractor spillover	Y	N	Y
Inconsistency checks	Y	N	N

[3] Keeling, J.; Bruchs, D. 2017. "Chapter 7: Refrigerator Recycling Evaluation Protocol." *The Uniform Methods Project: Methods for Determining Energy-Efficiency Savings for Specific Measures.* Golden, CO; National Renewable Energy Laboratory. NREL/SR-7A40-68563. http://www.nrel.gov/docs/fy17osti/68563.pdf.

Section 3 Sensitivity Analysis

Here we detail the various sensitivity analysis results for the RCD and select products NTG study (MA20R28-B-NTGRCDP), the products NTG study (MA20X04-E-PRODNTG), and the R&A NTG study (MA20X06-B-RENOADDNTG) for four topics: free-ridership, participant spillover, inconsistent response methodology, and contractor scoring.

3.1 FREE-RIDERSHIP

Here we discuss the following where it is applicable:

- Market event type categorization
- Intent scoring
- Influence scoring
- Overall participant free-ridership

3.1.1 Market Event Type

Participant market event type questions identify whether the high-efficiency equipment installed was due to: new construction or major renovation; replacing still-functioning equipment (early replacement); replacing failed equipment (replace-on-failure, or ROF); or because the respondent installed a piece of equipment not previously used in their home (new equipment). Responses to these questions identify whether participants should be asked the timing questions in the free-ridership (FR) battery.[4] Event-type questions should be asked for all stand-alone measures, including thermostats. Table 2 below shows the market event questions that are part of the FR battery.

Table 2: Market Event-Type Questions

Market Event			
ET1	Was the high efficiency <measure> <rebated and> installed as part of a new construction project? (select one)		
	01	Yes	[new construction] SKIP TO Next Section
	02	No	
	88	Don't know	
	99	Refused	

[4] The event type questions can also be used for other purposes—for example, to give the PAs information to update the percent of units that are early retirement.

ET2	Did the high efficiency <measure> you installed replace any existing <equipment> or was it a new type of equipment that you did not have in your home before? (select one) 01 Replaced existing equipment 02 New equipment [new equipment] SKIP TO Next Section 88 Don't know 99 Refused
ET3	[ASK IF ET2=1, DK, R] Which of the following best describes the condition of your old equipment? (read list) 01 The old equipment was working with no need of repair 02 The old equipment was working with need of minor repair 03 The old equipment was working with need of major repair [ROF] 04 The old equipment was no longer working [ROF] 88 Don't know 99 Refused
ET4	[ASK IF ET3 = 1, 2, DK, and R] Do you think your old equipment would have lasted another two years? (select one) 01 Yes [ER] 02 No [ROF] 88 Don't know [ROF] 99 Refused [ROF]

3.1.1.1 Summary of Results

RCD and select products NTG study (MA20R28-B-NTGRCDP)

Don't Know/Refused Response Review

Reviewing the frequency of "don't know" and "refused" responses showed that rates of "don't know" and "refused" are reasonable and not an issue for ET1, ET2, and ET3. However, 25% of respondents could not answer ET4, which asked if they thought their equipment would have lasted another two years. As a result, this group's equipment was deemed a "replace on failure" (ROF). Per the methodology, these respondents should have been skipped past the timing questions, but they were not, so a timing score was calculated. The evaluation team reviewed the impact of including and excluding these timing scores for the ROF equipment. The results of this analysis can be found in Section 3.1.2.1.2 Timing Score.

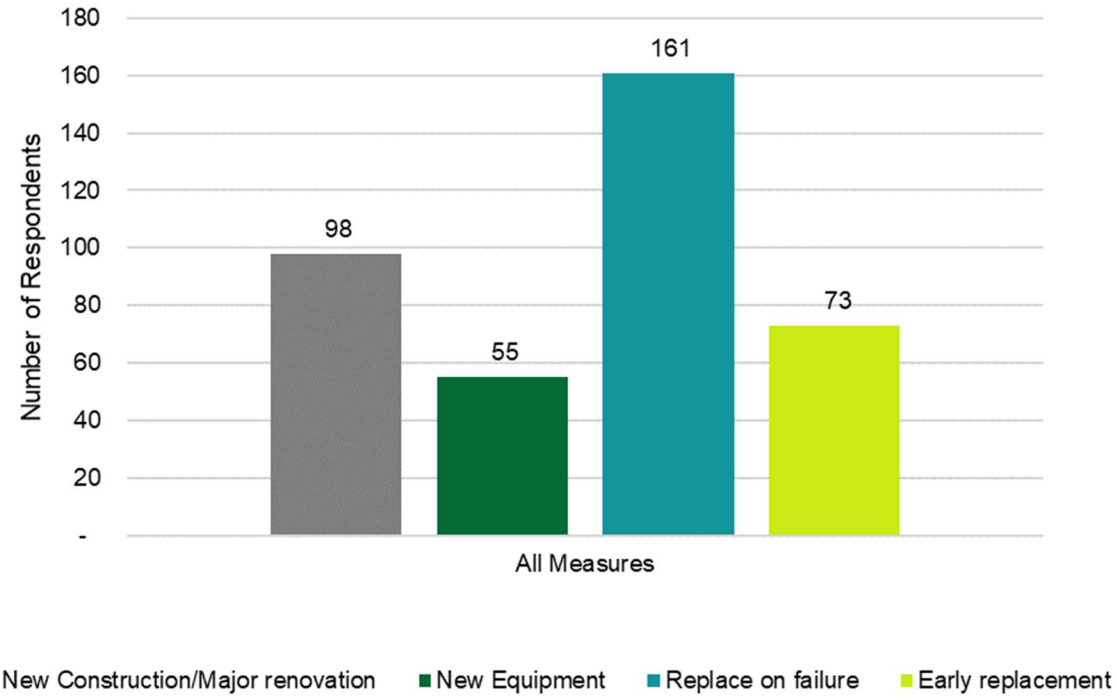
Table 3: Frequency of "Don't Know" and "Refused" Responses in RCD and Select Products NTG Research Study Market Event Questions

Market event questions	Measure Count	Don't know/Refused (n)	Don't know/Refused %
ET1 (was installed as part of NC or major reno)	387	3	1%
ET2 (did it replace existing or was it new to home)	289	1	0%
ET3 (condition of old equipment)	234	5	2%
ET4 (would equipment last another two years)	128	32	25%

Market Event Assignment

After applying the market event designations, most measures were reported to be ROF (161 of 387 respondents), followed by "new construction/major renovation" (91 respondents of 387 respondents). New equipment was reported at the lowest rate (55 of 387 respondents).

Figure 1: RCD and Select Products NTG Research Study Market Event Categorization (n=387)[5]



[5] Not all measures received market event questions. Weatherization, direct install, and thermostats are excluded.

Products NTG study (MA20X04-E-PRODNTG)

Don't Know/Refused Response Review

In the products NTG study, no respondents answered "don't know" or "refused" to ET2 and ET3. The first market event question did not ask about dehumidifiers (ET1, if the installation was part of a new construction or a major renovation). The entire market event question battery was not asked for advanced power strips. When asked if their dehumidifiers would last another two years, 36% of respondents answered, "don't know" or "refused" (ET4). As with the RCD and select products NTG study, the ROF respondents should have been skipped past the timing questions, but they were not, so a timing score was calculated. The evaluation team reviewed the impact of including and excluding these timing scores for the ROF equipment. The results of this analysis can be found in Section 3.1.2.1.2 Timing Score.

Table 4: Frequency of Don't Know Responses in Retail and RCD Initiatives NTG Study Market Event Questions

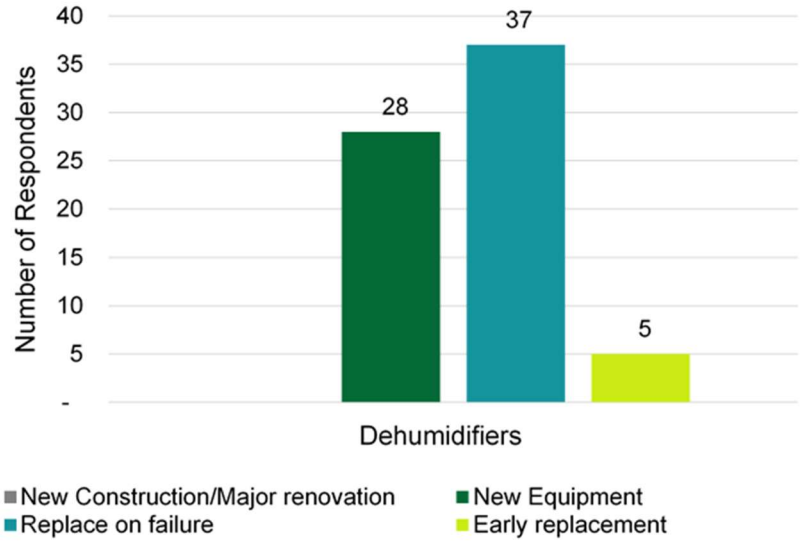
Market event questions	Measure Count	Don't know/Refused (n)	Don't know/Refused %
ET1 (was installed as part of NC or major reno)*	N/A	N/A	N/A
ET2 (did it replace existing or was it new to home)**	73	0	0%
ET3 (condition of old equipment)**	45	0	0%
ET4 (would equipment last another two years)**	11	4	36%

* Not asked of APSs or dehumidifiers.
** Not asked of APSs.

Market Event Assignment

After applying the market event designations, most dehumidifiers were identified as ROF (37 of 70 respondents), followed by new equipment (28 respondents of 70 respondents), with early replacement far behind at five of 70 respondents.

Figure 2: Market Event Categorization, Excluding the New Construction/Major Renovation Designation (n=70)



R&A NTG study (MA20X06-B-RENOADDNTG)

Don't Know/Refused Response Review

In the R&A study, only one respondent answered "don't know" or "refused" to questions ET2 and ET3; however, 26% answered "don't know" or "refused" to ET4. ET1 was not asked because the program is for additions and renovations. Per the methodology, replace on failure and new construction/major renovation respondents should have skipped the timing questions. As with the previous two studies, there was no skip instruction, so a timing score was calculated for these respondents. The evaluation team reviewed the impact of including and excluding these timing scores for the ROF equipment. The results of this analysis can be found in Section 3.1.2.1.2 Timing Score.

Table 5: Frequency of Don't Know Responses in the R&A Program NTG Study Market Event Questions

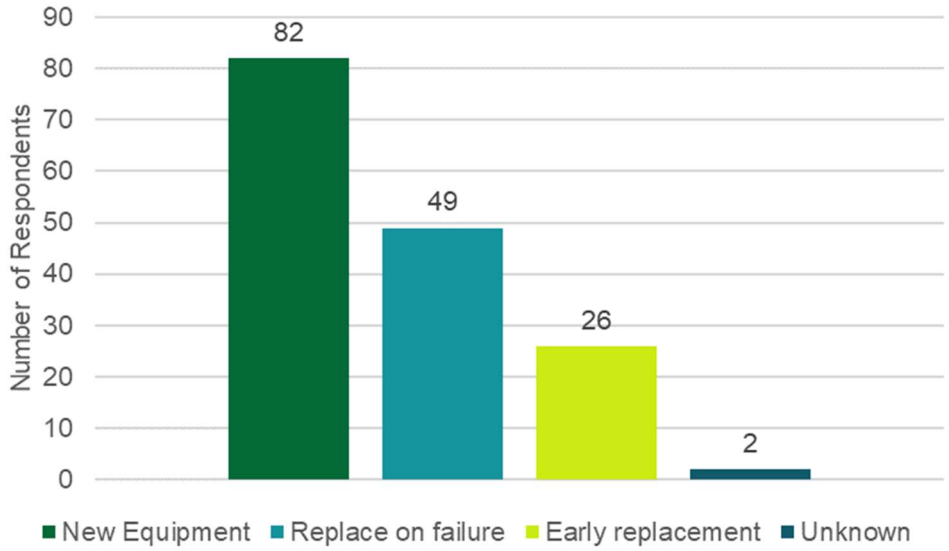
Market event questions	Measure Count	Don't know/Refused (n)	Don't know/Refused %
ET1 (was installed as part of NC or major reno)	N/A	N/A	N/A
ET2 (did it replace existing or was it new to home)	162	1	1%
ET3 (condition of old equipment)	76	1	1%
ET4 (would equipment last another two years)	50	13	26%

Market Event Assignment

By the rule of the algorithm in Table 2, all projects under the R&A NTG study would be classified as a "new construction/major renovation" market event. However, the survey still asks the market

event questions, except ET1 in Table 2 above. Using ET2, ET3, and ET4 from Table 2, we can determine the market event outside the "new construction/major renovation" designation. Figure 3 below shows measures being classified as "new equipment" were mentioned most (82 or 159 respondents), followed by "replace on failure" (49 of 159 respondents).

Figure 3: Market Event Categorization Excluding New Construction/Major Renovation Designation (n=159)



We recommend that the questions be asked for all major renovations.

3.1.1.2 Evaluation Team Questions, Responses, and Resolutions

Here we summarize (1) questions related to measuring market event type asked by the evaluation teams as they implemented the market event-type battery, (2) the *Consistent Methodology for Self-Reported Residential Net-to-Gross Measurement* team's responses, and (3) the resolution for each question.

Question: Does event type make sense for direct install or weatherization measures? Also, does it make sense for thermostats as a stand-alone measure? There is a general note in the guidance document alluding to this, but it is not measure-specific.

Response: Market event questions can be asked for all standalone measures, including thermostats. Thermostats can replace existing equipment that is not working, making these questions relevant for thermostats and other similar equipment.

Resolution: Modified explanation of eligibility to read, *"Market event type questions do not need to be asked for some types of measures, like weatherization and insulation, where event type does not impact timing.*

Question: Is the frequency of "don't know" and "refused" responses for the event-type questions reasonable?

Response: A participant could be assigned one of four categories: new construction, new equipment, replace-on-failure, and early replacement. Of the three studies, only the RCD and select products NTG study asked for the entire market event battery. Per the results shown above in Section 3.1.1.1, the number of "don't know" or "refused" responses for ET1, ET2, and ET3 are negligible; however, ET4, which asks if they believe the equipment would have lasted another two years, had a higher instance of "don't know" or "refused" (25%). Only one respondent from the R&A NTG study answered, "don't know."

Resolution: No resolution is needed, although given the prevalence of "don't know" or "refused" responses to ET4 in the RCD and select products NTG study, it may be worth continuing to explore the impacts on the timing score since that is skipped in this scenario (see Section 3.1.2.1 below).

3.1.2 Overall Participant Free-ridership Score

Participant free-ridership is calculated by taking the mean of intent and influence scores in all three studies examined (as shown in Figure 4). This section breaks down the two components, looking at frequencies of "don't know" and "refused" responses and free-ridership results using alternate approaches.

Figure 4: Free-ridership Calculation

Free-ridership = mean(intent score, influence score)

Table 6 shows free-ridership results calculated using the "Consistent Methodology" mean intent and influence score. The minimum and maximum of the two scores illustrate the sensitivity of FR results to alternate approaches. The difference between the minimum, mean, and maximum varies greatly, indicating high sensitivity to the scoring approach. Using the mean is the most reasonable approach given this high sensitivity.

Table 6: Overall Free-ridership Comparison of Using the Mean of Intent and Influence vs. the Min and Max

Study	Respondent (n)	Measure Count	FR Using Mean (Reported)	FR Using Min	FR Using Max
RCD and select products NTG study	554	623	31%	10%	37%
Products NTG study - Dehumidifiers	70	70	42%	12%	72%
Products NTG study - Leave Behind APS*	44	44	9%	9%	9%
Products NTG study - Purchased APS	72	72	15%	2%	29%
R&A NTG study	120	468	34%	17%	56%

*The mean, min, and max results for the leave-behind APS measure are the same because influence is not calculated.

Table 7 shows the frequency of missing intent and influence scores when calculating overall free-ridership. Results show no instances of missing intent scores and only two cases of missing influence scores in the RCD and select products NTG study.

Table 7: Frequency of Missing Intent and Influence Scores

Study	Respondent (n)	Measure Count	Missing Intent	Missing Influence
RCD and select products NTG study	554	623	0	2
Products NTG study - Dehumidifiers	70	70	0	0
Products NTG study - Leave Behind APS*	44	44	0	n/a
Products NTG study - Purchased APS	72	72	0	0
R&A NTG study	120	468	0	0

*Influence is not calculated for the leave-behind APS measure.

The following section discusses the sensitivity analysis of free-ridership scoring, including the three intent scoring components (timing, quantity, and efficiency) and the calculation of the influence score.

3.1.2.1 Intent Score

The intent score incorporates a series of questions to determine whether, in the absence of the program, the respondent:

- would have purchased any type of equipment at the same time, at a later date, or not at all (timing)
- would have purchased the same, lesser, or greater quantity of the high-efficiency equipment (quantity), or
- would have purchased the same or lesser efficiency equipment (efficiency)

There are no instances of missing intent scores within any of the studies. There are, however, missing components within the intent score, which is to be expected because not all components apply to all measures or studies.

Table 8 below shows the frequency of missing timing, quantity, and efficiency scores by study. Missing quantity scores for the RCD and select products NTG study is high because most measures would not be purchased in quantities greater than one (e.g., heating and cooling equipment, water heaters, etc.), so those questions are not asked. The efficiency scoring results are also higher because efficiency is not applicable for weatherization-type measures. Leave-behind APS in the Products NTG study are primarily treated as direct-install measures, so unless the participant is aware of APS before their home energy assessment (and only ten were), all components are skipped, and free-ridership is set to zero. The R&A study asked only about timing and efficiency.

Table 8: Frequency of Missing Intent Component Scores

Study	Measure Count	Missing timing	Missing quantity	Missing efficiency
RCD and select products NTG study	623	98	570	266
Products NTG study - Dehumidifiers	70	0	67	0
Products NTG study - Leave Behind APS*	44	34	36	n/a
Products NTG study - Purchased APS*	72	0	0	n/a
R&A NTG study**	120	52	n/a	0

*Efficiency is not calculated for APS measures in the Products NTG study.
**Quantity is not calculated for any measures in the R&A NTG study.

The intent score takes the minimum value from the three scoring components. The evaluation team looked at which scoring component was used most often. Table 9 shows that, where applicable, efficiency is the largest driver of the three scores. Quantity was the smallest driver across all three studies.

Table 9: Frequency of Intent Component Scores Used for Calculating Intent Score

Study	Measure Count	Timing	Quantity	Efficiency
RCD and select products NTG study	623	279	7	486
Products NTG study - Dehumidifiers	70	51	3	62
Products NTG study - Leave Behind APS	44	10	5	n/a
Products NTG study - Purchased APS	72	59	44	n/a
R&A NTG study	468	331	n/a	337

“Consistent Methodology” calculates an intent score using the minimum timing, quantity, and efficiency scores. Table 10 would show the intent results if the mean or maximum of these scores were used instead of the minimum. Any calculation other than the minimum would result in an increase in free-ridership. The increased effect is highest for purchased APS under the products NTG study, an increase of 10% in free-ridership. The next highest are measures within the R&A NTG study (8% increase). Using the mean had the lowest effect on the RCD and select products NTG study (2% increase). Applying the maximum value doubled the net difference in free-ridership compared to the mean.

Table 10: Intent Scoring Comparing the Use of the Minimum of Timing, Quantity, and Efficiency vs. the Mean and Max

Study	Measure Count	FR Using Min (Reported)	FR Using Mean	FR Using Max
RCD and select products NTG study	623	31%	33%	35%
Products NTG study – Dehumidifiers	70	42%	47%	53%
Products NTG study - Leave Behind APS	44	9%	13%	16%
Products NTG study - Purchased APS	72	15%	25%	35%
R&A NTG study	468	34%	42%	50%

3.1.2.1.1 Evaluation Team Questions, Responses, and Resolutions

Question: It is not clear in the methodology that survey language for incentive-based measures should be modified for direct install measures, as those are received free.

Response: While it is impossible to define wording for every permutation, for the single-family and multifamily surveys, some questions had to be changed slightly for direct install equipment. The *Consistent Methodology for Self-Reported Residential Net-to-Gross Measurement* survey questions referred to equipment for which the customer received a rebate; however, for direct install equipment, which customers get for free, the wording should be changed to highlight the free nature of the equipment.

Resolution: For incentive-based measures questions, the evaluation team clarified that the wording of questions for direct install measures received free as part of an audit will need to be changed. Evaluation teams should include the phrase *"If you had not received the free <measure(s)> during the audit, …"*

3.1.2.1.2 Timing Score

The timing score determines whether the program influenced when participants installed the measure (Table 11). While the methodology states that timing questions should be skipped if the measure is categorized as new construction/major renovation or replace on failure based on the market event question battery, results show that timing questions were still asked, and scoring was applied for all studies. The following results show the impact of including and excluding timing results deemed new construction/major renovation or replace on failure and the frequency of "don't know" and "refused" responses to timing questions.

Table 11: Timing Questions

Timing (Evaluator: if NC or ROF, do *not* ask participants the timing questions. New equipment and early replacement should be asked the timing questions.)			
T1	Without the <list of all rebates/incentives/support received> from <Sponsor>, how likely is it that you would have installed any type of <equipment> at the same time? Would you say …		
	01	Not at all likely	FR_timing = 0 (skip to 'influence')
	02	Slightly likely	
	03	Somewhat likely	
	04	Very likely	FR_timing = 1.0 (skip to Q1)
	88	Don't know	FR_timing = 0.5 (skip to Q1)
	99	Refused	FR_timing = 0.5 (skip to Q1)
T2	Without the <list of all rebates/incentives/support received> from <Sponsor>, when do you think you would have installed the <equipment>? Would you say …		
	01	Within 6 months of when you did	FR_timing = 1.0
	02	Between 6 months and 12 months of when you did	FR_timing = 0.5
	03	More than 1 year after when you did	FR_timing = 0 (skip to 'influence')

88	Don't know	FR_timing = 0.5
99	Refused	FR_timing = 0.5

Table 12 below would show overall free-ridership results if the market event designations of new construction/major renovation or replace on failure resulted in either skipped timing questions or deemed non-free riders. For the RCD and select products NTG study, free-ridership results decreased by 6%, supporting the methodology's assumption that timing is not a factor if the equipment was ROF, as they would have had to replace it anyway. For the other two studies, removing ROF timing results marginally increased free-ridership (<2%), indicating some program influence on timing even with failing equipment.

As discussed, the methodology assumes that failing equipment would have no program impact on timing. While there is no consideration of deeming ROF measures non-free riders because of this assumption, Table 12 includes free-ridership results for ROF measures deemed non-free riders for comparative purposes. As expected, free-ridership would go down across all studies. This reduction was most significant for both the RCD and select products NTG study and the Products NTG study, respectively, with a 52% and 58% reduction in free-ridership.

Table 12: Overall Free-ridership of Market Event New Construction/Major renovation and ROF Designations with Timing Removed/ Deemed Non-free rider

Study	Measure Count	Count of NC and ROF Designations	Overall FR (reported)	FR with ROF/NC removed	FR with ROF/NC deemed as non-FR**
RCD and select products NTG study	623	153	31%	25%	16%
Products NTG study *	70	37	42%	44%	24%
R&A NTG study	468	131	34%	35%	30%

* Market event questions are not asked for APSs.
** This approach is not recommended and is included solely for comparative purposes.

3.1.2.1.2.1 *Summary of Results*

RCD and select products NTG study

Just under one-tenth of respondents (10%) answered the timing questions with "don't know" or "refused," resulting in the imputation of 0.5 free-ridership. Thermostats had the highest "don't know/refused" response rate at 35% (21 of 60 respondents). Next was electric HVAC (15 of 60 respondents) and weatherization (12 of 60 respondents). If "don't know" and "refused" were treated as non-free riders, overall NTG would increase by 3%.

Table 13: Timing Free-ridership with "Don't know" and "Refused" Excluded and Deemed as Non-Free Riders

Questions	Number of Eligible R's	Don't Know /Refused (n)	Don't Know/ Refused %	Reported				Don't know/Refused Responses Removed		Don't know/Refused Responses Deemed Non-Free riders	
				FR	P SO	C SO	NTG Ratio	FR	NTG Ratio	FR	NTG Ratio
T1 Without the <rebates> from <Sponsor>, how likely is it that you would have installed any type of <equipment> at the same time?	527	9	2%	31%	12%	4%	85%	30%	86%	28%	88%
T2 Without the <rebates> from <Sponsor>, when do you think you would have installed the <equipment>?	243	51	21%								

Products NTG study

In the Products NTG study, timing was asked for all three measures (dehumidifiers, leave-behind APS, and purchased APS). Table 14 below shows the breakdown of "don't know" and "refused" responses by measure type. Most respondents were able to answer whether they would have installed any of the measures at the same time absent the program. However, instances of "don't know" and "refused" increased for the follow-up question asking when they would have installed the measure. Timing free-ridership saw minimal changes when removing the "don't know" and "refused" responses from the free-ridership calculation but did see an increase in NTG when treating those responses as non-free riders, ranging between a 1% and 4% increase.

Table 14: Timing Free-ridership with "Don't know" and "Refused" Excluded and Deemed as Non-Free Riders by Measure

	Questions	Number of Eligible R's	Don't Know /Refused (n)	Don't Know/ Refused %	Reported				Don't know/Refused Responses Removed		Don't know/Refused Responses Deemed Non-Free riders	
					FR	P SO	C SO	NTG Ratio	FR	NTG Ratio	FR	NTG Ratio
Dehumidifiers	**T1** Without the <rebates> from <Sponsor>, how likely is it that you would have installed any type of <equipment> at the same time?	70	1	1%	42%	1%	n/a	59%	43%	58%	40%	61%
	T2 Without the <rebates> from <Sponsor>, when do you think you would have installed the <equipment>?	27	6	22%								

	Questions	Number of Eligible R's	Don't Know /Refused (n)	Don't Know/ Refused %	Reported				Don't know/Refused Responses Removed		Don't know/Refused Responses Deemed Non-Free riders	
					FR	P SO	C SO	NTG Ratio	FR	NTG Ratio	FR	NTG Ratio
Leave Behind Advanced Power Strips	**T1** Without the <rebates> from <Sponsor>, how likely is it that you would have installed any type of <equipment> at the same time?	8	0	0%	9%	0%	n/a	91%	9%	91%	8%	92%
	T2 Without the <rebates> from <Sponsor>, when do you think you would have installed the <equipment>?	5	1	20%								
Purchased Advanced Power Strips	**T1** Without the <rebates> from <Sponsor>, how likely is it that you would have installed any type of <equipment> at the same time?	72	2	3%	15%	4%	n/a	89%	16%	88%	11%	93%
	T2 Without the <rebates> from <Sponsor>, when do you think you would have installed the <equipment>?	60	22	37%								

R&A NTG study

In the R&A study, timing scoring questions were asked about all measures except for duct or air sealing. They were not asked for HVAC or domestic hot water measures if respondents indicated the equipment needed major repair or was no longer working. Of those that received timing questions, there was only one response of "don't know" or "refused," which had an effect that was too small to show in the timing free-ridership score when either removed or treated as a non-free rider.

Table 15: Timing Free-ridership with "Don't know" and "Refused" Excluded and Deemed as Non-Free Riders

Questions	Number of Eligible R's	Don't Know /Refused (n)	Don't Know/ Refused %	Reported				Don't know/Refused Responses Removed		Don't know/Refused Responses Deemed Non-Free riders	
				FR	P SO	C SO	NTG Ratio	FR	NTG Ratio	FR	NTG Ratio
T1 Without the <rebates> from <Sponsor>, how likely is it that you would have installed any type of <equipment> at the same time?	416	0	0%	34%	2%	17%	85%	34%	85%	34%	85%
T2 Without the <rebates> from <Sponsor>, when do you think you would have installed the <equipment>?	123	1	14%								

3.1.2.1.2.2 *Evaluation Team Questions*

There were no questions about timing scores from the evaluation teams.

3.1.2.1.3 Quantity Score

The quantity score is the least populated score of the three intent scores. It is not asked where quantity is one or missing, such as a furnace, boiler, tune-up, water heater, or thermostat installations, of which it is unlikely that more than one would be installed. Table 16 shows the quantity score questions defined in the methodology.

Table 16: Quantity Questions

Quantity (Evaluator: Ask only if quantity purchased/received is greater than 1; for example, you would not need to ask if a customer purchased one furnace/boiler or received a tune-up. For insulation/weatherization type measures ask this series, but the wording should be revised to "amount of" instead of "quantity of.")			
Q1	Without the <list of all rebates/incentives/support received> from <Sponsor>, how likely is it that you would have installed the exact same [quantity of / amount of] high efficiency <measure>? Would you say it is …		
	01	Not at all likely	FR_qty = 0
	02	Slightly likely	
	03	Somewhat likely	
	04	Very likely	FR_qty = 1.0
	88	Don't know	FR_qty = 0.5
	99	Refused	FR_qty = 0.5
Q2a	[non-insulation/weatherization type measures] Without the rebate or assistance from <Sponsor>, how many high efficiency <measure> would you have installed?		
	__	Quantity would have installed	FR_qty = (Q2a/installed quantity)
	88	Don't know	FR_qty = 0.5
	99	Refused	FR_qty = 0.5
Q2b	[insulation/weatherization type measures] Without the rebate or assistance from <Sponsor>, what percent of the high efficiency <measure> would you have installed?		
	__	Percent would have installed [0-100]	FR_qty = Q2b
	88	Don't know	FR_qty = 0.5
	99	Refused	FR_qty = 0.5

3.1.2.1.3.1 *Summary of Results*

RCD and select products NTG study

The quantity questions are asked for this study only when the measure quantity is greater than one. 20 of 132 respondents, or 15%, answered "don't know" or "refused" to the quantity questions, resulting in the imputation of a 0.5 free-ridership score. Weighted against all respondents, the resulting free-ridership change when excluding these responses or deeming them as non-free riders was an NTG ratio increase of 1%. The difference between the two approaches was small (0.13%).

Table 17: Quantity Free-ridership with "Don't know" and "Refused" Excluded and Deemed as Non-Free Riders*

Questions	Number of Eligible R's	Don't Know /Refused (n)	Don't Know/ Refused %	Reported				Don't know/Refused Responses Removed		Don't know/Refused Responses Deemed Non-Free riders	
				FR	P SO	C SO	NTG Ratio	FR	NTG Ratio	FR	NTG Ratio
Q1 Without the <list of all rebates/incentives/support received> from <Sponsor>, how likely is it that you would have installed the exact same [quantity of / amount of] high efficiency <measure>?	132	3	2%								
Q2a [non-insulation/weatherization type measures] Without the rebate or assistance from <Sponsor>, how many high efficiency <measure> would you have installed?	86	13	15%	31%	12%	4%	85%	30%	86%	29%	86%
Q2b [insulation/weatherization type measures] Without the rebate or assistance from <Sponsor>, what percent of the high efficiency <measure> would you have installed?	46	4	9%								

*The free-ridership difference between removing "don't know" and "refused" responses and deeming them as a non-free riders was 0.13%. Rounding resulted in a 30% and 29% free-ridership score for each.

Products NTG study

There were no "don't know" or "refused" responses for the quantity free-ridership scoring questions for dehumidifiers and leave-behind advanced power strips. The number of eligible respondents for these two measures was also very small, substantially less for purchased advanced power strips (n=40). Of these, only one respondent provided an answer of "don't know" or "refused." This had little to no effect on the free-ridership score when removed or treated as a non-free rider.

Table 18: Quantity Free-ridership with "Don't know" and "Refused" Excluded and Deemed as Non-Free Riders

	Questions	Number of Eligible R's	Don't Know /Refused (n)	Don't Know/ Refused %	Reported				Don't know/Refused Responses Removed		Don't know/Refused Responses Deemed Non-Free riders	
					FR	P SO	C SO	NTG Ratio	FR	NTG Ratio	FR	NTG Ratio
Dehumidifiers	**Q1** Without the <list of all rebates/incentives/support received> from <Sponsor>, how likely is it that you would have installed the exact same [quantity of / amount of] high efficiency <measure>?	3 (only asked if qty>1)	0	0%								
	Q2a [non-insulation/ weatherization type measures] Without the rebate or assistance from <Sponsor>, how many high efficiency <measure> would you have installed?	0	0	0%	42%	1%	n/a	59%		n/a		n/a
	Q2b [insulation/weatherization type measures] Without the rebate or assistance from <Sponsor>, what percent of the high efficiency <measure> would you have installed?	0	0	0%								
Leave Behind Advanced Power Strips	**Q1** Without the <list of all rebates/ incentives/ support received> from <Sponsor>, how likely is it that you would have installed the exact same [quantity of / amount of] high efficiency <measure>?	8	0	0%								
	Q2a [non-insulation/weatherization type measures] Without the rebate or assistance from <Sponsor>, how many high efficiency <measure> would you have installed?	3	0	0%	9%	0%	n/a	91%		n/a		n/a
	Q2b [insulation/weatherization type measures] Without the rebate or assistance from <Sponsor>, what percent of the high efficiency <measure> would you have installed?	n/a	n/a	n/a								

	Questions	Number of Eligible R's	Don't Know /Refused (n)	Don't Know/ Refused %	Reported				Don't know/Refused Responses Removed		Don't know/Refused Responses Deemed Non-Free riders	
					FR	P SO	C SO	NTG Ratio	FR	NTG Ratio	FR	NTG Ratio
Purchased Advanced Power Strips	**Q1** Without the <list of all rebates/incentives/support received> from <Sponsor>, how likely is it that you would have installed the exact same [quantity of / amount of] high efficiency <measure>?	40	1	3%								
	Q2a [non-insulation/ weatherization type measures] Without the rebate or assistance from <Sponsor>, how many high efficiency <measure> would you have installed?	40	0	0%	15%	4%	n/a	89%	16%	88%	15%	89%
	Q2b [insulation/ weatherization type measures] Without the rebate or assistance from <Sponsor>, what percent of the high efficiency <measure> would you have installed?	n/a	n/a	n/a								

R&A NTG study

Quantity scoring is not included in this study. Intent free-ridership is calculated only using timing and efficiency.

3.1.2.1.3.2 *Evaluation Team Questions*

There were no questions about quantity scores from the evaluation teams.

3.1.2.1.4 Efficiency Score

The efficiency scoring question is not asked for all measures or all studies. Efficiency is not applicable for tune-ups, insulation, weatherization, air sealing, or direct install measures. If the efficiency scoring question responses are answered with "don't know" or "refused," the free-ridership scoring is set to 0.5. Table 19 shows the efficiency scoring question.

Table 19: Efficiency Scoring Question

Efficiency (Evaluator: Ask only if efficiency applies; for tune-up, insulation/weatherization, or air sealing type measures, etc., skip to 'influence' questions)			
E1	Without the <list of all rebates/incentives/support received> from <Sponsor>, how likely is it that you would have installed the exact same high-efficiency level of <measure>? Would you say …		
	01	Not at all likely	FR_eff = 0
	02	Slightly likely	FR_eff = 0.25
	03	Somewhat likely	FR_eff = 0.5
	04	Very likely	FR_eff = 1.0
	88	Don't know	FR_eff = 0.5
	99	Refused	FR_eff = 0.5

3.1.2.1.4.1 *Summary of Results*

RCD and select products NTG study

Seven percent of respondents said "don't know" or "refused" to the free-ridership efficiency question (24 of 359), which resulted in the imputation of a free-ridership score of 0.5. Weatherization measures did not receive the efficiency question. Across the other measures, the number of "don't know" and "refused" responses varied little. Table 20 below shows the reported NTG ratio and the NTG ratios when "don't know" and "refused" responses are removed and deemed non-free riders. Excluding these 24 responses increased NTG by 1% and deeming them non-free riders increased the NTG ratio by 3%, from 85% to 88%.

Table 20. Efficiency Free-ridership with "Don't know" and "Refused" Excluded and Deemed as Non-Free Riders

Questions	Number of Eligible R's	Don't Know /Refused (n)	Don't Know/ Refused %	Reported				Don't know/Refused Responses Removed		Don't know/Refused Responses Deemed Non-Free riders	
				FR	P SO	C SO	NTG Ratio	FR	NTG Ratio	FR	NTG Ratio
E1 Without the <list of all rebates/incentives/support received> from <Sponsor>, how likely is it that you would have installed the exact same high efficiency level of <measure>?	359	24	7%	31	12%	4%	85%	1%	86%	3%	88%

Products NTG study

The efficiency scoring is not applicable for advanced power strips. For dehumidifiers, the efficiency question is only applicable to respondents who also recycled and recalled doing so at an event. Since there was only one respondent who both rebated and recycled a dehumidifier, no alternative free-ridership scoring could be calculated.

Table 21: Efficiency Free-ridership with "Don't know" and "Refused" Excluded and Deemed as Non-Free Riders

Questions	Number of Eligible R's	Don't Know /Refused (n)	Don't Know/ Refused %	Reported				Don't know/Refused Responses Removed		Don't know/Refused Responses Deemed Non-Free riders	
				FR	P SO	C SO	NTG Ratio	FR	NTG Ratio	FR	NTG Ratio
E1 Without the <list of all rebates/incentives/support received> from <Sponsor>, how likely is it that you would have installed the exact same high efficiency level of <measure>?	1	0	0%	31	12%	4%	85%	n/a		n/a	

R&A NTG study

Three percent of R&A NTG study respondents answered "don't know" or "refused" to the efficiency question. Removing responses of "don't know" and "refused" from the efficiency scoring did not significantly impact their free-ridership score. Due to rounding, deeming these responses as non-free riders increased the NTG ratio by 1%.

Table 22: Efficiency Free-ridership with "Don't know" and "Refused" Excluded and Deemed as Non-Free Riders*

Questions	Number of Eligible R's	Don't Know /Refused (n)	Don't Know/ Refused %	Reported				Don't know/Refused Responses Removed		Don't know/Refused Responses Deemed Non-Free riders*	
				FR	P SO	C SO	NTG Ratio	FR	NTG Ratio	FR	NTG Ratio
E1 Without the <list of all rebates/incentives/support received> from <Sponsor>, how likely is it that you would have installed the exact same high efficiency level of <measure>?	468	15	3%	34%	2%	17%	85%	34%	85%	34%	86%

*The free-ridership difference between imputing "don't know" and "refused" responses as 0.5 and deeming them as non-free riders was 0.007%. Rounding resulted in an 85% and 86% NTG ratio for each.

3.1.2.1.4.2 *Evaluation Team Questions*

Question: The efficiency question may need to be modified if the technology type varies by situation. For example, in the case of a thermostat replacement, the FR would depend on the type of thermostat being replaced.

Response: Rather than focus on what the thermostat replaced, ask what the respondent would have done without the incentive, then generate an efficiency score based on choice options.

Resolution: Modified the methodology to ask: *"Without the rebate or assistance from <Sponsor> for your thermostat, what type of a thermostat would you have likely installed?*

01 Would not have installed any thermostat FR_eff = 0.0
02 A non-programmable thermostat (you turn it on and off and change temperature by hand) FR_eff = 0.0
03 A programmable thermostat (you program temperature settings that change automatically at different times of the day, but with no remote control) FR_eff = 0.25
04 A wi-fi programmable thermostat (you can program and control it remotely with your smartphone) FR_eff = 0.5
05 A wi-fi smart thermostat (you can control it remotely with your smartphone but it also learns your schedule and preferences and adjusts the temperature based on those) FR_eff = 1.0
88 Don't know FR_eff = 0.5
99 Refused FR_eff = 0.5

3.1.2.2 Influence Score

The algorithm uses the maximum-rated score for program attributes to create an influence score. Table 23 shows the influence questions within "Consistent Methodology."

Table 23: Influence Questions

Influence: (The evaluator should review logic models and program theory, and conduct staff interviews to inform the list of elements to include)	
X1	On a scale of 0 to 10 where 0 is 'not at all influential' and 10 is 'very influential,' how influential was < attribute from list below> on your decision to install the high efficiency <measure>? [NOTE: program features asked about should be tailored to the program being evaluated; list max of 6 factors] a. the program rebate b. the program's marketing materials c. [Evaluation team will need to insert other features of the program design/delivery] __ Rating 66 Not applicable 88 Don't know 99 Refused
X2	[ASK IF PROGRAM THEORY SHOWS CONTRACTOR/SALESPERSON PLAY A PROMINENT ROLE IN DELIVERING AND PROMOTING THE PROGRAM] Using the same 0 to 10 scale where 0 = 'not at all influential' and 10 is 'very influential,' how influential was the contractor or salesperson on your decision to install the high efficiency <measure>? __ Rating 66 Not applicable 88 Don't know

	99 Refused
X3	[ASK IF CONTRACTOR/SALESPERSON HAS THE HIGHEST INFLUENCE SCORE AND PROGRAM THEORY SHOWS CONTRACTOR/SALESPERSON PLAY A PROMINENT ROLE IN DELIVERING AND PROMOTING THE PROGRAM] Which of the following best describes how you selected the new high efficiency <measure>? (Select one) [rotate options 1 through 3] 01 I did some research on <measure>s and made my own choice 02 My contractor suggested one <measure> efficiency level, and I agreed – (use Contractor FR results) 03 My contractor suggested various <measure> efficiency levels, and I chose one (mean of contractor and customer FR results) 04 Something else (specify)

There will always be a maximum, minimum, or mean unless a participant says "don't know" or "refused" to all attributes. In none of the studies did a respondent answer "don't know" or "refused" to all program attributes, so analysis of the effect of these "don't know" or "refused" responses was not applicable. For comparison purposes, Table 24 shows the overall free-ridership score if the mean or minimum is used instead of the maximum. Using the mean will inherently increase free-ridership, even more so using the minimum. It is difficult to assess the impact of using the mean influence because not all program aspects apply to all measures. Some measures have values across five or six program aspects while others may have only two or three; in the latter case, much greater weight is given to those few responses. The influence score only aims to determine if any aspect of the program affected the participant's decision-making to install what they did and when. While we show results using the min value below, which greatly increases free-ridership, using the min value would not accurately represent the program's effects.

Table 24: Free-ridership Using Maximum vs. Mean and Minimum Influence Score

Study	Measure Count	FR Using Maximum Influence (Reported)	FR Using Mean Influence	FR Using Min Influence
RCD and select products NTG study	623	31%	38%	47%
Products NTG study - Dehumidifiers	70	42%	58%	71%
Products NTG study - Leave Behind APS*	44	9%	9%	9%
Products NTG study - Purchased APS	72	15%	28%	41%
R&A NTG study	468	34%	58%	62%

*Influence questions are not asked for leave behind APS.

3.1.2.2.1 Evaluation Team Questions

There were no questions about influence scoring from the evaluation teams.

3.2 PARTICIPANT SPILLOVER SCORING METHODOLOGY ANALYSIS

Participant spillover can only be calculated if all scoring questions are answered. Responses of "don't know" and "refused" to key spillover questions resulted in no additional savings for the measure. Because of this, analysis is limited to frequency checks of "don't know" and "refused" responses. Figure 5 shows the participant spillover calculation, and Table 25 shows the participant spillover questions. Across the three studies, responses of "don't know" and "refused" were minimal and not identified as problematic.

Figure 5: Spillover Calculation

Participant Spillover score = (S4 + S5) / 2

Table 25: Participant Spillover Questions

Participant Spillover[6]	
S1	Since installing the equipment, have you made other energy-saving purchases or changes that did not receive a rebate through the Mass Save® program? (Select one) 01 Yes 02 No (skip to next section) 88 Don't know (skip to next section)
S2	Did your experience with the Mass Save program influence your decision to make these energy-saving purchases or changes? (Select one) 01 Yes 02 No (skip to next section) 88 Don't know (skip to next section)
S3	What energy-saving purchases or changes did you make? (Select all that apply) [Note: Make light bulbs and other upstream products available for interviewers and respondents, but these measures will not be counted towards spillover] 01 LEDs/Energy-efficient light bulbs 02 Energy-efficient appliance(s) (specify type) 03 HVAC equipment (specify type) 04 Attic, wall, or basement insulation 05 Duct sealing or duct insulation 06 Air sealing of leaks 07 Smart thermostat 08 Water heater 09 Other purchases or changes (specify)

[6] NOTE: To be able to calculate spillover, at a minimum, the survey demographics should collect information on water heater fuel, clothes dryer fuel, primary heating fuel, and type of cooling equipment in order to determine what prescriptive savings to apply.

[ASK S3b to S6 FOR EACH MEASURE OTHER THAN LIGHTING[7] SELECTED IN S3]

S3b [if S3 = 2 or 3 or 8 or 9] Can you describe the <equipment from S3> work in more detail? If applicable, include the quantity installed, the type of equipment, and the efficiency levels installed. (For AC probe on SEER rating, for furnaces or boilers probe on AFUE, for heat pumps probe on Heating season performance factor [HSPF] or SEER)

S3c How did you know the equipment was energy efficient and would have qualified for a rebate through the program?

[Record verbatim response]

Possible response categories for a web survey:

Efficiency rating or label of equipment, such as an "ENERGY STAR®" logo
Equipment dealer/retailer said it was energy efficient
Personal experience
Met utility rebate requirements
Third party report, such as Consumer Reports
Recommendations from the contractor/installer
Did not rely on any specific type of information
Internet/website
Other (specify)
Don't know

S3a [If S3=2 or 8 and ENERGY STAR not mentioned in S3b] Was the <equipment from S3> ENERGY STAR-labeled? (Select one)

01	Yes
02	No
88	Don't know

S4 How important was your experience with the Mass Save program on your decision to install <equipment from S3> that did not receive a rebate from <Sponsor>? Was it ...

01	Not at all important	score = 0
02	Slightly important	score = 3
03	Somewhat important	score = 6
04	Very important	score = 10
66	Not applicable	
88	Don't know	
99	Refused	

[7] LEDs are assumed to be counted as part of the upstream program. If the upstream program is ended, then they should be counted towards spillover.

S5	How likely is it that you would still have purchased/installed energy-efficient <equipment from <S3> if you had *not* already received a rebate for the energy-efficient <measure>? Would you have been ... to have installed the energy-efficient <equipment from S3>? 01 Not at all likely score = 10 02 Slightly likely score = 6 03 Somewhat likely score = 3 04 Very likely score = 0 66 Not applicable 88 Don't know 99 Refused
S6	**(Consistency question)** Why did you not submit an application for a rebate through the Mass Save program for this additional equipment? [record verbatim response] [Note: for respondents who indicate that the equipment did not qualify, no spillover should be attributed to the program]

3.2.1.1 Summary of Results

RCD and select products NTG study

As Table 26 shows below, less than 10% of eligible respondents recalled having made other energy-saving purchases that did not receive a rebate. All instances of "don't know" and "refused" responses were low across the participant spillover questions, suggesting respondents did not have recall issues. There were minimal missed opportunities for capturing additional energy savings.

Table 26: Don't know/Refused Response Count for Participant Spillover Questions

Questions	Number of Eligible Respondents	Don't Know/ Refused (n)	Don't know/ Refused %
S1 Since installing the equipment, have you made other energy-saving purchases or changes that did not receive a rebate through the Mass Save® program?	508	41	8%
S2 Did your experience with the Mass Save program influence your decision to make these energy-saving purchases or changes?	155	8	5%
S3 What energy-saving purchases or changes did you make?	84	0	0%
S3b [if S3 = 2 or 3 or 8 or 9] Can you describe the <equipment from S3> work in more detail?	84	open end	0%
S3c How did you know the equipment was energy efficient and would have qualified for a rebate through the program?	84	5	6%

Questions	Number of Eligible Respondents	Don't Know/ Refused (n)	Don't know/ Refused %
S3a [If S3=2 or 8 and ENERGY STAR not mentioned in S3b] Was the <equipment from S3> ENERGY STAR-labeled?	84	2	2%
S4 How important was your experience with the Mass Save program on your decision to install <equipment from S3> that did not receive a rebate from <Sponsor>?	83	1	1%
S5 How likely is it that you would still have purchased/installed energy-efficient <equipment from <S3> if you had not already received a rebate for the energy-efficient <measure>?	83	3	4%
S6 (Consistency question) Why did you not submit an application for a rebate through the Mass Save program for this additional equipment?	83	open end	0%

Products NTG study

Thirteen percent of dehumidifier respondents and 3% of advanced power strip respondents could not recall if they had made other energy-saving purchases that did not receive a rebate. Other responses of "don't know" and "refused" were very low for all other participant spillover questions, meaning respondents did not have recall issues. There were minimal missed opportunities for capturing additional energy savings.

Table 27: Don't know/Refused Response Count for Participant Spillover Questions for Dehumidifiers and Advanced Power Strips

	Questions	Number of Eligible Respondents	Don't Know/ Refused (n)	Don't Know/ Refused %
Dehumidifier	**Participant Spillover**			
	S1 Since installing the equipment, have you made other energy-saving purchases or changes that did not receive a rebate through the Mass Save® program?	70	9	13%
	S2 Did your experience with the Mass Save program influence your decision to make these energy-saving purchases or changes?	21	0	0%
	S3 What energy-saving purchases or changes did you make?	8	0	0%
	S3b [if S3 = 2 or 3 or 8 or 9] Can you describe the <equipment from S3> work in more detail?	4	0	0%
	S3c How did you know the equipment was energy efficient and would have qualified for a rebate through the program?	8	1	13%
	S3a [If S3=2 or 8 and ENERGY STAR not mentioned in S3b] Was the <equipment from S3> ENERGY STAR-labeled?	n/a	n/a	n/a

	Questions	Number of Eligible Respondents	Don't Know/ Refused (n)	Don't Know/ Refused %
	S4 How important was your experience with the Mass Save program on your decision to install <equipment from S3> that did not receive a rebate from <Sponsor>?	8	0	0%
	S5 How likely is it that you would still have purchased/installed energy-efficient <equipment from <S3> if you had not already received a rebate for the energy-efficient <measure>?	8	1	13%
	S6 (Consistency question) Why did you not submit an application for a rebate through the Mass Save program for this additional equipment?	8	0	0%
Advanced Power Strips	**Participant Spillover**			
	S1 Since installing the equipment, have you made other energy-saving purchases or changes that did not receive a rebate through the Mass Save® program?	72	2	3%
	S2 Did your experience with the Mass Save program influence your decision to make these energy-saving purchases or changes?	24	0	0%
	S3 What energy-saving purchases or changes did you make?	13	0	0%
	S3b [if S3 = 2 or 3 or 8 or 9] Can you describe the <equipment from S3> work in more detail?	4	0	0%
	S3c How did you know the equipment was energy efficient and would have qualified for a rebate through the program?	13	0	0%
	S3a [If S3=2 or 8 and ENERGY STAR not mentioned in S3b] Was the <equipment from S3> ENERGY STAR-labeled?	not asked	not asked	
	S4 How important was your experience with the Mass Save program on your decision to install <equipment from S3> that did not receive a rebate from <Sponsor>?	13	1	8%
	S5 How likely is it that you would still have purchased/installed energy-efficient <equipment from <S3> if you had not already received a rebate for the energy-efficient <measure>?	13	0	0%
	S6 (Consistency question) Why did you not submit an application for a rebate through the Mass Save program for this additional equipment?	115	open end	0%

R&A NTG study

As with the other two studies, beyond the initial recall of having made other energy-saving purchases that did not receive a rebate (11%), responses of “don’t know” and “refused” were minimal. Table 28 shows the breakdown of “don’t know” and “refused” responses to the participant spillover questions.

Table 28: Don't know/Refused Response Count for Participant Spillover Questions for the R&A NTG Study

Questions	Number of Eligible Respondents	Don't Know/ Refused (n)	Don't Know/ Refused %
S1 Since installing the equipment, have you made other energy-saving purchases or changes that did not receive a rebate through the Mass Save® program?	120	13	11%
S2 Did your experience with the Mass Save program influence your decision to make these energy-saving purchases or changes?	38	3	8%
S3 What energy-saving purchases or changes did you make?	26	3	12%
S3b [if S3 = 2 or 3 or 8 or 9] Can you describe the <equipment from S3> work in more detail?	7	0	0%
S3c How did you know the equipment was energy efficient and would have qualified for a rebate through the program?	NOT ASKED		
S3a [If S3=2 or 8 and ENERGY STAR not mentioned in S3b] Was the <equipment from S3> ENERGY STAR-labeled?	NOT ASKED		
S4 How important was your experience with the Mass Save program on your decision to install <equipment from S3> that did not receive a rebate from <Sponsor>?	26	3	12%
S5 How likely is it that you would still have purchased/installed energy-efficient <equipment from <S3> if you had not already received a rebate for the energy-efficient <measure>?	26	0	0%
S6 (Consistency question) Why did you not submit an application for a rebate through the Mass Save program for this additional equipment?	26	1	4%

3.2.1.2 Evaluation Team Questions

Question: We are looking only at electricity measures, yet respondents named measures that save space or water heating fuels. We don't have that heating fuel information. This leads to two questions:

1. What should we assume would be the heating fuels?
2. Do we convert all the fuels to kWh or MMBtu or other consistent unit for calculating SO?

Response: The *Consistent Methodology for Self-Reported Residential Net-to-Gross Measurement* study guides how to calculate the spillover influence score based on the participant's answers to questions about how influential their rebate was on the decision to install additional equipment outside of the program. However, the guidance does not delineate how to calculate the actual spillover rate used in the NTG formula. The evaluation team determined the best approach to calculate program level spillover would be by summing deemed savings for all spillover measures, converting to MMBtu, and dividing spillover savings by program savings, as Section 2.4.2 notes.

Resolution: Added to the methodology, *"To calculate spillover, at a minimum, the survey demographics should collect information on water heater fuel, clothes dryer fuel, primary heating*

fuel, and type of cooling equipment to determine what prescriptive savings to apply." To calculate program level spillover, sum the deemed savings for all spillover measures, converting to MMBtu, and divide the spillover savings by program savings

Question: For the web survey, do we offer choice options for respondents rather than ask the following as an open-end, *"How did you know the equipment was energy efficient and would have qualified for a rebate through the program?"*

Response: There is no recommended change to the methodology; however, if choice options are desired, those options would need to be available to all modes. Here are some choice options.

1. Efficiency rating or label of equipment, such as an ENERGY STAR® label or logo
2. Equipment dealer/retailer said it was energy efficient
3. Personal experience
4. Met utility rebate requirements
5. Third-party report, such as Consumer Reports
6. Recommendations from the contractor/installer
7. Did not rely on any specific type of information
8. Internet/website
9. Other (specify)
10. Don't know

Resolution: No methodology change. Added response categories to the methodology as optional guidance.

Question: Is spillover calculated separately for each measure so that each of the measures listed has a spillover score?

Response: It is calculated separately for each measure. The spillover rate equals the sum of the savings associated with the additional TRM (or other secondary source) spillover measure divided by the sum of the program savings. Savings should be converted to MMBtus to allow for spillover for both gas and electric measures.

Resolution: Added to the methodology, text clarification, ***"****A spillover score is calculated for each measure. If the score is 5 or lower, spillover is set to 0. If the score is higher than 5, the spillover savings is the appropriate TRM (or other secondary source) savings."*

Also added the text, *"The spillover rate is equal to the sum of the savings associated with the additional TRM (or other secondary source) spillover measure divided by the sum of the program savings. Savings should be converted to MMBtus to allow for spillover for both gas and electric measures."*

Question: Is the spillover calculation "*Participant Spillover score = (S4 + (10-S5)) / 2*" working correctly? We ask S4 and S5 as follows:

S4 How important was your experience with the Mass Save program on your decision to install <equipment from S3> that did not receive a rebate from <Sponsor>? Was it ...

01	Not at all important	score = 0
02	Slightly important	score = 3
03	Somewhat important	score = 6
04	Very important	score = 10
66	Not applicable	
88	Don't know	
99	Refused	

S5 How likely is it that you would still have purchased/installed energy-efficient <equipment from <S3> if you had *not* already received a rebate for the energy-efficient <measure>? Would you have been ... to have installed the energy-efficient <equipment from S3>?

01	Not at all likely	score = 10
02	Slightly likely	score = 6
03	Somewhat likely	score = 3
04	Very likely	score = 0
66	Not applicable	
88	Don't know	
99	Refused	

In the calculation equation, since we've already swapped the S5 scale to account for its "backwardness" why do we do the 10-S5? Doesn't that re-swap the scale? So, if a person said "not at all important" to S4 and "not at all likely" to S5, their scores would be 0 and 10, respectively. In the equation I'd have 0 + (10 – 10) / 2, so spillover would be zero. But the person just told us that they would be "not at all likely" to have installed the measure without the prior rebate. It seems like it should be more than zero. The same happens on the other extreme side of the scale: A person says 4 to both, so their scores become 10 and 0, with the equation being 10 + (10 – 0) /2 = 10. This indicates full spillover, even though they just said that they were very likely to install without the prior rebate. The midpoints of the scale make a little more sense, but I'm confused at the extremes, which makes me wonder if the midpoints are wrong.

Response: You are correct that including (10-S5) in the calculation incorrectly calculates the spillover score.

Resolution: Updated the spillover calculation to read "*Participant Spillover score = (S4 + S5)) / 2.*"

3.3 Inconsistent Response Methodology Analysis

The evaluation team assessed the frequency of inconsistent responses between intent and influence. An inconsistency is identified when (1) the respondent said the program did not influence their timing or intent to install the rebated equipment but rated the influence of a program factor as high (greater than 6 on a 0 to 10 scale) or (2) the respondent said the program affected their timing and intent to install the rebated equipment when they did but rated the influence across all program factors low (less than 4 on a 0 to 10 scale).

When responses are flagged as inconsistent, two follow-up questions are asked (Table 29), and a manual review is conducted to determine a resolution. The process of the manual review is subjective and relies upon experienced analysts to determine if the respondent's open-ended response to the influence of the program supports the existing free-ridership score. If there are any indications supporting the free-ridership score, the participant's score should be kept. If their response strongly contradicts the free-ridership results, the case is excluded. Only the RCD and select products NTG study included the programmed consistency check questions. The evaluation team found that less than 1% (5 of 623 respondents) received those questions for this study. All were able to be reconciled, and free-ridership was calculated.

Table 29: Consistency Questions

Consistency	
C1	**(CONSISTENCY QUESTION)** Did you make the decision to install the high efficiency <measure> before or after you learned of the rebate? (select one) 01 Before 02 After 88 Don't know 99 Refused
C2	**(CONSISTENCY QUESTION)** Please describe the influence that the program <list of all rebates/incentives/support received) had on your decision to install the high efficiency <measure>? [Record verbatim response]

A consistency check was computed for all three studies to look at intent and influence discrepancies between the program's impact on intent and influence. To do this, the evaluation team created an inconsistent data check. The following summary results break down the frequency of each check for each program. The values include:

- 0 – No inconsistency was identified
- 1 – The respondent said the program had no impact on their timing, quantity, or efficiency, but they rated a program attribute high (intent=0/influence>6)
- 2 – The respondent said the program did impact their decision to install the measure when they did, or the quantity, or the efficiency, but said program attributes did not influence their decision (intent=1/influence<4).

Most participant responses showed no inconsistency between intent and influence across all three studies. The most significant inconsistency was within the R&A program NTG study, where 26% of respondents indicated that the program had no impact on their timing, quantity, or efficiency decision-making but rated the influence of at least one program aspect as high. Because there were no programmed consistency checks within the survey, these inconsistencies were not reviewed, and no changes were made to their free-ridership.

Table 30: Frequency of Inconsistent Responses by Program

Questions	RCD and Select Products NTG Study	Products NTG Study - Dehumidifiers	Products NTG Study - Purchased APS*	R&A NTG Study
Not inconsistent	559	61	70	346
No intent impact, high influence	0	8	0	122
High intent impact, low influence	64	1	2	0
Total	623	70	72	468

*Because no influence scoring is calculated for leave behind APS, no inconsistencies were identified.

3.3.1.1 Summary of Results

RCD and select products NTG study

As noted above, the RCD and select products NTG study was the only one of the studies that had programmed consistency questions into the survey. Most responses (89%) were deemed consistent. Less than 1% said the program had no impact on their timing, quantity, or efficiency selection but rated the influence of a program attribute high (>6 rating). One-tenth of respondents said the program did impact their timing, quantity, or efficiency selection but rated all applicable program attributes low (<4 rating). The open-ended responses to the influence of the program supported the free-ridership results and no cases were excluded.

Table 31: Frequency of Consistency Check Questions

Inconsistency	Number of Eligible Respondents	Percent
Not inconsistent	559	89%
No intent impact, high influence	0	0%
High intent impact, low influence	64	11%
Total	**623**	**100%**

Products NTG study

The survey did not contain consistency check questions.

R&A NTG study

The survey did not contain consistency check questions.

3.3.1.2 Evaluation Team Questions

There were no consistency check questions from the evaluation teams.

3.4 CONTRACTOR FREE-RIDERSHIP METHODOLOGY ANALYSIS

The following contains an analysis of contractor free-ridership and the frequency with which it is used either in place of or averaged with the participant's free-ridership score. Contractor free-

ridership is calculated using the total weighted average free-ridership for the applicable measure across all contractors. The analysis found that contractor free-ridership is higher than participant free-ridership and is used in place of participant free-ridership 44% of the time in the RCD and select products NTG study and the 9% in the R&A NTG study. The following analysis is for the RCD and select products NTG study and the R&A NTG study. Contractors are not applicable in the products NTG study.

3.4.1 Contractor Free-ridership Score

Contractor free-ridership is determined by assessing the contractor's estimated percentage of measure sales had the program not existed. That percentage equals the free-ridership score. The contractor may also provide a direct estimate of measure sales had the program not existed. In that case, the contractor estimated sales had the program not existed divided by program sales of the measure is used to calculate free-ridership. Figure 6 below shows the contractor free-ridership calculation, and Table 32 shows the free-ridership battery.

Figure 6: Contractor Free-ridership Calculation

Free-ridership = CF1 or CF3/program sales

Table 32: Contractor Free-ridership Questions

Contractor Free-ridership	
CF0	For the next set of questions, we are going to ask you to think about your company's experiences with Mass Save rebates. Our records show that in <year> your company sold <QTY> residential high-efficiency <measure> that received rebates from Mass Save. Does this sound right? 01 Yes [program sales = QTY] skip to CF1 02 No 88 Don't know [program sales = QTY] skip to CF1
CF0a	About how many residential <measures> that received rebates from Mass Save did your company sell in <year>? ____ [0 TO 1000] [program sales = CF0a] 8888 Don't know [skip to next section] 9999 Refused [skip to next section]
CF1	If Mass Save had not offered any rebates in <year>, about what percentage of the <QTY> high efficiency <measure> that received rebates would you have sold that year? [Interviewer note: If the contractor only sold 1, then the response should be 0% if said would not have sold any, or 100% if said would have sold the same amount] ___ [0 TO 100] 888 Don't know [skip to next section] 999 Refused [skip to next section]
CF2	[IF CF1 < = 100] To confirm, you are estimating that in <year>, your company likely would still have installed <QTY * (CF1 / 100)> <measure> of the <QTY> you sold if Mass Save had not offered the rebates. Is that **roughly** correct? 01 Yes [skip to next section] free-ridership = CF1 02 No 88 Don't know [skip to next section] 99 Refused [skip to next section]
CF3	[IF CF2 = 2] Could you provide your best estimate of about how many of the <QTY> <measure> your company would still have sold in <year> if Mass Save had not offered the rebates? ___ [0 TO 1000] 8888 (Don't know)

3.4.2 Application of Contractor Free-ridership

The participant survey assessed program influence scores to determine who was most influential in their decision to install the energy-efficiency equipment and when to apply the respective free-ridership score. If their max influence score across program attributes (E1) is less than their contractor influence rating (E2), the survey (E3) asks the respondent to describe how they selected the new high-efficiency measure. If they said they made their own choice, the participant's free-ridership score is used. If the contractor suggested the equipment, the

contractor's free-ridership score is used instead. If they say it was a combination of both factors, the mean is used between the customer and contractor free-ridership score.

Table 33 shows the participant survey results and what scores are used for free-ridership. Almost half (44%) of participants said the contractor suggested the equipment, meaning the free-ridership score was applied. One quarter said they made their own choice, and another quarter said it was a combination.

Table 33: Participant and Contractor Free-ridership Scoring Assignment

Study	FR Assignment	RCD and select products NTG study		R&A NTG study	
		Measure Count	Percent	Measure Count	Percent
I did some research and made my own choice	Participant FR	156	25%	133	28%
My contractor suggested efficiency level, and I agreed	Contractor FR	272	44%	176	38%
My contractor suggested various efficiency levels, and I chose one	Average of Participant/Contractor FR	162	26%	130	28%
Something else	Review open end	33	5%	29	6%
Total		**623**	**100%**	**468**	**100%**

3.4.2.1 Summary of Results

Unlike the participant survey, for which responses of "don't know" and "refused" of scoring related questions are set to 0.5 free ridership, if contractors answer "don't know" or "refused" to CF1 or CF3, no free-ridership score is calculated, and the participant free-ridership is retained. Therefore, for contractors, "don't know" and "refused" analysis was not applicable beyond the presentation of frequencies.

RCD and select products NTG study

If contractors answer "don't know" or "refused" to CF1 and CF3, no free-ridership score is calculated, and the participant free-ridership is retained.

Table 34 shows the number of “don’t know” and “refused” responses by question. A total of 42 measure-level contractor free-ridership scores (28%) could not be calculated due to these responses.

Table 34: Contractor Free-ridership with "Don't know" and "Refused" Excluded and Deemed as Non-Free Riders

Questions	Number of Eligible R's	Don't Know /Refused (n)	Don't Know/ Refused %	Reported				Don't know/Refused Responses Removed		Don't know/Refused Responses Deemed Non-Free riders	
				FR	P SO	C SO	NTG Ratio	FR	NTG Ratio	FR	NTG Ratio
CF0 Our records show that in <year> your company sold <QTY> residential high-efficiency <measure> that received rebates from Mass Save. Does this sound right?	154	14	9%								
CF1 If Mass Save had not offered any rebates in <year>, about what percentage of the <QTY> high efficiency <measure> that received rebates would you have sold that year?	151	42	28%								
CF2 [IF CF1 < = 100] To confirm, you are estimating that in <year>, your company likely would still have installed <QTY * (CF1 / 100)> <measure> of the <QTY> you sold if Mass Save had not offered the rebates. Is that roughly correct?	108	3	3%	31%	12%	4%	85%	n/a		n/a	
CF3 [IF CF2 = 2] Could you provide your best estimate of about how many of the <QTY> <measure> your company would still have sold in <year> if Mass Save had not offered the rebates?	6	0	0%								

Products NTG study

Contractor surveys were not applicable for this study.

R&A NTG study

The R&A NTG study contractor free-ridership battery did not follow "Consistent Methodology." Instead, it employed a simplified approach to obtain an average contractor free-ridership score at the measure level. "Consistent Methodology" is best suited for situations where it is possible to get both the homeowner and the contractor survey completed. This was not possible for this study due to limited program tracking data and difficulty in project recollection. Table 35 below shows the two alternative free-ridership questions that were asked instead of the full contractor free-

ridership battery shown in Table 32. There were no "don't know" or "refused" responses to these two questions; therefore, additional analysis was not applicable.

Table 35: Contractor Free-ridership with "Don't know" and "Refused" Excluded and Deemed as Non-Free Riders

Questions	Number of Eligible R's	Don't Know /Refused (n)	Don't Know/ Refused %	Reported				Don't know/Refused Responses Removed		Don't know/Refused Responses Deemed Non-Free riders	
				FR	P SO	C SO	NTG Ratio	FR	NTG Ratio	FR	NTG Ratio
FR1 Without the Mass Save Renovations and Additions Initiative incentive, how likely would you have been to install the same efficiency/amount of [MEASURE]??	263	0	0%	31%	12%	4%	85%		n/a		n/a
FR2. [IF FR1 = 3 or 4] Without the incentive from the Mass Save Renovations and Additions Initiative, what efficiency/amount of [MEASURE] would have been installed instead?	18	0	0%								

3.4.2.2 Evaluation Team Questions

Question: It wasn't clear if we should use an average of ALL contractors we speak to or use the score of the specific contractor who worked with that customer.

Response: We use all contractors' total weighted average free-ridership for the applicable measure. That score replaces ALL customers who say their contractor made the decision or was heavily influential in their decision-making.

Resolution: The methodology has been updated to clarify, *"Responses to these questions will be used to calculate a free-ridership score that will replace the participant free-ridership score or averaged with the participant's free-ridership score. Where participants rated the contractor highly influential in doing the project and indicated, they used the contractor's equipment recommendations, use the contractor's free-ridership score in place of the participant's free-ridership score. Specifically, for all customer respondents who said their contractor decided or influenced it heavily, replace the customer's free-ridership score with the weighted free-ridership of all contractors for the given measure."*

Question: Should the contractor survey focus on the most active housing sector for each measure? To limit the survey length and reduce response fatigue, we focused on the housing sector in which they were most active (single-family or multifamily). We added a question, "Which housing type does your company specialize or focus on?" which was asked if the contractor indicated that they installed measures in both single-family and multifamily dwellings. We then

asked about only one of the housing types for each measure, prioritizing multifamily over single-family after noticing fewer instances of multifamily contractors compared to single-family.

Response: Where contractors are involved in multiple sectors (e.g., multifamily, single-family), consider asking contractors about just one sector. This approach reduces respondent burden by allowing contractors to focus on the most active sector. Evaluators should add a question such as "In what sector does your organization complete the most projects?" to gather which sector for the survey.

Resolution: Added clarification to the methodology but did not make any core changes.

Question: If a contractor estimates their measure quantity sold without the program to be higher than program sales, should those results be excluded from the analysis?

Response: While the response could be deemed illogical, it should not be removed from the free-ridership calculation.

Resolution: No change to the methodology document.

3.4.3 Contractor Spillover Score

The contractor spillover questions are used to assess the program's influence on contractors' recommendations and the influence the program and other factors had on the sales of program-eligible equipment outside the program. Figure 7 below shows the contractor spillover calculation followed by the spillover question battery in Table 36.

Figure 7: Contractor Spillover Calculation

Contractor Spillover = (Non-program sales x (Influence score / 10)) / Program Sales

Table 36: Participating Contractor Spillover Questions

NS1	Think of all your sales of residential <measure>, including standard-efficiency and high-efficiency and rebated and non-rebated units. About how many did your company sell **in total** in <year>? ____ Total installations 8888 Don't know [skip to next section] 9999 Refused [skip to next section]
NS2	About how many of the sales of residential <measure> in <year> were high-efficiency, including rebated and non-rebated units? ___ Number of installs 8888 Don't know [skip to next section] 9999 Refused [skip to next section]
NS3	Did all the high-efficiency <measure> that your company sold in <year> receive a Mass Save rebate of some kind? 01 Yes [skip to next section]

	02 No 88 Don't know [skip to next section] 99 Refused [skip to next section]
NS4	About what percentage of the high-efficiency <measure> that your company sold in <year> did not receive a Mass Save rebate even though they would have qualified for one? ___ Percent [0 TO 100] 888 Don't know 999 Refused
NS5	[ASK IF NS2 AND NS4 < 888] To confirm, that means about <(NS2) * (NS4 / 100)> of the <NS2> high-efficiency <measure> your company sold in <year> did not receive **a Mass Save rebate** even though they would have qualified for one. Is that **roughly** correct? 01 Yes 02 No 88 Don't know 99 Refused
NS6	[ASK IF NS5 = 2, 88, OR 99] About how many of the sales of residential high-efficiency <measure> in <year> did not receive a rebate even though they qualified? ____ [OPEN END NUMERIC] 8888 Don't know 9999 Refused [NOTE: IF NONE SOLD OUTSIDE PROGRAM ((NS6 = 0, 8888, OR 9999) OR (NS3 = 1) OR (NS4 = 0)), SKIP TO NEXT MEASURE/SECTION]
NS7	Next, I'd like you to think about the energy-efficient equipment you sold that did <u>not</u> receive a rebate. Using a 0 to 10 scale where 0 is 'not at all influential' and 10 is 'very influential,' how influential were the Mass Save rebates on your sales of high-efficiency <measure> in <year>? __ 0 to 10 Rating 88 Don't know 99 Refused
NS8	Using that same scale (where 0 is 'not at all influential' and 10 is 'very influential'), how influential was the program support such as marketing, advertising, education and training on your company's sales of these high-efficiency <measures>? __ 0 to 10 Rating 88 Don't know 99 Refused

3.4.3.1 Summary of Results

RCD and select products NTG study

As shown in Table 37 below, 40% of respondents could not estimate their total measure sales to allow the program to capture any spillover. Participating contractors also had difficulty determining

how many of their sales qualified for a program incentive but did not receive one (38% could not respond). The high "don't know" and "refused" responses indicate the spillover is a conservative estimate.

Table 37: Don't know/Refused Response Count for Contractor Spillover Questions

Questions	Measure Count	Don't Know/Refused (n)	Don't Know/Refused %
NS1 Think of all your sales of residential <measure>, including standard-efficiency and high-efficiency and rebated and non-rebated units. About how many did your company sell **in total** in <year>?	154	62	40%
NS2 About how many of the sales of residential <measure> in <year> were high-efficiency, including rebated and non-rebated units?	69	9	13%
NS3 Did all the high-efficiency <measure> that your company sold in <year> receive a Mass Save rebate of some kind?	83	7	8%
NS5 [ASK IF NS2 AND NS4 < 888] To confirm, that means about <(NS2) * (NS4 / 100)> of the <NS2> high-efficiency <measure> your company sold in <year> did not receive **a Mass Save rebate** even though they would have qualified for one. Is that **roughly** correct?	18	0	0%
NS6 [ASK IF NS5 = 2, 88, OR 99] About how many of the sales of residential high-efficiency <measure> in <year> did not receive a rebate even though they qualified?	16	6	38%

Over 10% of respondents did not answer questions about the influence of the rebates and program support. Four respondents (16%) did not know or refused to answer the influence of the program rebates, and three respondents (12%) did not know or refused to answer how influential the program support was on sales of high-efficiency equipment.

Table 38: Don't know/Refused Response Count for Contractor Spillover Questions

Questions	Measure Count	Don't Know/ Refused (n)	Don't know/ Refused %
NS7 Next, I'd like you to think about the energy-efficient equipment you sold that did not receive a rebate. Using a 0 to 10 scale where 0 is 'not at all influential' and 10 is 'very influential,' how influential were the Mass Save rebates on your sales of high-efficiency <measure> in <year>?	25	4	16%
NS8 Using that same scale (where 0 is 'not at all influential' and 10 is 'very influential'), how influential was the program support such as marketing, advertising,	25	3	12%

Questions	Measure Count	Don't Know/ Refused (n)	Don't know/ Refused %
education and training on your company's sales of these high-efficiency <measures>?			

Products NTG study

Contractor surveys were not applicable for this study.

R&A NTG study

The R&A NTG study did not follow the standard methodology questions. Instead, spillover was simplified to ask for the total number of renovation and addition projects completed, followed by what percent received incentives. All respondents provided a response or an estimate, so there were no "don't know" or "refused" responses.

Table 39: Don't know/Refused Response Count for Contractor Spillover Questions

Questions	Number of Eligible Respondents	Don't Know/ Refused (n)	Don't know/ Refused %
P5 Approximately how many renovation and addition projects did you complete in Massachusetts in 2019?	31	0	0%
P6 Approximately what percent of your renovation and addition projects completed in 2019 received incentives from the Mass Save Renovations and Additions Initiative?	31	0	0%
NS7 Next, I'd like you to think about the energy-efficient equipment you sold that did not receive a rebate. Using a 0 to 10 scale where 0 is 'not at all influential' and 10 is 'very influential,' how influential were the Mass Save rebates on your sales of high-efficiency <measure> in <year>?	31	0	0%

3.4.3.2 Evaluation Team Questions

Question: An area was identified with some redundancy in the contractor free-ridership battery. CF0 and, if needed, CF0a ask about the quantity of measures that received rebates through the program. NS3 or NS4 in the spillover battery ask if all high-efficiency measures sold received a rebate, then ask the approximate percentage that did not receive a rebate. By asking questions CF0 and, if needed, CF0a, it is not necessary to ask NS3 or NS4; we could calculate whether all measures received a rebate (NS3), and if not, what percent received a rebate (NS4).

Response: Agree there is redundancy with NS3 and NS4 and that those values can be calculated using CF0, CF0a, NS1, and NS2.

Resolution: The methodology has been updated to remove NS3 and NS4 and clarifies, *"NS3 can be calculated from NS2-CF0 and NS4 can be calculated from (NS2-CF0)/NS2."*

Question: During the analysis, we replaced the estimated measure quantity of high-efficiency units sold (NS2) with the total measure quantity sold (NS1) if NS2 was greater than NS1. We replaced NS2 with program sales if NS2 was less than program sales. Similar to the previous issue, this erroneous reporting was rare (n=2); however, this is a logical quality assurance practice without reducing the number of valid responses.

Response: Logically, NS2 should not be greater than NS1. Added a note to this effect. Also added a note to NS1 using the information collected in CF0 and CF0a. We recommend keeping NS1 and NS2 with self-report information and not trying to reconcile all the information. While we understand we can match part of NS2 with the program data, self-report between NS1 and NS2 would be "off" by similar amounts.

Resolution: Added introductory sentence to NS1 in the methodology. "*Moving on from the <CF0 or CF0a> units of high efficiency <measure>that received a Mass Save incentive to talk about **all** your sales of residential <measure>, including standard-efficiency and high-efficiency and rebated and non-rebated units. About how many did your company sell **in total** in <year>?"*

Question: The NTG framework briefly mentions nonparticipant contractors but does not come out and recommend an approach for using nonparticipant contractor research within the methodology. Should we use nonparticipant contractor data in the NTG study?

Response: Questions are focused on participating contractors. As of the writing of this methodology, there is no consistently utilized protocol for incorporating nonparticipating contractors. There are several challenges in talking with nonparticipating contractors that make developing a protocol difficult. These include double-counting with participating customers' trade allies, ensuring nonparticipating contractors have some exposure to the program, and the difficulty of identifying the nonparticipating contractor population.

Resolution: Added clarification to the methodology, but no core changes were made.

Question: Is the nonparticipant spillover calculated for every spillover measure the contractor sells and then summed up to a final nonparticipant value?

Response: Yes, this is correct.

Resolution: Added clarification to the methodology, but no core changes were made.

Section 4 Revised Residential Self-reported NTG Survey Questions and Algorithms

This section replicates Section 4 of *Consistent Methodology for Self-Reported Residential Net-to-Gross Measurement* (MA19X03-B-RSRNTG) which has been now updated with the revisions described in Section 3 of this report. It supersedes the methodology presented in the MA19X03-B-RSRNTG report.

The batteries of questions presented here are to be used in measuring self-reported NTG from participating end-users or contractors for residential programs for which such measurement is appropriate. The batteries described in this section apply to downstream programs other than appliance recycling. Appliance recycling NTG studies should follow the protocols outlined in the Uniform Methods Project.[8]

4.1 Background

The NTG batteries consist of the following five different groups of questions, which are explained in more detail in the remainder of this chapter. We provide guidance to question wording in each section, but evaluators will need to tailor questions to the specific program designs and needs. The more detailed algorithms for these five question batteries can be found in Appendix A of this document. The calculation of spillover in the participant spillover algorithm was updated due to a missing bracket. The participating contractor spillover algorithm has been updated to remove the duplicate questions identified by the evaluation team.

- **Participant market event type** – used to determine whether the reason for purchasing the equipment was due to new construction, the addition of new equipment (to an existing home), replace on failure (ROF), or early replacement. This series of questions identifies participants who should be skipped out of the timing question (ROF and new construction).
- **Participant free-ridership** – includes questions on timing, quantity, and efficiency elements; an influence adjustment; and consistency checks. We recommend asking the participant free-ridership questions as soon as possible after receiving the program-eligible equipment to maximize recall, balancing that with the need to use the most recent participants for setting prospective NTG ratios.
- **Participant spillover** – questions ask about whether participants have purchased similar equipment since participation and the type and efficiency of equipment purchased. Respondents are then asked about the importance of the program on their decision to

[8] Keeling, J.; Bruchs, D. (2017). Chapter 7: Refrigerator Recycling Evaluation Protocol. The Uniform Methods Project: Methods for Determining Energy-Efficiency Savings for Specific Measures. Golden, CO; National Renewable Energy Laboratory. NREL/SR-7A40-68563. http://www.nrel.gov/docs/fy17osti/68563.pdf

purchase the additional efficient equipment and the likelihood of purchasing the additional efficient equipment had they not also purchased the rebated equipment. A sufficient period should elapse before asking the spillover questions (e.g., 12 months) to allow the respondent time to have made purchases on their own.

- **Contractor free-ridership** – contains questions to ask participating contractors/ salespersons about FR for measures for which contractors/salespersons play a prominent role in delivering and promoting the equipment. The questions confirm program sales and determine whether the sales would have occurred had the program rebates not been available.
- **Participating Contractor Spillover** – contains questions on the sale of program-eligible equipment that did not receive a rebate through the program, the influence of the program on the contractor's recommendations, and the influence of the program and other factors on the number of energy-efficient sales outside the program.

Together, these algorithms are used to calculate an NTG ratio as follows:

NTG = (1 – FR) + participant spillover + contractor spillover.

The free-ridership component is taken from either the participant, the contractor, or a combination of the two. We provide instructions in Appendix A on when participant free-ridership should be used and when contractor free-ridership should replace or be averaged with participant results.

In the table below, we indicate which algorithm components are applicable to different types of downstream measures.

Table 40: Algorithm Components

Algorithm component	Rebated and major renovation measures	Audit measures (direct install and leave-behind)	New construction
Market event	Y (excluding weatherization measures)	Y (excluding weatherization measures)	N
Free-ridership - Intent	Y	Y	Y (excluding timing)
Free-ridership - Influence	Y	Y	Y
Participant spillover	Y	Y	Y
Contractor free-ridership	Y	N	Y
Contractor spillover	Y	N	Y
Inconsistency checks	Y	Y	Y

4.2 PARTICIPANT MARKET EVENT TYPE

Participant market event type questions are included in the NTG batteries to identify whether the high-efficiency equipment installed was due to new construction/major renovation, to replace existing equipment that was still functioning (early replacement), because existing equipment failed (ROF), or because the respondent installed a piece of equipment that was not previously used in their home (new equipment). Responses to these questions are used to identify whether participants should be asked the timing questions in the FR battery.[9] Event type questions should be asked for all stand-alone measures, including thermostats.

While the questions do identify rebates that were installed as part of a new construction home, we do not expect many instances where this will occur. This is because the equipment is captured through the new construction program and is therefore not eligible for rebates in another program. We include these questions in the algorithm to capture all the possible scenarios so interviews can flow smoothly, and to will allow us to measure the extent of any new construction cases that may receive a rebate outside the new construction program. Cases identified as new construction through this algorithm should be reviewed by the team responsible for evaluating the new construction program.

The participant market event type questions were determined by reviewing the existing event type battery used in the prior C&I omnibus SR survey (TXC49), the results of the findings of the 2019 study *C&I Gas Equipment Early Replacement* (MA19C11-G-ERCUSTSRV)[10], and event type questions identified in the literature review. The recommended market event type questions can be found in Table 41. The algorithm for determining event type can be found in Appendix A.

CAVEAT

Market event type questions do <u>not</u> need to be asked for some types of measures, like weatherization and insulation, where event type does not impact timing.

Table 41. Market Event Type Questions

Market Event			
ET1	Was the high efficiency <measure> installed as part of a new construction project? (select one) 01 Yes [new construction] SKIP TO Next Section 02 No 88 Don't know 99 Refused		
ET2	Did the high efficiency <measure> you installed replace any existing <equipment> or was it a new type of equipment that you did not have in your home before? (select one)		

[9] The event type questions can also be used for other purposes—for example, to give the PAs information to update the percent of units that are early retirement.

[10] Interview findings from the MA C&I Early Replacement Study prepared by the C&I Market Assessment Team (Cadeo and Navigant), February 4, 2020.

	01	Replaced existing equipment	
	02	New equipment	[new equipment] SKIP TO Next Section
	88	Don't know	
	99	Refused	
ET3	[ASK IF ET2=1, DK, R] Which of the following best describes the condition of your old equipment? (read list)		
	01	The old equipment was working with no need of repair	
	02	The old equipment was working with need of minor repair	
	03	The old equipment was working with need of major repair	[ROF]
	04	The old equipment was no longer working	[ROF]
	88	Don't know	
	99	Refused	
ET4	[ASK IF ET3 = 1, 2, DK, and R] Do you think your old equipment would have lasted another two years? (select one)		
	01	Yes	[ER]
	02	No	[ROF]
	88	Don't know	[ROF]
	99	Refused	[ROF]

4.3 PARTICIPANT FREE-RIDERSHIP

The participant FR score consists of two components: an intent score and an influence score. These are followed by consistency checks.

CAVEAT

The wording in the questions shown below are for incentive-based measures. The wording of questions for direct install measures received free as part of an audit will need to be changed to include the phrase "If you had not received the free <measure(s)> during the audit, …".

4.3.1 Intent Score

The intent score incorporates a series of questions to determine whether, in the absence of the program, the respondent:

- would have purchased any type of equipment at the same time, a later date, or not at all (timing)
- would have purchased the same, lesser, or greater quantity of the high-efficiency equipment (quantity)
- would have purchased the same or lesser efficiency equipment (efficiency).

Questions used to determine timing, quantity, and efficiency are contained in Table 42, along with exceptions that evaluators should make depending upon the type of high-efficiency measure purchased.

Table 42. Timing, Quantity, and Efficiency (Intent) Questions

Timing (Evaluator: if NC or ROF, do *not* ask participants the timing questions. New equipment and early replacement should be asked the timing questions.)			
T1	Without the <list of all rebates/incentives/support received> from <Sponsor>, how likely is it that you would have installed <u>any</u> <u>type</u> of <equipment> at the same time? Would you say …		
	01	Not at all likely	FR_timing = 0 (skip to 'influence')
	02	Slightly likely	
	03	Somewhat likely	
	04	Very likely	FR_timing = 1.0 (skip to Q1)
	88	Don't know	FR_timing = 0.5 (skip to Q1)
	99	Refused	FR_timing = 0.5 (skip to Q1)
T2	Without the <list of all rebates/incentives/support received> from <Sponsor>, when do you think you would have installed the <equipment>? Would you say …		
	01	Within 6 months of when you did	FR_timing = 1.0
	02	Between 6 months and 12 months of when you did	FR_timing = 0.5
	03	More than 1 year after when you did	FR_timing = 0 (skip to 'influence')
	88	Don't know	FR_timing = 0.5
	99	Refused	FR_timing = 0.5
Quantity (Evaluator: Ask only if quantity purchased/received is greater than 1; for example, would not need to ask if a customer purchased one furnace/boiler or if they received a tune-up. For insulation/weatherization type measures ask this series, but the wording should be revised to "amount of" instead of "quantity of.")			
Q1	Without the <list of all rebates/incentives/support received> from <Sponsor>, how likely is it that you would have installed the <u>exact same [quantity of / amount of]</u> high efficiency <measure>? Would you say it is …		
	01	Not at all likely	FR_qty = 0
	02	Slightly likely	
	03	Somewhat likely	
	04	Very likely	FR_qty = 1.0
	88	Don't know	FR_qty = 0.5
	99	Refused	FR_qty = 0.5
Q2a	[non-insulation/weatherization type measures] Without the rebate or assistance from <Sponsor>, how many high efficiency <measure> would you have installed?		
	__	Quantity would have installed	FR_qty = (Q2a/installed quantity)
	88	Don't know	FR_qty = 0.5
	99	Refused	FR_qty = 0.5

Q2b	[insulation/weatherization type measures] Without the rebate or assistance from <Sponsor>, what percent of the high efficiency <measure> would you have installed?		
	__	Percent would have installed [0-100]	FR_qty = Q2b
	88	Don't know	FR_qty = 0.5
	99	Refused	FR_qty = 0.5
Efficiency (Evaluator: Ask only if efficiency applies; for tune-up, insulation/weatherization type measures, air sealing type measures, etc. skip to 'influence' questions)			
E1	Without the <list of all rebates/incentives/support received> from <Sponsor>, how likely is it that you would have installed the exact same high efficiency level of <measure>? Would you say …		
	01	Not at all likely	FR_eff = 0
	02	Slightly likely	FR_eff = 0.25
	03	Somewhat likely	FR_eff = 0.5
	04	Very likely	FR_eff = 1.0
	88	Don't know	FR_eff = 0.5
	99	Refused	FR_eff = 0.5

The Efficiency question may need to be modified if the technology type varies by situation. For example, in the case of a thermostat replacement the FR would depend on the type of thermostat being replaced. In such a case, consider revising the efficiency question as follows: "Without the rebate or assistance from <Sponsor> for your thermostat, what type of a thermostat did your smart thermostat replace?" Response options could include "would not have installed any thermostat" (FR_eff=0), "a non-programmable thermostat (you turn it on and off and change temperature by hand)" (FR_eff=0), "a programmable thermostat (you program temperature settings that change automatically at different times of the day, but with no remote control)" (FR_eff=0.25), "a wi-fi programmable thermostat (you can program and control it remotely with your smartphone)" (FR_eff=0.5), or "a wi-fi smart thermostat (you can control it remotely with your smartphone but it also learns your schedule and preferences and adjusts the temperature based on those)" (FR_eff=1.0) along with the standard values for Don't know and Refused.

The intent score is the minimum of the timing, quantity, and efficiency scores. Note that if one or more of these questions are not asked because they are not applicable, the evaluator should conduct a sensitivity analysis to determine what impact only using one or two of these three scores has on the intent score.

4.3.2 Influence Score

The influence score shown in Table 43 asks how influential different program attributes were in the participant's decision to install the high-efficiency measure through the program. The number of program attributes in question X1 will vary depending upon the program's design. The evaluator should review logic models and program theory and conduct staff interviews to inform the list of elements to include. The more typical elements programs used to influence customer decision-making include information; incentives or rebates; interaction with program staff (technical assistance); building audits or assessments; and financing. Depending on the design of the program and what role contractors play in promoting energy efficient equipment through the program, questions X2 and X3 may get asked. These questions capture the influence the

contractor or salesperson had on the decision and how the decision was made on what efficiency of equipment to install. The evaluator should make the decision on whether questions X2 and X3 get asked.

Table 43. Influence Questions

Influence: (The evaluator should review logic models and program theory, and conduct staff interviews to inform the list of elements to include)	
X1	On a scale of 0 to 10 where 0 is 'not at all influential' and 10 is 'very influential,' how influential was < attribute from list below> on your decision to install the high efficiency <measure>? [NOTE: program features asked about should be tailored to the program being evaluated; list max of 6 factors] a. the program rebate b. the program's marketing materials c. [Evaluation team will need to insert other features of the program design/delivery] __ Rating 66 Not applicable 88 Don't know 99 Refused
X2	[ASK IF PROGRAM THEORY SHOWS CONTRACTOR/SALESPERSON PLAY A PROMINENT ROLE IN DELIVERING AND PROMOTING THE PROGRAM] Using the same 0 to 10 scale where 0 = 'not at all influential' and 10 is 'very influential,' how influential was the contractor or salesperson on your decision to install the high efficiency <measure>? __ Rating 66 Not applicable 88 Don't know 99 Refused
X3	[ASK IF CONTRACTOR/SALESPERSON HAS THE HIGHEST INFLUENCE SCORE AND PROGRAM THEORY SHOWS CONTRACTOR/SALESPERSON PLAY A PROMINENT ROLE IN DELIVERING AND PROMOTING THE PROGRAM] Which of the following best describes how you selected the new high efficiency <measure>? (Select one) [rotate options 1 through 3] 01 I did some research on <measure>s and made my own choice 02 My contractor suggested one <measure> efficiency level, and I agreed – (use Contractor FR results) 03 My contractor suggested various <measure> efficiency levels, and I chose one (average of contractor and customer FR results) 04 Something else (specify)

The influence score is based on the program attribute that is rated highest in question X1. The influence score calculation is shown in Table 44.

Table 44. Calculation of the Influence Score

Rating	Score
Not at all influential – 0	1.0
1	0.9
2	0.8
3	0.7
4	0.6
5	0.5
6	0.4
7	0.3
8	0.2
9	0.1
Very influential – 10	0.0

4.3.3 Participant Free-ridership Score

The Participant FR score is the average of the intent and influence scores (Intent Score + Influence Score / 2).

4.3.3.1 Participant and Contractor Free-ridership Reconciliation

If the participant is asked questions X2 and X3, their response to these questions may impact whether and how the contractor FR score is used. If the respondent says the contractor was the most influential factor in their decision (i.e., X2 is rated higher than any program attributes in X1), and the respondent in question X3 says the contractor suggested one measure efficiency level, and the customer agreed, the evaluator should use the contractor FR score instead of the customer's responses to the intent and influence questions. If the respondent in question X3 said the contractor was the most influential program element, and the contractor suggested various measure efficiency levels, and the customer chose the measure, the evaluator should average the contractor and customer FR results. See the Participant and Contractor Free-ridership Reconciliation algorithm in Appendix A for details.

4.3.4 Consistency Questions

The FR battery, shown in Table 45, also includes two consistency questions to ensure that responses to the intent and influence questions are consistent. To minimize respondent burden, these questions are only asked in cases where responses are inconsistent to the intent and influence questions. For example, if the respondent says in the intent questions that they would not have purchased any equipment within a year, or they would not have purchased the same quantity and efficiency of equipment, but then say in the influence questions that the program elements asked about (e.g., program rebate, audit, marketing materials) were not influential (a rating of 0, 1, 2, or 3), the consistency questions should be triggered. As another example, if the respondent says in the intent questions that they would have installed the exact same efficiency and quantity of equipment at the same time, but rate one or more of the program elements in the influence questions as highly influential in their decision to install the efficient equipment (a rating of 7, 8, 9, or 10), the consistency questions should be triggered.

While not all jurisdictions like or incorporate consistency checks, we are proposing to include these questions until the battery has been thoroughly tested to ensure the questions capture the necessary elements. These questions could be removed in future studies, assuming they are not found to be heavily relied upon to make adjustments.

Table 45. Consistency Questions

Consistency	
C1	**(CONSISTENCY QUESTION)** Did you make the decision to install the high efficiency <measure> before or after you learned of the rebate? (select one) 01 Before 02 After 88 Don't know 99 Refused
C2	**(CONSISTENCY QUESTION)** Please describe the influence that the program <list of all rebates/incentives/support received) had on your decision to install the high efficiency <measure>? [Record verbatim response]

The team conducting the evaluation should examine responses to the consistency questions and use expert judgment to try to resolve the inconsistency and determine if the program should receive any credit. If the inconsistency cannot be resolved, we recommend removing the respondent from the FR estimate. The evaluation team should document decisions made along with the rationale during reporting.

4.4 PARTICIPANT SPILLOVER

In the participant spillover questions, respondents are asked the type and efficiency of equipment purchased, the importance of the program on their decision to purchase the additional efficient equipment, and the likelihood of purchasing the additional efficient equipment had they not also purchased the rebated equipment. The participant spillover questions are heavily based on the prior TXC34 Massachusetts spillover battery.

In the event that Massachusetts is able to conduct the participant FR questions soon (e.g., within six months) after the equipment purchase, the evaluator should consider using a separate survey to estimate participant spillover so that the participants have enough time to make a purchase on their own.

The spillover questions can be found in Table 46.

Table 46. Participant Spillover Questions

Participant Spillover[11]	
S1	Since installing the equipment, have you made other energy-saving purchases or changes that did not receive a rebate through the Mass Save® program? (Select one) 01 Yes 02 No (skip to next section) 88 Don't know (skip to next section)
S2	Did your experience with the Mass Save program influence your decision to make these energy-saving purchases or changes? (Select one) 01 Yes 02 No (skip to next section) 88 Don't know (skip to next section)
S3	What energy-saving purchases or changes did you make? (Select all that apply) [Note: Make light bulbs and other upstream products available for interviewers and respondents, but these measures will not be counted towards spillover] 01 LEDs/Energy-efficient light bulbs 02 Energy-efficient appliance(s) (specify type) 03 HVAC equipment (specify type) 04 Attic, wall, or basement insulation 05 Duct sealing or duct insulation 06 Air sealing of leaks 07 Smart thermostat 08 Water heater 09 Other purchases or changes (specify)
[ASK S3b to S6 FOR EACH MEASURE OTHER THAN LIGHTING[12] SELECTED IN S3]	
S3b	[if S3 = 2 or 3 or 8 or 9] Can you describe the <equipment from S3> work in more detail? If applicable, include the quantity installed, the type of equipment, and the efficiency levels installed. (For AC probe on SEER rating, for furnaces or boilers probe on AFUE, for heat pumps probe on Heating season performance factor [HSPF] or SEER)

[11] NOTE: To be able to calculate spillover, at a minimum, the survey demographics should collect information on water heater fuel, clothes dryer fuel, primary heating fuel, and type of cooling equipment in order to determine what prescriptive savings to apply.

[12] LEDS are assumed to be counted as part of the upstream program. If the upstream program is ended, then they should be counted towards SO.

S3c How did you know the equipment was energy efficient and would have qualified for a rebate through the program?

[Record verbatim response]

Possible response categories:

Efficiency rating or label of equipment, such as an "ENERGY STAR®" logo
Equipment dealer/retailer said it was energy efficient
Personal experience
Met utility rebate requirements
Third party report, such as Consumer Reports
Recommendations from the contractor/installer
Did not rely on any specific type of information
Internet/website
Other (specify)
Don't know

S3d [If S3=2, 3 or 8 and ENERGY STAR not mentioned in S3b] Was the <equipment from S3> ENERGY STAR-labeled? (Select one)

01	Yes
02	No
88	Don't know

S4 How important was your experience with the Mass Save program on your decision to install <equipment from S3> that did not receive a rebate from <Sponsor>? Was it ...

01	Not at all important	score = 0
02	Slightly important	score = 3
03	Somewhat important	score = 6
04	Very important	score = 10
66	Not applicable	
88	Don't know	
99	Refused	

S5 How likely is it that you would still have purchased/installed energy-efficient <equipment from <S3> if you had *not* already received a rebate for the energy-efficient <measure>? Would you have been ... to have installed the energy-efficient <equipment from S3>?

01	Not at all likely	score = 10
02	Slightly likely	score = 6
03	Somewhat likely	score = 3
04	Very likely	score = 0
66	Not applicable	
88	Don't know	
99	Refused	

S6	**(Consistency question)** Why did you not submit an application for a rebate through the Mass Save program for this additional equipment? [record verbatim response] [Note: for respondents who indicate that the equipment did not qualify, no spillover should be attributed to the program]

Based on responses to S3b, evaluators should align efficiency specifications with TRM specifications to determine measure savings. Secondary sources should be used to estimate savings for non-like measures if they are not available in the MA TRM.

Where spillover is found, a follow-up telephone or web survey may be used to allow participants to submit pictures. The pictures could help determine if the equipment qualifies for spillover and the savings that may be attributed to the program. Incentives may need to be considered if this approach is used.

The spillover score is calculated for each measure as follows. If the score is 5 or lower, spillover is set to 0. If the score is higher than 5, the spillover savings is the appropriate TRM (or secondary sources) savings.

Participant Spillover score = (S4 + S5) / 2

The spillover rate equals the sum of the savings associated with the additional TRM (or secondary sources) spillover measure divided by the sum of the program savings. To allow for spillover for both gas and electric measures, savings should be converted to MMBTUs.

Caveat

To avoid double-counting, evaluation teams will need to review spillover mentioned by participants and contractors at the measure-level. If spillover is calculated for the same measures by both participants and contractors, the evaluator should subtract any savings reported by the participants from the contractor spillover estimate. We base this conservative approach on the assumption that participating design professionals or vendors were reporting the same information as the participant's spillover project. This is consistent with how the C&I SR NTG calculation works.

Where the evaluator has evidence that spillover may be high, this may warrant a separate study.

4.5 CONTRACTOR FREE-RIDERSHIP

For measures or programs where the program theory shows contractors play a prominent role in delivering and promoting the program, such as HVAC equipment in residential retrofit, contractors

will be asked about the influence of the program on their sales. This approach was utilized in the recent *Residential HVAC Net-to-Gross and Market Effects* study (TXC34).

The contractor FR battery starts by confirming the contractor's sales of the program-qualifying equipment. The contractor is then asked how their sales would have changed if the program rebates had not been available. The questions can be found in Table 47.

Table 47. Contractor Free-ridership Questions

Contractor Free-ridership	
CF0	For the next set of questions, we are going to ask you to think about your company's experiences with Mass Save rebates. Our records show that in <year> your company sold <QTY> residential high-efficiency <measure> that received rebates from Mass Save. Does this sound right? 01 Yes [program sales = QTY] skip to CF1 02 No 88 Don't know [program sales = QTY] skip to CF1
CF0a	About how many residential <measures> that received rebates from Mass Save did your company sell in <year>? ____ [0 TO 1000] [program sales = CF0a] 8888 Don't know [skip to next section] 9999 Refused [skip to next section]
CF1	If Mass Save had not offered any rebates in <year>, about what percentage of the <QTY> high efficiency <measure> that received rebates would you have sold that year? [Interviewer note: If the contractor only sold 1, then the response should be 0% if said would not have sold any, or 100% if said would have sold the same amount] ___ [0 TO 100] 888 Don't know [skip to next section] 999 Refused [skip to next section]
CF2	[IF CF1 < = 100] To confirm, you are estimating that in <year>, your company likely would still have installed <QTY * (CF1 / 100)> <measure> of the <QTY> you sold if Mass Save had not offered the rebates. Is that **roughly** correct? 01 Yes [skip to next section] free-ridership = CF1 02 No 88 Don't know [skip to next section] 99 Refused [skip to next section]
CF3	[IF CF2 = 2] Could you provide your best estimate of about how many of the <QTY> <measure> your company would still have sold in <year> if Mass Save had not offered the rebates? ___ [0 TO 1000] 8888 (Don't know) **Free-ridership = CF3/program sales**

Responses to these questions will be used to calculate a free-ridership score that will replace the participant FR score or may be averaged with the participant score. Where participants rated the contractor highly influential in doing the project and indicated they used the contractor's equipment recommendations, use the contractor FR score in place of the participant FR score. Specifically, for all customer respondents who said their contractor made the decision or influenced it heavily, replace the customer's free-ridership score with the weighted average free-ridership of all contractors for the given measure.

When the contractor was highly influential in the participant's decision but ultimately the customer made the equipment decision themselves with input from the contractor, the participant and contractor FR scores will be averaged. This allows the program to factor in the impact the contractors have had on the selection of the program qualifying-equipment.

Caveat

This study has been designed to run concurrently with the participant FR study. If there is ample time, the evaluation team could run the contractor FR study upon completion of the participant FR study. In doing so, only measures where customers indicated the contractor was highly influential in the decision would need to be included in the contractor FR study.

4.6 PARTICIPATING CONTRACTOR SPILLOVER

The contractor spillover questions are asked of participating contractors.[13] These questions are used to assess the influence of the program on the contractor's recommendations, and the influence of the program and other factors on the number of sales of program-eligible equipment outside the program, after capturing sales information. This approach is similar to that used in the prior TXC34 Massachusetts study.

Some questions, such as NS2, are not needed for measures such as insulation, where there is no distinction between high-efficiency and standard-efficiency.

Questions are focused on participating contractors. As of the writing of this methodology, there is no consistently utilized protocol for incorporating nonparticipating contractors. There are a number of challenges in talking with nonparticipating contractors make developing a protocol for them difficult. These include double counting with participating customers' trade allies, ensuring nonparticipating contractors have some exposure to the program, and the difficulty of identifying the nonparticipating contractor population.

The questions can be found in Table 48.

[13] The evaluation team could also conduct surveys with nonparticipating contractors to assess the influence of programs on contractors' sales of efficient equipment. Depending upon the program theory or changes in the market, these questions may be considered for future studies.

Table 48. Participating Contractor Spillover Questions

NS1	Moving on from the <CF0 or CF0a> units of high efficiency <measure> that received a Mass Save incentive to talk about **all** your sales of residential <measure>, including standard-efficiency and high-efficiency and rebated and non-rebated units. About how many did your company sell **in total** in <year>? ____ Total installations 8888 Don't know [skip to next section] 9999 Refused [skip to next section]
NS2	About how many of the sales of residential <measure> in <year> were high-efficiency, including rebated and non-rebated units? ___ Number of installs [number should be less than NS1] 8888 Don't know [skip to next section] 9999 Refused [skip to next section]
NS3	[ASK IF NS2 < 8888 AND CF0 < 888] To confirm, that means about <(NS2-CF0)> of the <NS2> high-efficiency <measure> your company sold in <year> did not receive **a Mass Save rebate** even though they would have qualified for one. Is that **roughly** correct? 01 Yes 02 No 88 Don't know 99 Refused
NS4	[ASK IF NS3 = 2, 88, OR 99] About how many of the sales of residential high-efficiency <measure> in <year> did not receive a rebate even though they qualified? ____ [OPEN END NUMERIC] 8888 Don't know 9999 Refused [NOTE: IF NONE SOLD OUTSIDE PROGRAM ((NS4 = 0, 8888, OR 9999) OR (NS2=CF0) OR (NS2-CF0)= 0)), SKIP TO NEXT MEASURE/SECTION]
NS5	Next, I'd like you to think about the energy-efficient equipment you sold that did <u>not</u> receive a rebate. Using a 0 to 10 scale where 0 is 'not at all influential' and 10 is 'very influential,' how influential were the Mass Save rebates on your sales of high-efficiency <measure> in <year>? __ 0 to 10 Rating 88 Don't know 99 Refused
NS6	Using that same scale (where 0 is 'not at all influential' and 10 is 'very influential'), how influential was the program support such as marketing, advertising, education and training on your company's sales of these high-efficiency <measures>? __ 0 to 10 Rating 88 Don't know 99 Refused

Where contractors are involved in multiple sectors (e.g., multifamily, single-family) consider asking contractors about one. This approach reduces respondent burden by allowing contractors to focus on the most active sector. Evaluators should add a question such as "In what sector does your organization complete projects?" to gather which sector for the survey.

The following general guidance is offered for collecting savings for spillover measures. When spillover measures are incentivized through a Massachusetts program, evaluators should use the lowest savings in the Massachusetts TRM. If the measure does not include any savings values in the Massachusetts TRM, evaluators should utilize sources such as the Illinois TRM (IL TRM)[14]. Evaluators should adopt a 50% discount to those savings to provide a conservative estimate. If a measure is not incentivized through a Massachusetts program, evaluators can use other sources for savings, such as the IL TRM.

The influence score is based on the maximum rating in NS5 and NS6. The contractor spillover score is then calculated as:

(Non-program sales x (Influence score / 10)) / Program Sales

Caveat

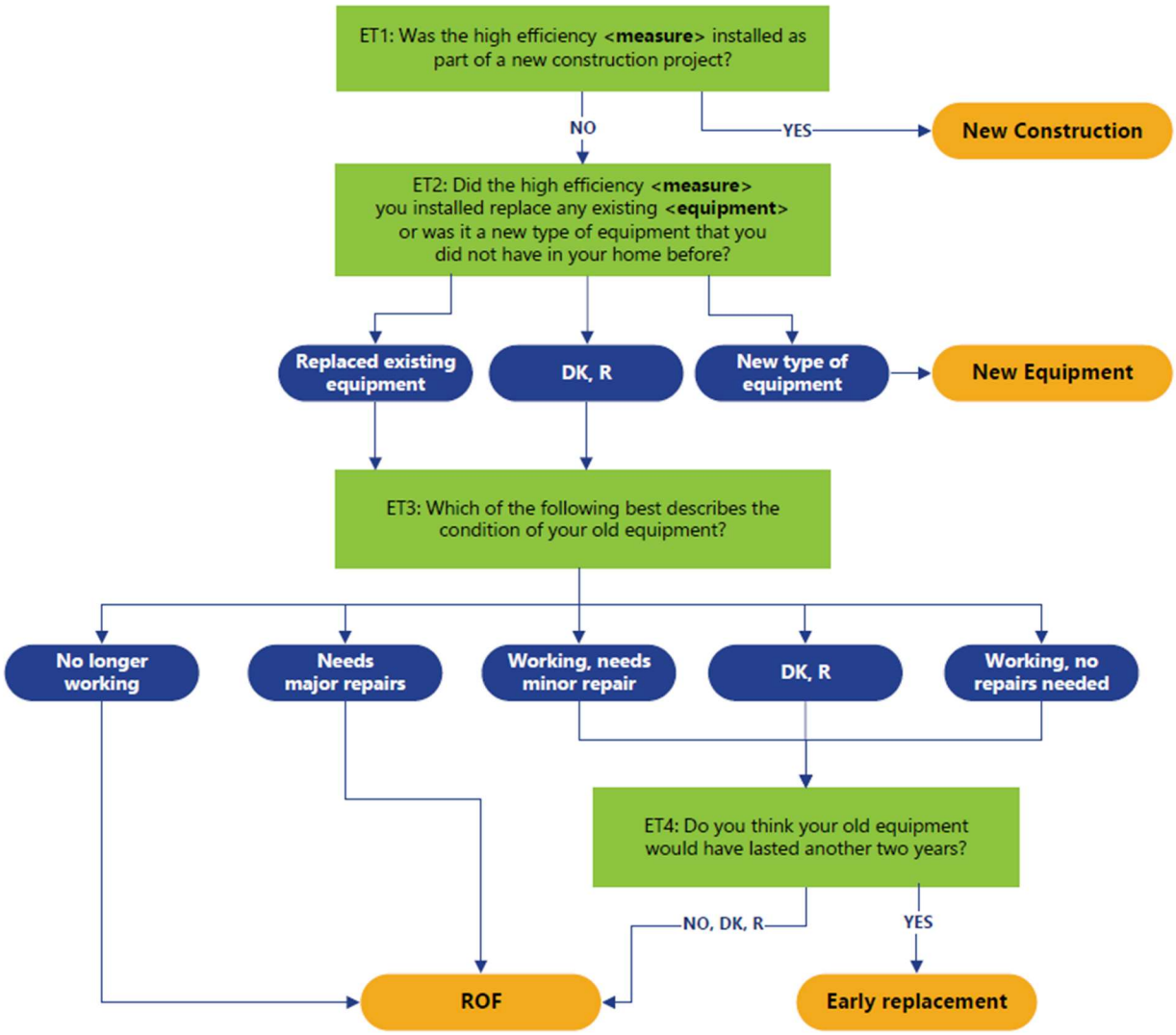
As mentioned above in the Participant SO section above, evaluation teams will need to review any spillover mentioned by participants and contractors at the measure-level to avoid any double-counting of savings. Any spillover calculated for the same measures that is reported by both participants and contractors will need to be reviewed and handled appropriately. The evaluator should subtract any savings reported by the participants from the contractor spillover estimate to conservatively avoid any double counting.

[14] Illinois Statewide Technical Reference Manual - Illinois Energy Efficiency Stakeholder Advisory Group (ilsag.info); https://www.ilsag.info/technical-reference-manual/

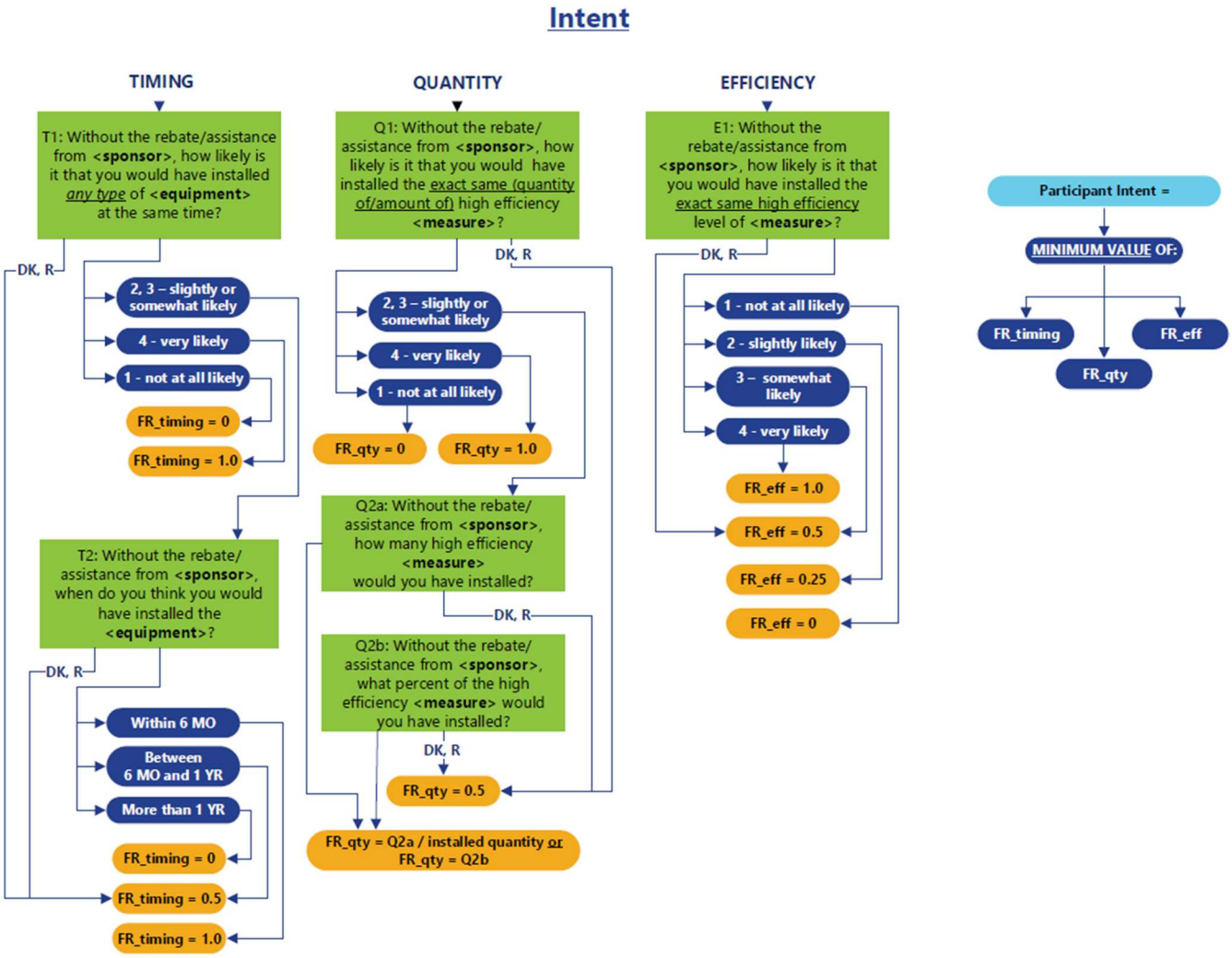
Appendix A Scoring Algorithms

This section contains the algorithms and score mapping for the downstream self-report rebate programs. Two of these algorithms have been updated from the original methodology document to reflect the edits mentioned in Section 3. The spillover calculation in the participant spillover algorithm was updated due to a missing bracket. The participating contractor spillover algorithm has been updated to remove the duplicate questions identified by the evaluation teams.

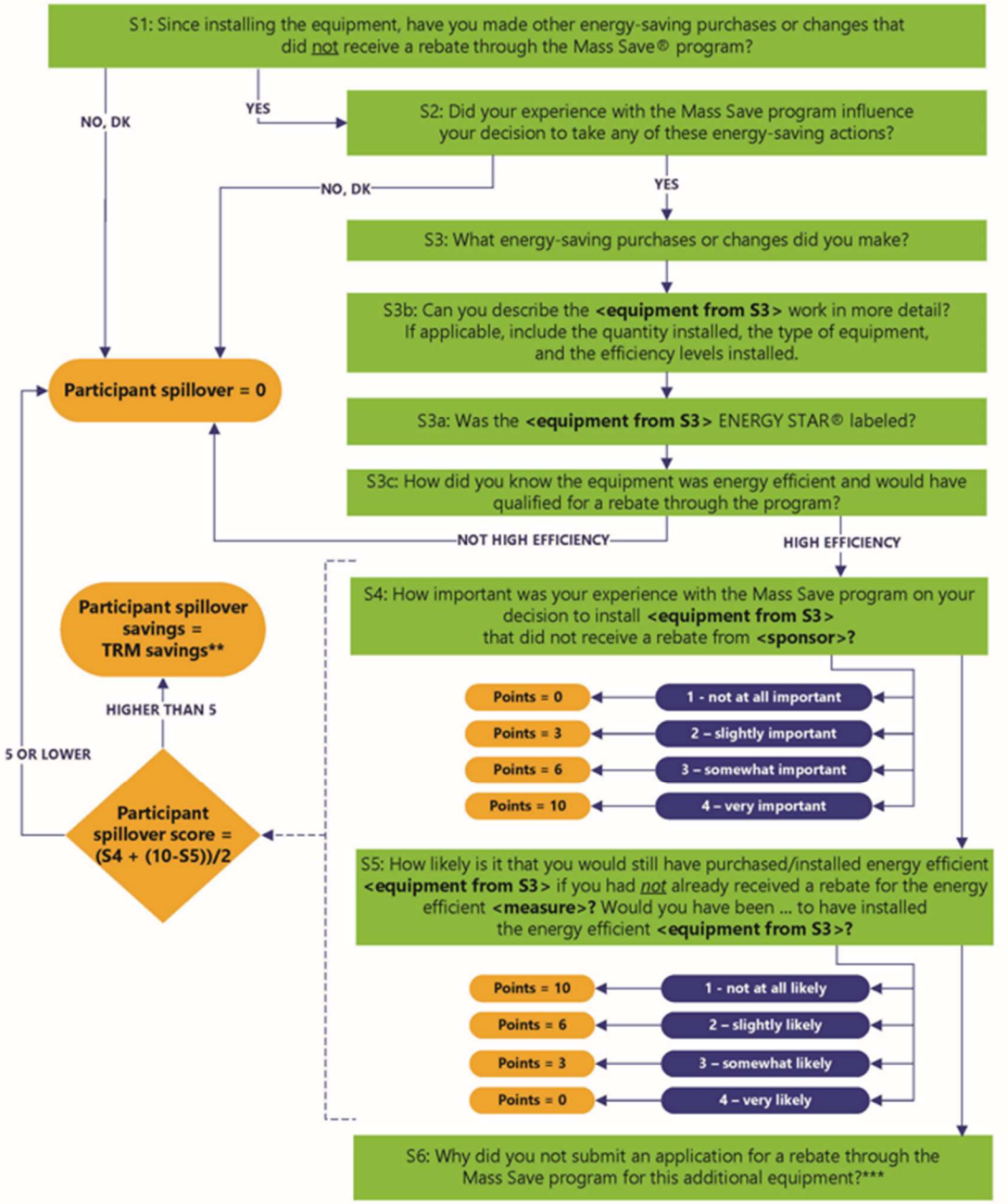
A.1 Event Type Algorithm



A.2 PARTICIPANT FREE-RIDERSHIP ALGORITHM



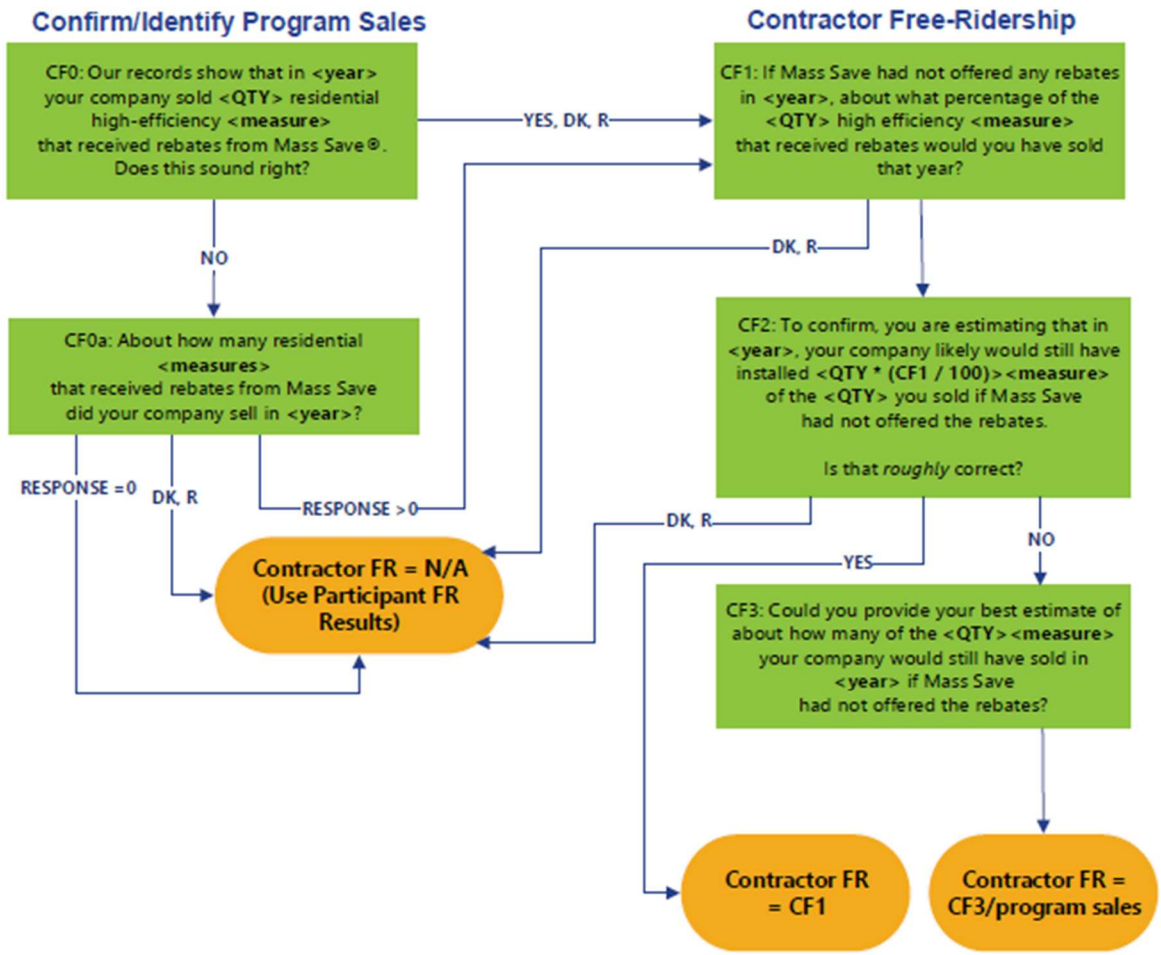
A.3 PARTICIPANT SPILLOVER ALGORITHM



* Efficiency specifications use TRM specifications for measures eligible for PA programs. For measures not offered by the PAs, the team will determine if the measures were higher than federal standards. Customers were explicitly asked if appliances were ENERGY STAR-qualified.
** Secondary sources will be used to estimate savings for non-like measures if they are not available in the MA TRM. Spillover rates are equal to the sum of the savings associated with the spillover measures divided by the sum of the savings associated with respondents participating measures.
*** Consistency check.

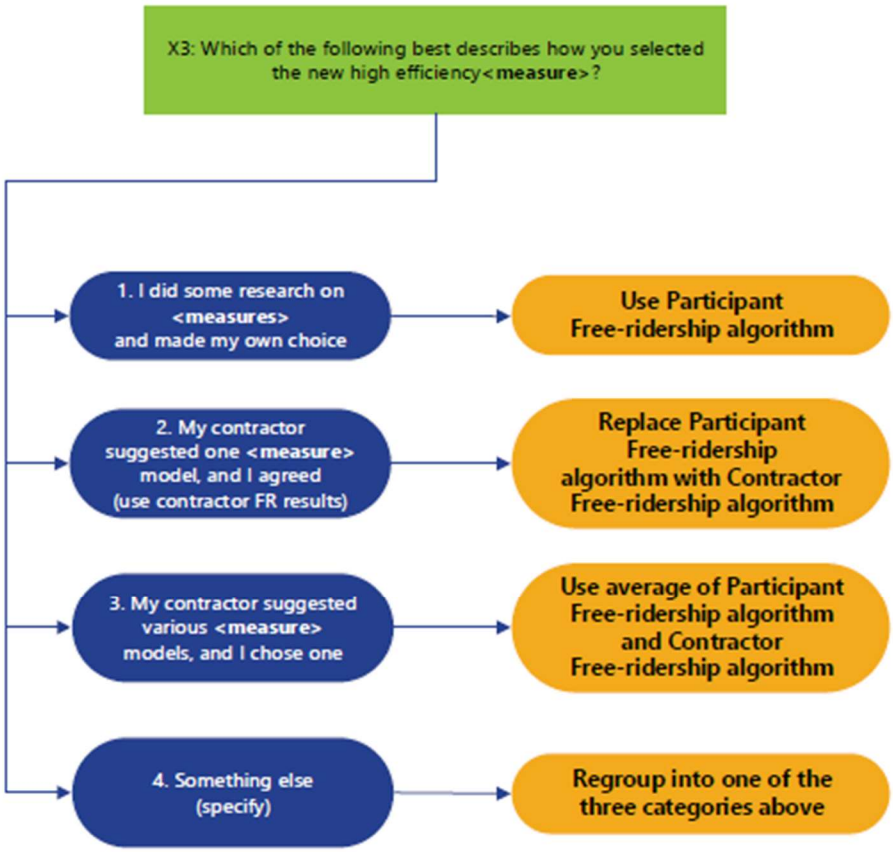
A.4 CONTRACTOR FREE-RIDERSHIP ALGORITHM

If program theory shows that the contractor/salesperson plays a prominent role in delivering and promoting the program.

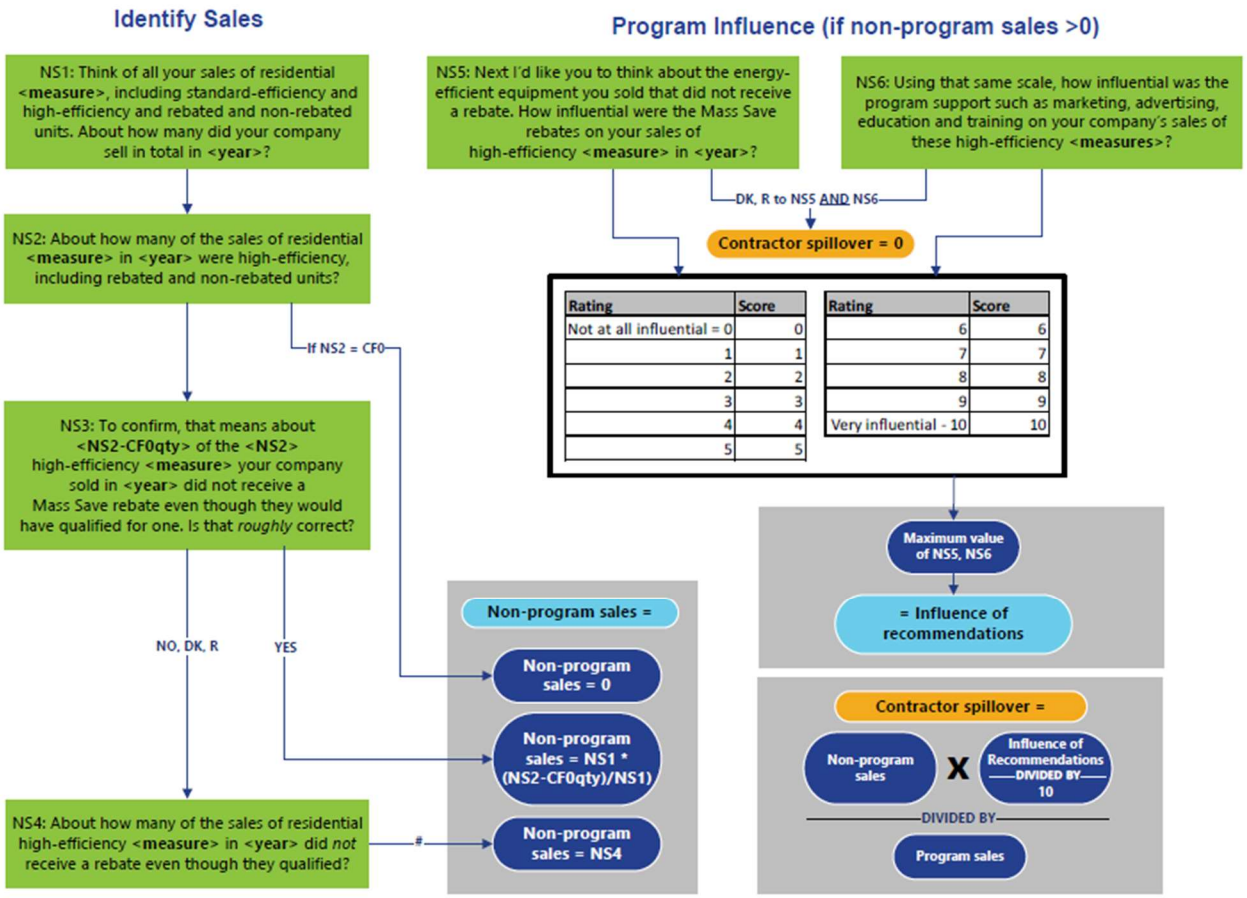


A.5 PARTICIPANT AND CONTRACTOR FREE-RIDERSHIP RECONCILIATION

When the contractor was rated the most influential program element in Participating FR and the program theory shows that the contractor pays a prominent role in the program's delivery, use this algorithm.



A.6 PARTICIPATING CONTRACTOR SPILLOVER ALGORITHM



Appendix B Appliance Recycling Free-ridership Analysis

Comparative analysis of the self-report methodology document is limited for appliance recycling programs because the NTG calculations do not follow the self-report methodology document. Instead, they follow the approach outlined in *The Uniform Methods Project: Methods for Determining Energy-Efficiency Savings and for Specific Measures*.[15] The Uniform Methods Project (UMP) provides "model protocols for determining energy and demand savings that result from specific energy-efficiency measures implemented through state and utility programs." Chapter 7 is specific to appliance recycling programs and provides guidance for estimating retrospective gross and net savings based on the respondent's actions had the appliance not been recycled through the program.

Table 49 below outlines the free-ridership scoring schematic used for refrigerators, freezers, and dehumidifiers for the Appliance Recycling NTG Update Study. While we can assess the impacts of "don't know" and "refused" survey responses on NTG ratios, we cannot perform the same response analysis level as the other three studies discussed in this report.

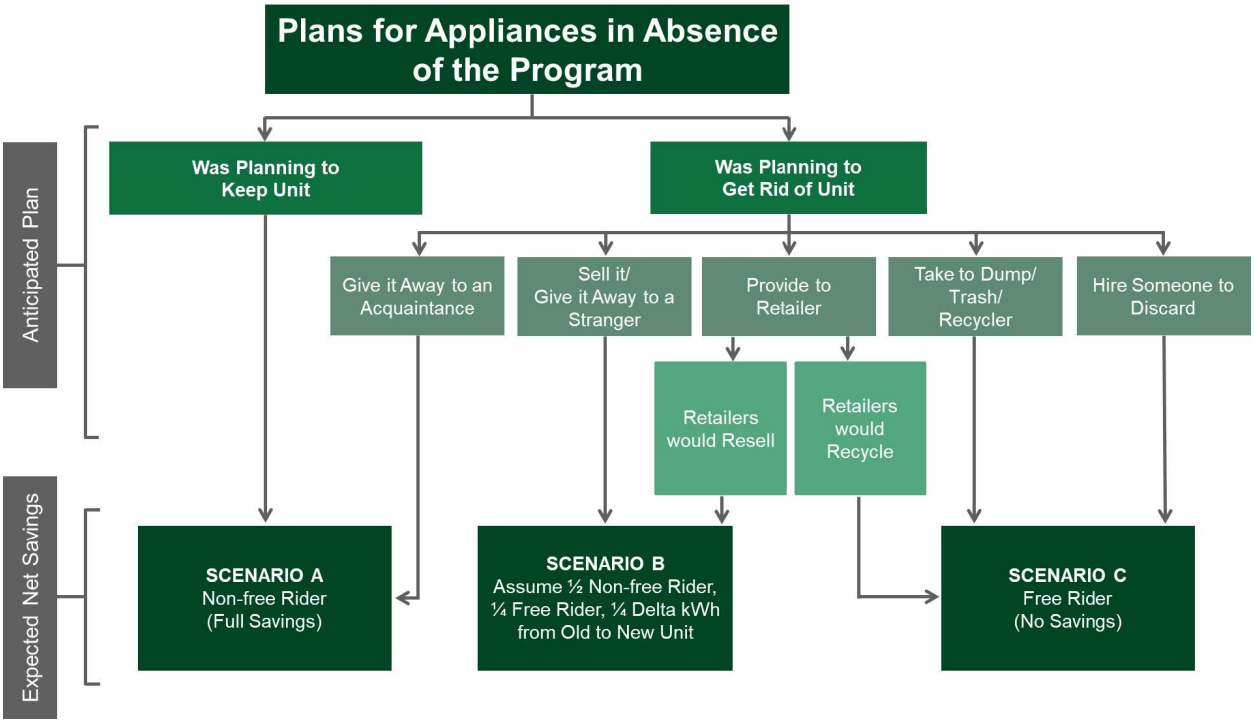
Table 49: UMP Free-ridership Scoring

Self-reported Disposition in the Absence of the Program	Free-ridership
Kept by the household	0%
Given away for free to an acquaintance	0%
Sold/given to charity, classified ads, internet, etc.	See algorithm 0% to 50%
Provided to retailer (new or used)	Fridge/freezer (if >10 years 100% FR, if <=10 years old see algorithm discussion) Dehumidifier (if >8 years 100% FR, if <=8 years old see algorithm discussion)
Surrendered to municipality/hired a hauler	100%

[15] Keeling, J.; Bruchs, D. 2017. "Chapter 7: Refrigerator Recycling Evaluation Protocol." The Uniform Methods Project: Methods for Determining Energy-Efficiency Savings for Specific Measures. Golden, CO; National Renewable Energy Laboratory. NREL/SR-7A40-68563. http://www.nrel.gov/docs/fy17osti/68563.pdf.

The appliance recycling free-ridership scoring algorithm pulled from the Appliance Recycling Net Savings Update[16] is shown below.

Figure 8: Appliance Recycling Program Free-ridership Scoring Algorithm



B.1 Free-ridership Scoring Methodology Review

The methodology reported here is pulled directly from the Appliance Recycling NTG Update Study report. As part of that report, the evaluation team reported on the impact of "don't know" and "refused" responses. This section assesses that information to show the impact of the programs reported above. Appliance recycling NTG studies before the Appliance Recycling NTG Update Study included additional contextual questions for respondents who said they did not know what they would have done had there been no program. Those questions aimed to determine if participants had tried to recycle the appliance before participating in the program and why they ultimately used the Mass Save program. The PA and EEAC working group requested these contextual questions be removed because they did not directly factor into UMP's NTG algorithm. Absent these questions, respondents who answered "don't know" or "refused" were entered into Scenario B in the algorithm (see Figure 8 above).

[16] Appliance Recycling Net Savings Update (MA20X03-E-ARNTG), July 23, 2021. https://ma-eeac.org/wp-content/uploads/MA-20X03-E-Appliance-Recycling-NTGReport_FINAL_2021.07.23-clean.pdf

To assess the impact of "don't know" or "refused" responses on free-ridership, NMR created three treatment groups. The first group excluded "don't know" or "refused" responses from the analysis entirely. The second group assigned them as non-free riders. The third group followed a modified algorithm in which "don't know" or "refused" respondents were entered into Scenario B.

Table 50: NMR's Free-ridership Analysis for "don't know" or "refused" Responses

Treatment Group	Refrigerators	Freezers	Dehumidifiers
FR classification - Method 1 (DKs removed)	44%	48%	39%
FR classification - Method 2 (DKs treated as non-free riders)	51%	58%	46%
FR classification - Method 3 (DKs treated through algorithm)	46%	50%	41%
Mean of Method 1 and 2	47%	53%	42%
Number of Don't Know's	12	14	20
Total Sample Size	96	75	154

B.1.1 Market Event Type

The market event battery is the only questionnaire similarity between the Appliance Recycling NTG Update Study and the other evaluations reviewed. While the Appliance Recycling NTG Update Study does not seek to categorize or report market event because timing is not a part of UMP's NTG methodology, it can be categorized for analysis purposes using the condition of the existing equipment question.

The survey assumes that recycled equipment is not from new construction and is replacing existing equipment. However, the survey asks about the condition of the existing equipment (ET3 in Table 2 above). From that question, we can determine if the equipment was replaced on failure (ROF) or early replacement (ER). Most equipment was defined as an early replacement: respondents reported that 83% of refrigerators, 92% of freezers, and 67% of dehumidifiers were working with minor or no need of repair.

Table 51: ARP – Condition of Working Equipment

Condition of Existing Equipment	Fridge	Freezer	Dehumidifier	Status
It was working with no need of repair	61%	76%	48%	ER
It was working with need of minor repair	22%	16%	19%	ER
It was working with need of major repair	13%	7%	9%	ROF
It was no longer working	1%	1%	18%	ROF
Don't know	3%	0%	6%	
Total	100%	100%	100%	
Respondents (n)	96	75	174	

B.1.1.1 Participant Spillover

There is no participant spillover for the Appliance Recycling NTG Update Study.

B.1.1.2 Consistency Questions

There are no consistency questions in the Appliance Recycling NTG Update Study.

B.1.1.3 Contractor Free-ridership and Spillover

Contractor free-ridership and spillover does not apply to the Appliance Recycling NTG Update Study.