Boston | Headquarters

617 492 1400 tel
617 497 7944 fax
800 966 1254 toll free

1000 Winter St
Waltham, MA 02451

# Memorandum

# AIC Custom Initiative NTG Research

**To:** Fernando Morales, AIC, and Jennifer Morris, ICC Staff

**From:** Zach Ross and Adriana Kraig, Opinion Dynamics

**Date:** September 21, 2021

**Re:** Updated AIC Custom Initiative Net-to-Gross Ratios

## Introduction

In 2020 and 2021, the evaluation team conducted research with Custom Initiative participants to update the net-to-gross ratio (NTGR) for this initiative for future application. We developed the NTGR using self-reported information from computer-assisted web interviewing (CAWI) surveys with program participants. We used participant survey responses to develop estimates of free-ridership (FR) and participant spillover (PSO). This memo presents an updated initiative-level NTGR estimate, which is summarized in Table 1.

Table 1. Updated Custom Initiative NTGR from 2020-2021 Research

| Offering | Number of Responses (n) | Free-Ridership (FR) | Participant Spillover (PSO) | NTGR (1-FR+PSO) |
|---|---|---|---|---|
| Custom (Electric) | 35 | 0.214 | 0.000 | 0.786 |
| Custom (Gas) | 10 | 0.200 | 0.000 | 0.800 |

## Data Collection and Sampling Methodology

The evaluation team fielded CAWI surveys with customers who participated in the Custom Initiative in late 2019 and 2020. The survey focused on satisfaction with program processes, and attribution (free-ridership and spillover). The sample of Custom projects came from program participants from October 1, 2019 through December 31, 2021.

The data extract included 213 unique projects. As in previous evaluations, we sampled by project contact, rather than by project, because many customers completed more than one project. These customers generally submitted the same contact name for each of the different projects. To reduce respondent burden and to facilitate question wording, we asked each contact only about the project with the largest savings. Note that we also dropped contacts for whom no valid email was available (since this was a web survey). We formed a sample frame of 155 unique customer contacts for the survey (see Table 2).

Table 2. Custom Initiative - Data Supporting 2020-2021 NTGR Research

| Initiative | Number of Survey Completes (*n*) | In Sample | Number of Projects in Population | % of Projects Covered in Survey | % of Electric Savings Covered In Survey | % of Gas Savings Covered in Survey |
|---|---|---|---|---|---|---|
| Custom | 43 | 155 | 213 | 20% | 19% | 23% |

## NTGR Overview

Net impact evaluation is generally described in terms of determining program attribution. Program attribution accounts for the portion of gross energy savings associated with a program-supported measure or behavior change that would not have been realized in the absence of the program. The share of program-induced savings, indicated as a NTGR, is made up of FR and PSO. FR is the portion of the program-achieved verified gross savings that would have been realized absent the program and its interventions. PSO occurs when participants take additional energy-saving actions that are influenced by the program interventions but did not receive program support.

The formula to calculate the NTGR is:

$$NTGR = 1 - FR + PSO$$

The Illinois evaluation teams have worked with the Illinois Commerce Commission (ICC) and the Illinois Stakeholder Advisory Group (SAG) to create a standard Illinois Statewide NTG approach for use in Illinois energy efficiency evaluation, measurement, and verification work. Per the NTG Methods attachment to the Illinois TRM,[1] all NTG data collection and analysis activities for program types covered by the attachment must conform to the statewide NTGR methods. Our survey covered all inputs required by IL-TRM V9.0. Therefore, this evaluation conforms with the requirement of Version 9 of the TRM.

## Free-Ridership (FR)

### Methodology

Free-riders are program participants who would have installed the same energy efficiency measure(s) or taken the same energy-saving actions without program support. FR estimates are based on a series of questions that explore the influence of the program on participants' purchasing decisions as well as actions the participant likely would have taken had the program not been available.

As prescribed by the Core Non-Residential Protocol in the NTG Methods attachment, we implemented two FR algorithms for the Custom Initiative evaluation.[2] The algorithms consist of three scores: (1) influence of program components (PC) score, (2) overall program influence (PI) score, and (3) no-program (NP) score (counterfactual), as well as a timing adjustment. Each sub-score serves as a separate estimator of FR and can take on a value of 0 to 1, where a higher score means a higher level of FR. The overall free-ridership score for a project is the average of the three scores, combined with a timing adjustment. Depending on the algorithm, the timing adjustment is applied to either the no-program score component or the preliminary overall FR score (average of the three sub-scores). The FR score for each project thus ranges from 0 (no FR) to 1 (100% FR).

The three scores included in the algorithms, their variations, and the timing adjustment are described below.

---

[1] Illinois Statewide Technical Reference Manual for Energy Efficiency, Version 9.0. Volume 4: Cross-Cutting Measures and Attachments. Dated: September 25, 2020.

[2] In this memo, we present results from two specifications of FR for the Custom Initiative. We select one algorithm as our choice to calculate program free-ridership and justify our choice of algorithm.

1. **Influence of Program Components.** This score is based on a series of questions that ask respondents to rate the importance of program components in their decision to install the energy-efficient equipment, using a scale of 0 to 10 (where 0 is "Not at all important" and 10 is "Very important").

   Program Components considered include such items as the availability of the incentive, information from program marketing materials, a recommendation from program staff, and previous program experience. Other components, such as payback period and previous experience with incented equipment, could qualify as program components based on responses to follow-up questions included in the survey. We estimate the Program Components score as follows:

   Equation 1. Program Components Score

$$PC\ Score = 1 - \left(\frac{PF_{max}}{10}\right)$$

   where:

   - $PF_{max}$ is the highest score given to a program factor.

   Greater importance of the program components means a lower level of FR. In this approach, if a respondent rated the program rebate 10 out of 10, the recommendation of program staff 8 out of 10, and the information from program materials 8 out of 10, $PF_{max}$ would be 10 and the PC score would be 0.

2. **Program Influence.** This score is based on a survey question asking the respondent to rate the importance of the program compared to the importance of other factors in their decision to implement the energy-efficient equipment. To do so, respondents were asked to divide 100 points between the program and other, non-program factors. This score is estimated as:

   Equation 2. Program Influence Score

$$PI\ Score = 1 - \left(\frac{Points\ Given\ to\ Program}{100}\right)$$

   More points allocated to the program means a lower level of FR. For example, if a respondent gave the program 70 points out of 100, the PI score would be 0.30.

3. **No-Program Score.** This score is based on the likelihood that the exact same energy-efficient equipment would have been installed without the program, using scale of 0 to 10 (where 0 is "Not at all likely" and 10 is "Very likely") and is calculated as follows:

   Equation 3. No-Program Score

$$NP\ Score = 1 - \left(\frac{Likelihood\ to\ Install\ Same\ Equipment}{10}\right)$$

   A greater likelihood of participating without the program means a higher level of FR. For example, if the participant provides a likelihood rating of 7 to install the same equipment in the absence of the program, their NP FR score would be a 0.70.

   In the first FR algorithm, the NP score incorporates a timing adjustment (discussed next) as follows:

$$NP\ Score\ Adjusted = \left(\frac{Likelihood\ to\ Install\ Same\ Equipment}{10}\right) * Timing\ Adjustment$$

4. **Program Timing Adjustment.** There are two ways to calculate the program timing adjustment in accordance with the NTG Methods attachment, and they incorporate information from one or two survey questions.

   - The first question asks (1) whether the installation would have occurred at the same time without the program; and (2) if the installation would have occurred later, how much later.
   - The second question asks the respondent to provide a likelihood, on a 0 to 10-point numeric scale, of installing the same energy efficient equipment within 12 months of when it was actually installed.

   The two timing adjustments are referred to as Timing Adjustment 1 and Timing Adjustment 2 and are described below.

   *Timing Adjustment 1*

   Timing Adjustment 1 uses only the first question. In this adjustment, later purchases without the program means a lower level of FR. This adjustment is calculated on a 0 to 1 scale. A timing adjustment of 1 means that there is no evidence that the program changed the time frame in which the project would have occurred, while a lower value of the timing adjustment means that the program caused the project to occur sooner. The timing adjustment provides the program with some credit for accelerating the project. Timing Adjustment 1 is calculated as follows:

   *Timing Adjustment 1 = 1 − (Number of Months Expedited – 6) / 42*

   Timing Adjustment 1 is used in the first specification of the algorithm.

   *Timing Adjustment 2*

   Timing Adjustment 2 uses both timing adjustment questions. In this adjustment, later purchases without the program means a lower level of FR, but the likelihood of implementing within a certain timeframe without the program is also taken into account. Like Timing Adjustment 1, this adjustment is calculated on a 0 to 1 scale, and a timing adjustment of 1 means that there is no evidence that the program changed the time frame in which the project would have occurred. A lower value of the timing adjustment means that the program caused the project to occur sooner. Timing Adjustment 2 is calculated as follows:

   *Timing Adjustment 2 = 1 – ((Number of Months Expedited – 6) / 42)*(10 – Likelihood of Implementing within 1 Year) / 10))*

   Timing Adjustment 2 is used in the second specification of the algorithm and is multiplied by the average of the Program Components (PC), Program Influence (PI), and No-Program (NP) scores.

This evaluation implemented and analyzed the following two FR algorithms.

- **Algorithm 1:** (PC Score + PI Score + [NP Score * Timing Adjustment 1]) / 3
- **Algorithm 2:** (PC Score + PI Score + NP Score) / 3) * Timing Adjustment 2

In the first algorithm, the NP score is adjusted by Timing Adjustment 1 and then an average of the PC score, the PI score, and the adjusted NP scores is calculated. In the second algorithm, the average of the PC, PI, and

NP scores are taken, and this average is adjusted by Timing Adjustment 2. Table 3 summarizes the differences between the two FR algorithms.

Table 3. Free-Ridership Algorithm Specifications

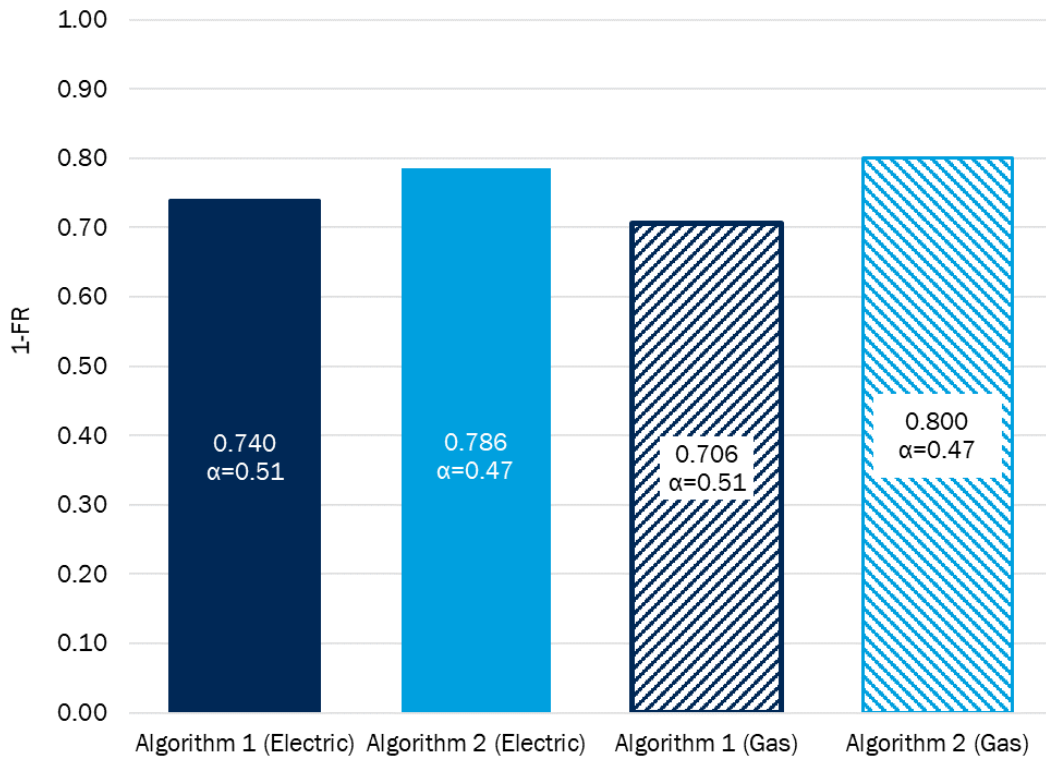
| Free-Ridership | Variant Used | | | | |
|---|---|---|---|---|---|
| Algorithm Specification | Program Component Score | Program Influence Score | No-Program Score | Adjusted No-Program Score | Overall Timing Adjustment |
| Algorithm 1 | ✓ | ✓ | | ✓ | |
| Algorithm 2 | ✓ | ✓ | ✓ | | ✓ |

We used Cronbach's alpha as a tool to help us evaluate the different algorithms, as it is a tool that examines the consistency of tests that measure the same construct. As each of the three scores incorporated into the final FR estimate serves as a separate estimate of FR, we used Cronbach's alpha to examine the internal consistency of the three scores for each specification, working from the basis that a higher degree of internal consistency is desirable for the algorithm. We also examined and compared FR results across algorithms.

Figure 1 presents FR estimates for the Custom Initiative for the two FR algorithms discussed above. The figure also shows the associated Cronbach's alphas. The evaluation team selected Algorithm 2 as our specification for reported NTGR in this evaluation (Algorithm 2 is circled in the figure below). We base this decision on our professional judgement that Algorithm 2's mathematical application of the timing adjustment is the most conceptually valid.

As we attempted a census of all participants for this evaluation, there are no error bounds around our estimates of FR for the Custom Initiative.

Figure 1. Custom FR and Cronbach's Alphas by Algorithm




n=34 for electric, n=10 for gas

We note that the Cronbach's alpha scores for this evaluation are relatively low. A general rule of thumb is that a Cronbach's alpha of 0.7 or higher indicates an acceptable level of internal consistency;[3] neither algorithmic approach produces this level of alpha.

While not relevant to the decision to choose between Algorithm 1 and 2 for application, we offer the following thoughts on this finding:

- Use of Cronbach's alpha as a tool to decide between various forms of the Illinois FR algorithms for application may have run its course. While earlier in the development of the Illinois FR algorithms, as many as six separate specifications of the algorithm, some with substantial differences, existed for analysis, only two remaining specifications now exist. Furthermore, the only difference in the component scores for the algorithms (what Cronbach's alpha compares) is the introduction of the timing adjustment into the NP score for Algorithm 1. As a result, we would expect only relatively small

[3] In measuring any underlying construct (intelligence, program influence, etc.), if you increase the number of items or questions, you increase Cronbach's alpha. Researchers in other fields, such as sociology or psychology, often rely on a battery of 10 or more questions. However, due to a number of factors such as respondent fatigue, we typically are able to ask about three to four key NTGR questions. Thus, simply due to the relatively small number of NTGR questions, our alphas are more likely to be smaller than those reported in other fields. As a result, the normal thresholds for alpha (e.g., 0.70) that are typically based on larger batteries of questions might not apply.

differences in Cronbach's alpha between specifications of the algorithms moving forward, and it is unlikely these differences will provide clear and convincing evidence of the superiority of one algorithm over another.

- Use of Cronbach's alpha as an absolute tool to assess the quality of the FR algorithms, however, continues to make sense. The low alphas exhibited by this evaluation may indicate that the algorithm subscores exhibit a substantial degree of internal variance. While it is not clear that this is a problematic finding (some theories of the NTG algorithms discussed in the Illinois NTG Working Group suggest that this is a "feature," rather than a "bug," in that variance between the independent estimates of FR leads to a more robust overall FR estimate), it is likely worth further discussion in the Working Group.

## Participant Spillover

### Methodology

Participant Spillover (PSO) refers to the installation of energy efficient measures by program participants who were influenced by the program but did not receive an incentive. An example of PSO is a customer who installed incented equipment in one facility and, as a result of the positive experience, installs additional equipment at another facility but does not request an incentive (outside PSO). In addition, the participant may install additional equipment, without an incentive, at the same facility because of the program (inside PSO).

We examined both inside and outside PSO in projects from Custom Initiative offerings using participant responses to the CAWI surveys and follow-up telephone interviews. We conducted an engineering analysis of participant responses to determine the savings associated with measures identified as SO.

After calculating the PSO savings reported by participants in our sample, we used Equation 4 to develop the program PSO rate.

Equation 4. Participant Spillover Rate

$$PSO\ Rate = \frac{Total\ Net\ PSO\ Savings_{Participant\ Sample}}{Total\ Ex\ Post\ Gross\ Program\ Savings_{Participant\ Sample}}$$

### Results

Table 4 presents the results of the PSO analysis. We did not find verifiable participant spillover in this evaluation.

Table 4. 2019-2020 Custom Initiative PSO Analysis

| Initiative | Spillover Savings | Sample Savings | Participant Spillover |
|---|---|---|---|
| Custom (Electric) | 0 MWh | 9,015 MWh | 0.0% |
| Custom (Gas) | 0 therms | 467,060 therms | 0.0% |