DNV·GL

ENERGY

RHODE ISLAND PIGGYBACKING DIAGNOSTIC STUDY

FINAL

Date January 14, 2020

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

EXECUTIVE SUMMARY

National Grid is the only investor-owned utility (IOU) in Rhode Island (RI) and serves approximately 90% of the state. National Grid is also one of the largest utilities operating in Massachusetts (MA), where it funds a substantial amount of evaluation work. Regulations and program designs are similar in both states, so historically, RI evaluations have leveraged the evaluation efforts conducted in MA ("piggybacking") out of a desire to reduce evaluation costs and when RI-specific results did not exist or were outdated. However, evaluators have done so relatively unsystematically, and have not previously tried to rigorously assess the validity of the practice. This study is an attempt to put the strategy of piggybacking on firmer ground.

The primary objective of this study is to develop guidance on when it is appropriate to "piggyback" or combine RI evaluation efforts with MA studies or adopt MA results as a proxy for RI versus stand-alone RI studies. The report recommends which approaches National Grid should use for commercial and industrial (C&I) measure groups and residential programs. Table ES-1 provides basic descriptions for the approaches.

Table ES-1. Piggybacking Approaches: Basic Descriptions

Approach Number	Approach Name	Description
1	Direct Proxy	Use MA results directly for RI
2	Shared Algorithm	Calculate savings using data collection results from MA, applied to an independent RI sample using similar formulas
3	Pooled Sample	Collect data from MA and RI sites. Create a sample from both MA and RI so that the combined sample is large enough to meet precision requirements in RI
4	Independent Sample	Conduct data collection and analysis on an independent RI sample using the same tools as MA
5	Independent Study	Conduct a completely independent study that leverages nothing directly from MA

These approaches follow a loose hierarchy of decreasing assumptions and increasing rigor as one moves from Approach 1 to Approach 5. As such, using a higher numbered approach in lieu of a lower numbered approach is usually possible and remains technically sound. In particular, any other approach could replace Approach 1. Approach 5 could be used instead of Approach 4, which could be used instead of Approach 3.

None of this report's recommendations should be interpreted as recommending the same evaluation firm conduct both the RI and MA evaluations. Issues related to evaluation firms are practical issues rather than hard requirements. Because of the pooled sampling, from a practical perspective, Approach 3 implies a single firm will conduct both the RI and MA portions of the evaluation. Also from a practical perspective, if separate firms conduct the RI and MA evaluations, they will probably not utilize Approaches 3 or 4. This is because separate (often competing) firms do not always share all of their methods. This report is neutral to these practical considerations.

Table ES-2 lists our recommended approaches by C&I measure groups. We recommend adopting Approach 4 for most C&I measure types. Most of the previous C&I evaluations used Approach 3 (pooled sample), but without adjustments made for measure mix or participant differences. Prescriptive lighting was an exception; it used Approach 5. Prescriptive gas was another exception, which used Approach 1 and Approach 3 depending on measure.

Table ES-2. Recommended Approaches: C&I Measure Groups

Measure Group	Recommended Approach
Prescriptive Lighting	**Approach 4** – Independent Sample or **Approach 5** – Independent Study
Upstream Lighting	**Approach 4** – Independent Sample
Custom Electric Non-lighting	**Approach 4** – Independent Sample
Custom Electric Lighting	**Approach 4** – Independent Sample
Small Business Electric	**Approach 3** – Pooled Sample, with adjustments for participants Or **Approach 1** – Direct Proxy if limited to non-lighting
Prescriptive Non-lighting	**Approach 4** – Independent Sample or **Approach 3** – Pooled Sample if done on individual measure types
Custom Gas	**Approach 4** – Independent Sample
Prescriptive Gas	Insufficient evidence to make strong recommendation

Table ES-3 lists our recommended approaches for residential programs. We recommend continuing to use Approach 4 for most residential programs. In many cases, the previous residential evaulations used Approach 4. Many also utilized billing analysis or other econometric techniques, for which a pooled sample does not substantially reduce evaluation costs. The following table lists several recommendations for each program. The first recommendation listed is our recommendation if current conditions persist. Secondary recommendations include brief descriptions of situational changes that would support the decision to use that approach.

Table ES-3. Recommended Approaches: Residential Programs

Program	Recommended Approach
Lighting	**Approach 4** – Independent Samples **or** **Approach 2** – Shared Algorithm (with adjustments)
Behavioral Programs	**Approach 4** – Independent Samples **or** **Approach 5** – Independent Studies
EnergyWise Single Family	**Approach 4** – Independent Samples **or** **Approach 5** – Independent Studies **or** **Approach 3** – Pooled Sample (if no billing analysis & next study shows similar results for RI and MA)
Residential Cooling & Heating	Insufficient evidence to make strong recommendation
Consumer Products	Appliance Recycling: **Approach 2** – Shared Algorithm **or** **Approach 3** – Pooled Sample (if field data collection used) Other Measures: **Approach 1** – Direct Proxy
Income Eligible Single Family	**Approach 4** – Independent Samples **or** **Approach 5** – Independent Studies; **Approaches 1, 2, or 3** (if next study has similar results for RI and MA)
EnergyWise Multi-family	**Approach 4** – Independent Samples **or** **Approach 2** – Shared Algorithm (if not using billing analysis)
New Construction, Code Compliance, and Building Characteristics	**Approach 4** – Independent Samples **or** **Approach 5** – Independent Studies
Demand Response Programs	**Approach 4** – Independent Samples **or** **Approach 3** – Pooled Samples (if small participant population or constrained data)

An overarching recommendation that is primarily applicable to the residential studies reviewed in our meta-analysis is that evaluators should always report precisions or variance statistics (standard error or standard deviation) for final evaluation metrics such as realization rates. Not only do these statistics help place the findings for that study in better context, they facilitate cross-study comparisons in the future.

Method

To generate these recommendations, DNV GL completed the following activities:

- Compared and analyzed data from National Grid's available RI and MA tracking and billing data, the American Community Survey (ACS), and the Bureau of Labor Statistics (BLS)
- Interviewed RI program staff
- Conducted a meta-analysis of 75 previous RI or MA studies.

Limitations

The study attempted to utilize all information that was available during the analysis period. Not all information types were available for all C&I measure groups and residential programs. For example, some of the residential studies did not list confidence intervals or error values, so DNV GL could not utilize statistical meta-analytic techniques on them. We also had only high-level summaries of RI residential tracking data. National Grid produces new studies on a regular basis, and some of the most recent studies were not completed in time for this study to utilize the information within them.

The recommendations in this study should be interpreted as technical guidelines. While this study describes the evaluation cost savings for the different approaches and considers program size as a factor in our recommendations in several places, the recommendations can never factor in all possibilities that might be relevant in the future. The recommendations here are made mostly from a technical and evaluation rigor perspective. Many recommendations call for activities that will increase evaluation costs. This study is meant to provide guidance to National Grid and the Rhode Island Energy Efficiency and Resource Management Council (RI EERMC) from the technical and rigor perspective to help them make final decisions about balancing increased costs, rigor, and other contextual and practical considerations.

Disclosure

To maintain full disclosure, DNV GL is one of National Grid's evaluation contractors. An unintended outcome of this study is to recommend more expensive evaluation methods, which DNV GL could benefit from. However, we believe the recommendations in this report are supported by objective evidence.

1 INTRODUCTION

National Grid is the only investor-owned utility (IOU) in Rhode Island (RI) and serves approximately 90% of the state. National Grid is also one of the largest utilities operating in Massachusetts (MA), where it funds a substantial amount of evaluation work. Regulations and program designs are similar in both states, so historically, RI evaluations have leveraged the evaluation efforts conducted in MA ("piggybacking") out of a desire to reduce evaluation costs and when RI-specific results did not exist or were outdated. However, evaluators have done so relatively unsystematically, and have not previously tried to rigorously assess the validity of the practice. This study is an attempt to put the strategy of piggybacking on firmer ground.

This report presents results of DNV GL's analysis of National Grid Rhode Island's practice of leveraging MA energy efficiency evaluation efforts to supplement and/or reduce the cost of RI energy efficiency evaluation efforts. The practice is colloquially referred to as "piggybacking". This study was completed by DNV GL for National Grid and the Rhode Island Energy Efficiency and Resource Management Council (RI EERMC) to provide guidance to National Grid RI to determine under what conditions is it appropriate to leverage Massachusetts (MA) energy efficiency program evaluation efforts or to conduct completely separate RI studies.

Study Goal and Objectives

The goal of this study is to develop guidance for National Grid Rhode Island concerning when it is appropriate to leverage MA energy efficiency program evaluation efforts to supplement and/or reduce the cost of RI energy efficiency evaluation efforts.

To achieve this research goal, DNV GL completed the following research objectives:

1. Conducted interviews with National Grid staff to identified similarities and differences in MA and RI codes, programs, populations, implementation practices, and evaluation practices;
2. Assessed whether there are differences in demographic and firmographic characteristics of the population of MA and RI customers and participants that impact the ability to leverage MA evaluation results for RI evaluations;
3. Analyzed similarities in methods and findings for past evaluation studies that cover RI and MA.
4. Provided guidance on when piggybacking is justified and suggest which of several different approaches to piggybacking are appropriate, by measure category.

Study Milestones

DNV GL, National Grid, and the RI EERMC agreed to a revised work plan in July 2018. We issued a data request for RI program tracking and billing data on July 27, 2018. DNV GL received RI C&I and residential billing data on September 24, 2018. DNV GL received savings by measure type tables for residential programs in August 2019. From past evaluations with National Grid Rhode Island, DNV GL already had C&I tracking data for RI. We had access to MA billing and tracking data for both C&I and residential customers through DNV GL's MA Customer Profile studies.

In September 2018, DNV GL delivered an interim memo describing demographic differences and an initial review of the originally identified list of previous evaluation reports to meta-analyze. Responses to this initial deliverable redirected the project to focus more on similarities/differences of program participants rather than state populations and increase the use of past evaluation results. This feedback resulted in the addition of approximately 20 studies to the meta-analytic task.

DNV GL presented a set of general recommendations in December 2018. In response, National Grid and the RI EERMC requested more specific advice for each of the major measure groupings for C&I and Residential programs.

DNV GL received contact information for C&I program managers in May 2019. We conducted interviews with those staff on May 22nd. We received contact information for residential program managers in July 2019 and conducted those interviews on July 23rd and 25th.

DNV GL provided a draft report to National Grid on July 31, 2019. National Grid asked for extensive revisions to that report. A version of the report incorporating those revisions was sent to the EERMC in October 2019. This version includes revisions based on additional National Grid and EERMC comments to the October version.

Overview of Report

The remainder of the report is organized into the following sections:

- ***Piggybacking Approaches*.** Describes the different piggybacking approaches considered, strengths, limitations, and when to use them
- ***Methods*.** Describes the activities conducted to complete the objectives.
- ***Findings*.** Presents the results of the interviews with National Grid staff, then reports detailed commercial and residential findings. Each of the commercial and residential findings subsections has several divisions:
  - Results of in-depth interviews relevant to policy context
  - Comparisons of economic and demographic data
  - Comparisons of billing data, tracking data, and past evaluation results by major measure category
- ***Appendices*.** Contains additional detailed information on our methods and detailed residential demographic differences.

2 PIGGYBACKING APPROACHES

This report identifies measure groups and programs for which different forms of piggybacking is justified. It suggests which of several different approaches are appropriate and provides recommended steps to take to take when implementing a recommended piggybacking approach. The goal is to ensure the evaluation results are representative of RI, even when they leverage information from MA. To be representative of RI, MA results sometimes must be adjusted to account for known differences in the participant populations, measures installed, and other differences identified by this study that could produce differing evaluation results between MA and RI.

DNV GL's recommendations are based on the analysis of four sets of information:

- National Grid's billing and efficiency program tracking databases allowed for examination of population characteristics by measure type, program, and other key firm-o-graphic characteristics,
- Secondary research provided by the US Census and Bureau of Labor Statistics allowed for comparison of demographics and trends in key economic indicators between RI and MA over time,
- Results from interviews with National Grid program and evaluation staff identified similarities and differences between populations, programs, and implementation and evaluation practices that may influence the appropriateness of each recommended approach for a given measure group and program, and
- Examination of past impact evaluation results for RI and MA to determine whether impact results are statistically similar or different.

Figure 2-1. Overview of study methods

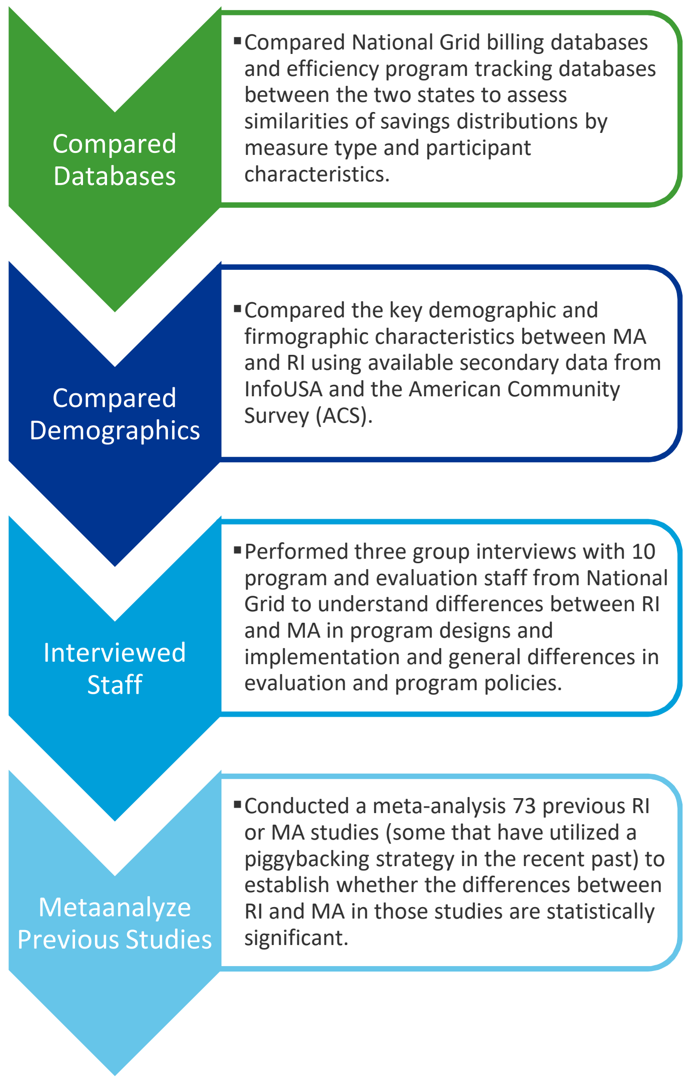


Figure 2.1 provides an overview of the study's methods. The remainder of this section contains the following information:

- Section 2.1 – Characterizes piggybacking efforts into 5 general approaches for leveraging MA evaluation studies to produce RI evaluation results. The section also identifies the approaches employed by each measure category in previous evaluations.
- Sections 2.2 and 2.3 – Discusses DNV GL identified criteria and conditions for selecting a given piggybacking approach.

2.1 Potential Piggybacking Approaches for RI Evaluations

DNV GL has identified the following 5 possible piggybacking approaches for leveraging MA evaluation studies to produce RI evaluation results:

- **Approach 1: "Direct Proxy"** apply MA-only evaluation results directly to RI
- **Approach 2: "Shared Algorithm"** apply parameters estimated from MA-only sample data to RI-specific sample frame and algorithms
- **Approach 3: "Pooled Sample"** use a sample that includes sites from both MA and RI and pools the results to achieve required statistical precisions in RI. Results might be reported by state, but RI uses the pooled result.
- **Approach 4: "Independent Sample"** uses MA research design, instruments and algorithms on a RI-only sample
- **Approach 5: No Piggybacking** or a completely independent study that does not directly leverage any existing MA study.

These approaches follow a loose hierarchy of decreasing assumptions and increasing rigor as one moves from Approach 1 to Approach 5. As such, using a higher numbered approach in lieu of a lower numbered approach is usually possible and remains technically valid. In particular, any other approach could replace Approach 1. Approach 5 could be used instead of Approach 4, which could be used instead of Approach 3.

None of this report's recommendations should be interpreted as recommending the same evaluation firm conduct both the RI and MA evaluations. Issues related to evaluation firms are practical issues rather than hard requirements. Because of the pooled sampling, from a practical perspective, Approach 3 implies a single firm will conduct both the RI and MA portions of the evaluation. Also from a practical perspective, if separate firms conduct the RI and MA evaluations, they will probably not utilize Approaches 3 or 4. This is because separate (often competing) firms do not always share all of their methods. This report is neutral to these practical considerations.

For each approach, DNV GL discusses the evaluation activities used, advantages, limitations, and identifies past evaluations that have employed each approach:

Approach 1: Direct Proxy

Approach 1 applies results from an evaluation previously conducted in MA to RI. This approach borrows the MA evaluation results (often gross savings realization rate) directly to derive the corresponding overall savings metrics for RI. It does not include data collection or analysis of RI sites or savings calculations. The only RI-specific information that are considered are top-line gross savings or basic participation values. For example, this approach could apply the realization rate for a MA program to the gross tracked savings from RI to calculate gross verified savings for RI or multiply a MA savings per measure by the number of installed measures in RI.

Evaluation activities leveraged

> This approach avoids almost all evaluation activities including sampling, development of data collection instruments, data collection, and analysis.

Advantages

The primary advantage of this approach is cost savings for RI because almost 100% of the evaluation study costs are assumed by MA. Incidental costs for RI would be those associated with transferring values from the MA study.

Limitations

This approach assumes the most similarities of MA and RI programs, measures, and populations to allow them to be directly transferrable. This level of similarity is unlikely for most programs given differences in measure mixes, populations, and previous evaluation results identified in this report.

Past applications

Some previous C&I prescriptive gas studies used this approach. National Grid reported that for new measures, it tends to use MA results directly at least until there is sufficient installation volume in RI to conduct an evaluation. This practice is a variation on Approach 1.

Approach 2: Shared Algorithm

This approach applies specific parameters estimated from a MA-only evaluation to a RI-specific sample frame and sometime a RI-specific savings algorithm. In contrast to Approach 1, Approach 2 employs intermediate evaluation parameters estimated by the MA study (such as hours of use (HOU), delta-watts (ΔW), and in-service-rate (ISR)) and applies the parameters to the RI population. In some cases, RI baselines and engineering algorithms may differ from MA as well. For Approach 2, the final savings estimates from the MA studies are not used, just selected parameters. This method isolates the MA parameters that are applicable to RI, and where there is evidence of a difference (e.g. known differences in HOU) uses some other source than MA for those parameters.

Evaluation activities leveraged

This approach leverages the development of data collection tools, data collection, and possibly analytic tools.

Advantages

This approach can provide substantial evaluation cost savings over other piggybacking approaches when multiple MA parameters can be used. It allows for corrections to be made to the intermediate parameters to account for measure and population differences between MA and RI. An advantage of this approach (over Approach 1) is the individual parameter estimates are more easily adjusted for measure and population differences than overall savings estimates.

Limitations

Approach 2 relies on confidence that parameters measured during data collection are the same in MA and RI. This approach also rests on the assumption that the same savings calculations can be used for all participants. As such, this method is generally not applicable to custom programs, where each measure is essentially unique. This approach is also not applicable when billing analysis or other econometric methods are used, as those derive savings a completely different way.

Past applications

A version of this approach was previously used for the Residential Consumer Products evaluation.

Approach 3: Pooled Sample

Approach 3 involves data collection from both RI and MA participants and produces results based on the combined sample. RI uses the pooled statistics as the official evaluation results, although results are often also reported separately by state. In the past, the majority of sites in the pooled sample have come from MA, and MA results (e.g., site level savings) have been combined with RI-specific results to calculate combined results.

Evaluation activities leveraged

Sampling, data collection instrument design, and data collection.

Advantages

Approach 3 is designed to provide the necessary statistical precisions at the pooled sample level at a much lower cost than if National Grid used only a RI-specific sample.

Limitations

This approach can deliver valid evaluation results, provided the pooled sample accounts for known differences in the sample frame such as the measure mix, key demographic/firm-o-graphic characteristics, and participant consumption levels. It assumes the implementation of the program including estimation of savings methods are similar across states.

Past RI applications

Most of the previous C&I evaluations have utilized a pooled sample approach but without adjustments for differences in measure mixes or customer characteristics.

Approach 4: Independent Sample

Approach 4 leverages the MA study design and research instruments, however, those elements are applied to an independent RI-specific sample. In most cases, the RI evaluation will be managed as an entirely separate research effort. However, if conditions permit, this approach might leverage MA evaluation administrative costs.

Evaluation activities leveraged

Data collection instrument design, possibly analytic tools, and possibly project administration.

Advantages

An independent sample is the simplest, surest way to make sure that the evaluation represents RI.

Limitations

This approach is not possible in cases where RI does not have the financial and manpower resources or the participation volume to do RI-only samples. A multi-year rolling sample in RI can partially overcome this limitation.

Past applications

Most of the previous residential evaluations have used Approach 4, without rolling samples. C&I custom evaluations are in the process of switching to this approach, utilizing the multi-year rolling sample technique.

Approach 5: Independent Study

This approach implements a completely stand-alone evaluation in RI that does not leverage any evaluation activities used in MA. Strictly speaking, it is the absence of piggybacking.

Evaluation activities leveraged

None.

Advantages

Approach 5 ensures RI-specific evaluation and findings.

Limitations

This approach is usually the most expensive approach because no previous evaluation activities or products are reused. The RI Evaluation team assumes 100% of evaluation cost. However, in cases where different evaluation firms are used, this approach can sometimes be less expensive than Approaches 3 or 4 because of differences in billing rates.

Past applications

The evaluation of the 2013-2014 RI behavioral programs appears to be an independent study. The EnergyWise evaluations, and Low income single family program evaluations also used independent study approaches.

Table 2-1 provides a summary of the five piggybacking approaches and their estimated evaluation cost savings. The table identifies an estimation of how much each approach would save National Grid, relative to an independent study.

Table 2-1. Summary of Piggybacking Approaches

Approach Number	Approach Name	Description	Evaluation activities leveraged	Estimated Evaluation Cost Savings
1	Direct Proxy	Use MA results directly for RI	All	100%
2	Shared Algorithm	Calculate savings using data collection results from MA, applied to an independent RI sample	Development of data collection tools, data collection, and possibly analytic tools	35%-90%
3	Pooled Sample	Collect data from MA and RI sites. Sample from MA and RI so that the combined sample is large enough to meet precision requirements	Some sampling development of data collection tools, some data collection, and some analysis	50%-75%
4	Independent Sample	Conduct data collection on an independent RI sample using same tools as MA	Development of data collection tools and some project management	25%-50%
5	Independent study	Conduct a completely independent study that leverages nothing directly from MA	None	0%

2.2 Recommendations by Approach - When Evaluation Activities Can be Piggybacked

As a general rule, each of the following should be as similar as possible when piggybacking:

- Program designs and evaluation goals
- Program delivery
- Savings baselines and calculations
- Measure mixes
- Participant demographics/firmographics

Similarities in these qualities ensure that the MA evaluation results and methods being borrowed by RI provide results that are representative of RI populations. Non-representative results can be inaccurate, which could cause the RI programs to look better or worse than they truly are.

To facilitate specific recommendations for which piggyback approach to use, DNV GL summarizes in the below table, criteria for when to use an approach, when to use it with some corrective adjustments, and when it should not be used. A more specific discussion of our reasoning follows.

Table 2-2. Piggybacking Approaches – When to Use

Approach	Name	When to Use	When to Adjust[1]	When Not to Use
1	Direct Proxy	Programs similar Measure mixes same Low rigor acceptable		Higher rigor needed
2	Shared Algorithm	Programs similar Different measure mixes	Different baselines Different algorithms Parameter values differ	Billing analysis Custom programs
3	Pooled Sample	Programs similar Program delivery same or savings algorithms same Few RI participants	Different measure mixes Participants differ	Different baselines Different algorithms Different delivery
4	Independent Sample	Similar data collection needs Many RI participants Higher rigor needed Different program delivery	Slightly different measures or variables Few RI participants	Cost constraints
5	Independent Study	Different program designs Different data collection needs		Cost constraints Programs similar

Approach 1 (Direct Proxy) assumes that everything about the MA program and evaluation is directly applicable to RI. DNV GL recommends reserving this method for situations where low evaluation rigor is acceptable, which generally means smaller programs with more static markets. From a purely technical perspective, any of the other approaches could be used in lieu of this approach.

Approach 2 (Shared Algorithm) assumes that program designs and savings calculations are similar. It also assumes that the values for the variables in the savings calculations verified in MA are applicable to RI. By applying the calculations to a RI-specific sample or population, the approach inherently controls for some differences in measure mixes, so this is a good approach to use when such differences are known to exist. Adjustments to this method can be made to account for differences in baselines or small differences in savings calculations (e.g., one state has a variable not in the other state). This approach can include using MA parameter values for some parameters and a different source (possibly primary RI research) of values for other parameters. For example, the savings for LED lighting is generally based on HOU x ISR x ΔW. If evaluators somehow know that ISR and ΔW could be expected to be different in RI but HOU is the same or has no evidence of difference, they could use HOU from MA and some other source for the values ISR and ΔW. The more MA values that can be used, the more this approach will save on evaluation costs. Once evaluators decide to conduct primary research in RI to estimate one of the parameters, there is likely a low incremental cost to use primary research for all of those parameters. Such a research approach is better categorized as Approach 4 (independent samples).[2] This approach is not applicable when billing analysis is used because that method generally does not utilize measure-specific savings algorithms. It is also not

[1] Such adjustments might or might not be possible for specific programs.

[2] Thus, there is some gray area between where Approach 2 ends and Approach 4 begins.

applicable to custom programs because each installation for such programs can be considered a unique measure that would not conform to a standardized savings algorithm.

Approach 3 (Pooled Sample) assumes that the MA sites are representative stand-ins for RI sites. This generally requires similar program designs and delivery, baselines, and savings calculations. Custom programs are a notable exception. Because savings calculations are essentially unique to each site, custom evaluations can be thought of as evaluating the accuracy of the engineering firms' savings estimates. Thus, custom programs delivered by the same vendors would qualify for this approach. In cases where the measure mixes or participant demographics differ, adjustments can be made to this approach to ensure the MA results retain representativeness to RI. If past evaluation results are statistically significantly different between RI and MA, that suggests the MA sites would not be good representatives of the RI sites. If the evaluation results are similar, it provides evidence of representativeness and helps justify Approach 3. Future decisions whether to use Approach 3 could be based on comparisons of evaluation results from past studies that used Approach 3, Approach 4, or Approach 5.

Approach 4 (Independent Sample) makes few assumptions about the similarities between MA and RI. The main criterion for when to use this approach is when the data collection needs are similar in both states. This method is good when higher rigor is required and there is a large RI participant population. In cases where there are few RI participants or the evaluation is extremely cost-constrained, this method would not be ideal, but multi-year rolling samples might be used to overcome these limitations. Adjustments can be made when the programs have slightly different measures or variables, such as by making minor edits to data collection instruments and econometric models. This is a technically valid approach to use in lieu of Approach 3.

Approach 5 (Independent Study) makes no assumptions about useful similarities between the programs or evaluation approaches in each state. This is a technically valid approach to use in lieu of any of the other approaches.

2.3 Recommendations by Approach – Corrective Actions

DNV GL has identified eight characteristics that evaluators should consider when choosing a piggybacking approach. The table lists when the characteristics should be the same, where adjustments could be made if not the same, or if the approach is robust to differences in that characteristic. These are respectively labeled "Same", "Adjust", or "Robust" (Table 2-3). Details regarding specific characteristics and adjustments follow the table.

Table 2-3. What Needs to be the Same, What Can You Adjust For

Characteristic	1 – Direct Proxy	2 – Shared Algorithm	3 – Pooled Sample	4 – Independent Sample
Program design	Same	Same	Same	Robust
Measures offered	Same	Adjust	Adjust	Adjust
Savings baselines	Same	Adjust	Same	Robust
Savings algorithms or estimation process	Same	Adjust	Same	Robust
Variables in the savings algorithms	Same	Adjust	Same	Adjust
Participants' measure mix	Same	Robust	Adjust	Robust
Participants' demo- or firmographics	Same	Robust	Adjust	Robust
Previous evaluation results	Same*	Adjust	Same	Robust

*Probably not available

Program designs – Similar program designs is a basic assumption to the practice of piggybacking. If programs designs are not similar, there is little reason to believe that the evaluation results of one are applicable to another. An example of a substantial program design difference is if one program is upstream and the other program is downstream.

Measures offered – Measures offered is, to some extent, a subcategory of program design. There must be some overlap in measures offered to believe that the evaluation results of one program apply to another. Furthermore, evaluations often compute metrics on a measure level, then aggregate those metrics to the program level. This practice is followed because different measures achieve different results. Thus, significant differences in the measures offered between two programs could suggest that they are not good representatives of each other.

Savings baselines – Baselines are an integral component to calculating both gross and evaluated savings. When baselines differ, the evaluation results of one program will not be directly applicable to the other, even for the same verified installed measures. Typically, savings is calculated by multiplying a difference in consumption by hours of use (HOU) by number of measures. Difference in consumption is calculated by subtracting the consumption of the installed measure from the consumption of a baseline measure. The consumption of the baseline measure and hours of use are often specified in a TRM.

Baseline consumption differences matter when evaluators verify the consumption (or efficiency) of installed measures. All else being equal, realization rate reduces to the verified consumption difference (verified savings) divided by the tracked consumption difference (tracked savings). When the baselines differ, neither verified nor tracked savings are the same for the same installed measure. In most cases, the differences will *not* offset when put into a ratio together.

Consider the lighting example in Table 2-4 below.

- Watts (W) installed, HOU, and Number of Fixtures are the same in tracking, but baselines and therefore ∆W are different.
- Evaluators find that both sites actually installed a slightly less efficient bulb, but HOU and fixture counts were confirmed.
- Verified ∆W differs between MA and RI because of the baseline difference, and that results in a difference in realization rate of 83% versus 86%.

Table 2-4. Baseline Differences Example

State	Tracked						Realization Rate
	W installed	W baseline	∆W	HOU	Number of Fixtures	Savings	
MA	30	60	30	1000	100	3,000,000	n/a
RI	30	65	35	1000	100	3,500,000	n/a
State	**Verified**						**Realization Rate**
	W installed	**W baseline**	**∆W**	**HOU**	**Num Fixtures**	**Savings**	
MA	35	60	25	1000	100	2,500,000	83%
RI	35	65	30	1000	100	3,000,000	86%

To calculate verified savings, evaluators could verify any, or all, of the variables that go into an energy savings calculation: consumption of installed measure, hours of use, or number of measures. Differences in HOU baselines could cause similar differences in calculated realization rates when evaluators verify hours of use. To generalize: for any variable assumed to have a constant baseline in the tracked savings calculations that is then verified by evaluators, if the constant value in one state differs from the constant value in the other state, different realization rates for the same installed measure can result.

Savings algorithms and **Parameters savings algorithms** – Savings algorithms matter for similar reasons as savings baselines. When there are differences in savings calculations, it is difficult to claim that one program is representative of the other. Consider the lighting example above. If MA also included an in-service rate variable in its savings calculations and RI did not, the MA savings would not match RI savings, even for projects that have the exact same configurations in all other ways. Having the same savings algorithms is also a direct assumption leveraged by Approach 2. If algorithms differ, then one cannot simply substitute MA values in the RI equations to calculate verified savings because the equations differ. A mixed approach that uses MA values for common parameters and values determined some other way for non-common parameters is sometimes possible.

Participant measure mix – The distribution of savings by measure type matters when one tries to apply the results of one evaluation directly to another. Similar to the reason why the measures offered matters – evaluators often look at different measure types individually because evaluation results often differ by measure type. Even in the case of a custom program that is implemented by the same contractors, those contractors might have better, or worse, results with some measures types. For example, chillers might receive a higher realization rate than split rooftop systems in an HVAC program, even if they aren't reported separately. If the states had substantially different mixes of chillers and rooftop systems installed by the program, the evaluation results of MA would not be a good representation of the results in RI unless the differences in installation rates were factored in. More than the other approaches, Approach 1 (direct proxy) and the historic use of Approach 3 (pooled samples) rest on the assumption that MA sites are representative of RI sites. A substantial difference in measure mixes can indicate a lack of that representativeness, which could invalidate the use of those approaches. However, some adjustments are possible.

There are two primary methods of adjusting for such differences in Approach 3. The first is how evaluators select the MA sample that will be pooled with the RI sample. Evaluators will know the characteristics of the (usually already completed) MA sample and the RI participant population. Sites with characteristics present in MA but not present in RI can be excluded from the pooled sample. For example, MA often has much larger sites in terms of energy consumption than RI. Evaluators already often use this variable to derive stratified samples, so they can exclude MA sites that are above the threshold of site sizes (plus perhaps some additional amount to account for reasonable variance) seen in RI.

The other way evaluators can make adjustments is by post-weighting results to make the proportions of savings from specific measure types in MA similar to those proportions in RI. For example, if 50% of MA savings are from measure X and 50% from measure Y, but the distribution in RI is 25/75, evaluators can apply weights to the MA sites to make the proportional mix match RI. Evaluators are cautioned to assess any implications to statistical precision this practice could cause.

The best that could be done for adjustments for Approach 1 is post-weighting, if results are reported in sufficient detail to make this possible.

Participant demographics and firmographics – Firmographics and demographics matter primarily because they can have a strong effect on measure mixes. However, to a lesser extent, it is possible that savings will differ for the same measure in different industries, particularly when savings depend on HOU and in-service rates. We also know that large (high consumption) customers tend to achieve deeper savings than smaller customers at least over time. Thus, participant demographic and firmographic differences could lead to nonrepresentative results.

Previous evaluation results – Almost by definition, if previous evaluations for each state results are significantly different, it means that one program may not be representative of the other. The underlying reason could be because of differences in study timing, differences in any of the previously mentioned characteristics, or truly represent different responses to the program or measure performance in MA and RI. When possible, evaluators should attempt to determine what caused the differences, including reconsidering the differences as the results of more studies become available. However, this is not always possible, and the conservative approach is to assume non-representativeness. This issue particularly affects Approach 1 (direct proxy) and the historic use of Approach 3 (pooled samples) where the results from MA sites were simply combined with RI sites without special sampling or post-weighting.

The following provides a more detailed discussion of our recommended adjustments to each approach to compensate when some of the previously described differences exist.

Approach 1: Direct Proxy

Ideally, previous evaluation results would be available that show that MA evaluation results are the same as RI evaluation results. However, in situations where this approach is a possibility, it is likely there will be little or no previous data to base the decision on.

Approach 2: Shared Algorithm

The Shared Algorithm approach has a basic assumption that the algorithms to compute savings are the same in both states. Elements related to the algorithms include: the actual algorithm/formula itself, which measures the algorithm applies to, savings baselines, the other variables besides savings baselines that are in the algorithm, and different values for the variables that go into the algorithm. We describe recommended adjustments that evaluators can make when these elements are not consistent across MA and RI.

- Savings algorithms differ: Evaluators should use the RI-specific algorithms.
- Different measures offered: There are two possible situations where measures offered could differ. Either MA offers a measure that RI does not, or vice-versa. When MA offers a measure that is not in RI, there is no adjustment necessary – the evaluation simply would not use that information from MA. When there is a measure unique to RI, the evaluation would have to find some other way to evaluate that particular measure. This could take the form of using the savings calculations values from some third state in the RI-specific calculations, or possibly conducting a more rigorous evaluation of that particular measure for RI only. It is uncommon for RI to have measures not already offered in MA, but they might be installed in different proportions.[3]
- Different savings baselines: Evaluators should use the RI-specific baselines in the gross savings calculations.
- Variables in savings algorithms differ: This has similar cases as different measures offered. Either MA has variables not used in RI, in which case those variables might be able to be ignored, or RI has variables not present in MA. When there are RI-unique variables, evaluators need some other method to determine a value to assign to them. In some cases, it might be possible to use a more elemental MA variable to determine the correct value for RI. Other options are the same as for unique measures – either find another state's values to substitute in or engage in a more rigorous evaluation technique to measure that particular variable. Unique RI variables are also uncommon.
- Previous evaluation results differ: This is the most likely case where evaluators will need to adjust Approach 2. This situation would occur when previous evaluations show that each state has different values for the variables that go into the savings calculations. For example, in the case of residential upstream lighting, LED penetration rates, by room type, differ for MA and RI. Because room type is a determinant of HOU, which is one of the variables directly used in savings calculations, we expect RI will have a different value for HOU than MA. Thus, we would recommend an adjustment rather than simply using the MA value. In this case, that adjustment could still utilize information gathered in MA. One could use the room-specific HOU from MA but weight the average HOU according to the RI-specific distribution of LED penetration by room type.

---

[3] If there is an overall MA parameter estimate that is statistically sampled for MA, but includes measures not present in RI, then evaluators will have to make a judgment call about how influential those unique MA measures are on the overall MA estimate. If information to make that judgment is not available, then evaluators likely will have to balance needed evaluation rigor with the risks involved in the potentially non-representative MA parameter. It is uncommon for MA programs to include a measure that not also included in the RI program.

Approach 3: Pooled Sample

The Pooled Sample approach depends on a basic assumption that MA sites can serve as representative stand-ins for RI sites. Elements related to this assumption include: the distribution of savings among offered measures (measure mix), and participant characteristics. Adjustments to measure mix and participant characteristics should be made to ensure that the MA sites selected by RI evaluators to pool with RI sites are representative of RI. This could result in the need to sample more sites from RI than has been typically done in previous studies to achieve necessary precision estimates.

- If there is a measure or characteristic not present in RI, then those sites should be removed from the MA sample frame before the MA sites are selected. For example, for Custom HVAC, we saw that there were no sites in RI as large as the largest MA sites. Those ultra-large MA sites should be excluded from the pooled sample.
- Evaluators can also post-weight MA results to make sure they represent RI-distributions. For example, if MA gets 50% of its savings from heat pumps and 50% from furnaces, but RI gets 75% from heat pumps and 25% from furnaces, then evaluators could post-weight the MA sites, so the MA average is based 75% on heat pumps and 25% on furnaces as in RI.
- Similar post-weighting approaches can be used to weight the average savings in MA reflective of the proportions of participant characteristic (e.g. usage, industry) that occur in RI.
- In some cases, it might be possible to piggyback by specific measure rather than an entire program or broader measure category. This adjustment would require sufficient participation per measure rather than measure category, to produce samples large enough to achieve required precisions.

Approach 4: Independent Sample

This method reuses data collection instruments. Technically, the programs do not need to be the same. There just needs to be some overlap in measures and the variables in the algorithms.

- If there are unique measures in one or other state, evaluators can add/subtract a small portion of the data collection instruments for those measures, but still leverage most of the instrument.
- When there are slightly different variables needed from data collection, similar small adjustments to data collection instruments can be made.

2.3.1 Recommendations by Evaluation Activity

We also divided and considered common evaluation activities and tools into six categories. The possibility of leveraging any of these evaluation activities across states or based on previous evaluations within a given state depends on the similarity of certain situational characteristics. Table 2-5 summarizes when piggybacking on each evaluation activity is possible. The sections below describe what each activity or evaluation element is and how it should be viewed when determining when piggybacking on the activity is warranted.

- **Evaluation Design**: This includes the evaluation design and decisions regarding what types of data collection and analyses will be used for the study. This activity typically requires between 5 and 10% of evaluation budgets. Decisions regarding overall approach are based on program design and evaluation goals. Reusing overall approaches requires that these are similar.
- **Sampling**: This includes the sample design and the algorithms and code used to identify the sample. These activities typically require between 5 and 15% of evaluation budgets. Sample design decisions

depend on the specific program measures, the distribution of savings across those measures, participant demographics/firmographics, program design, and evaluation goals. These all must be similar for an evaluator to be able to reuse sampling from one state to another.

- **Data collection instruments**: This includes the methods and data collection instruments and metering equipment used to design surveys, in-depth interviews, and onsite data collection, as well as the actual programs, worksheets, and other means of recording the data collected during those activities. Generation of data collection instruments typically consumes 5 to 15% of evaluation project budgets. The design of data collection instruments is based on program design and evaluation goals. Specific program measures, the distribution of savings by specific measure type, participant demographics/firmographics, and what specific data is available from program administrator databases can also affect specific data collection instrument decisions such as how to word some questions and skip patterns. Data collection needs determine whether instruments can be reused. There needs to be some overlap in measures and savings algorithms to allow for the reuse of instruments.
- **Data collection**: This comprises the actual labor required to collect the data, including site visits, telephone calls, recording of specific metering data and internal and internet searches to acquire secondary information. Pooling samples, as has commonly been done in RI C&I studies, achieves savings in this category. These activities typically require 25 to 50% of evaluation budgets. The viability of leveraging past data collection and combining across states depends on specific program measures, the distribution of savings across those measures, participant demographics/firmographics, and whether the previous data collection instruments gathered the same information as needed for the new study. The similarity of past evaluation results also factors into whether it is prudent to leverage data collection activities. When past evaluation results are statistically significantly different, it suggests there is some fundamental difference between the two states. Averaging inter-state results in such circumstances could lead to biased evaluation results for RI.
- **Data analysis based on collected data**: This includes analytic approaches, algorithms, workbooks, code, and other tools used to analyze primary data collected as part of the evaluation data collection step. Pooling samples across years and states also saves costs in this category because the realization rates from MA and other evaluation metrics are taken directly from the previous studies rather than being recomputed. This category typically requires 15 to 30% of evaluation budgets. The viability of leveraging past data analysis depends on specific program measures and whether the previous data collection instruments gathered the same information as needed for the new study. Leveraging this activity across states also requires that one is calculating the same performance metrics (e.g. annual savings or lifetime savings) and calculates the metrics the same way (e.g., use the same gross savings baselines).
- **Econometric analysis**: This includes the analytic approaches, algorithms, workbooks, code, and other tools used to conduct econometric analyses. Billing analyses and regression analyses fall into this category. When an evaluation uses econometric analysis, it typically requires between 25 and 50% of the project budget. Basic approaches (e.g. model specifications) can be reused when data structures differ, but much of the labor required for this category is in preparing the data for analysis. Furthermore, these methods often work by testing participant results to a comparison group. The comparison *is* the result, and it depends on the selection of the comparison group. Sometimes the comparison group is randomly determined at the beginning of the program, such as is common for home energy reports programs. Often, evaluators select the comparison group as part of the evaluation. In either case, Approach 2 (shared algorithm) and 3 (pooled sample) would almost never apply.

Table 2-5. Piggybacking Viability by Evaluation Activity

Similar Elements	Evaluation Design	Sampling	Data Collection Instruments	Data Collection	Data Analysis	Econo-metric Analysis
Program design	✓	✓	✓	✓	✓	✓
Evaluation goals	✓	✓	✓	✓	✓	✓
Program measures		✓	✓	✓	✓	
Savings distribution by measure types		✓		✓		
Participant characteristics		✓		✓		
Collected data					✓	
Past evaluation results				✓	✓	
Performance metrics and calculation methods			✓	✓	✓	

3 METHODS

The following provides an overview of the research approach DNV GL employed to complete this study. The approach leveraged information from the following sources to develop recommendations concerning which Piggyback approach was most appropriate for RI evaluations to adopt by measure category.

1. Analysis of National Grid billing and efficiency program tracking databases.
2. Secondary research to compare and contrast demographics and economic trends between RI and MA.
3. Comparison of past impact evaluation results for RI and MA including studies that previously employed piggybacking and separate evaluations completed in each state.

Specific research activities of this study included:

- Separating program incentivized measures and previous studies into measure categories.
- Comparing National Grid billing databases and efficiency program tracking databases between the two states.
- Compiling and comparing the key demographic and firmographic characteristics between two states (MA and RI) using available secondary data from the Bureau of Labor Statistics and the American Community Survey (ACS).
- Performing in-person and phone interviews with groups of National Grid program and evaluation staff to understand differences between RI and MA in program designs and implementation and general differences in evaluation and program policies.
- Conducting a meta-analysis on 73 existing RI or MA studies (some that have utilized a piggybacking strategy in the recent past) to establish whether the differences between RI and MA in those studies are statistically significant when considered as a whole.

3.1 Separating Measures into Measure Categories

DNV GL divided the C&I data into a series of measure categories identified after the presentation of general results in December 2018. These categories were based on a combination of input from National Grid Rhode Island, how previous evaluations divided measures, and our knowledge of how future evaluations intend to divide measures. Specific measure selection logic is documented in appendix Section 8.

C&I Measure Categories

DNV GL assigned C&I measures into each respective measure category as follows:

- ***Prescriptive Lighting.*** For the prescriptive lighting measure group, DNV GL identified records in the RI LCI tracking data that were listed as both prescriptive and lighting. For the MA comparison group, we identified records in the statewide database we compile annually that were listed as National Grid, electric, prescriptive and where end use equaled “LIGHTING”. We excluded measures that were in the C&I Multifamily Retrofit, C&I custom lighting, or C&I Small Business programs.
- ***Upstream Lighting.*** For the upstream lighting measure group, DNV GL identified records in the RI LCI Upstream Lighting data. For the MA comparison group, we identified records in the statewide database we compile annually that were listed as National Grid, electric, upstream, and where end use equaled "UPSTREAM LIGHTING”. We excluded records in C&I Multifamily Retrofit or C&I Small Business.
- ***Custom Electric Non-Lighting.*** For the custom electric non-lighting measure group, DNV GL identified records in the RI LCI tracking data that were listed as custom and not lighting, LED, or CHP. For the MA

comparison group, we identified records in the statewide database we compile annually that were listed as National Grid, electric, custom, and where end use equaled: "BUILDING SHELL" "COMPREHENSIVE DESIGN" "COMPRESSED AIR" "FOOD SERVICE" "HOT WATER" "HVAC" "MOTORS / DRIVES" "OTHER" "PROCESS" "REFRIGERATION".[4] We excluded records in C&I Multifamily Retrofit or C&I Small Business programs.

- ***Custom Electric Lighting.*** For the custom electric lighting measure group, DNV GL identified records in the RI LCI tracking data that were listed as custom lighting. For the MA comparison group, we identified records in the statewide database we compile annually that were listed as National Grid, electric, custom, and where end use equaled "LIGHTING". We excluded records in the C&I Multifamily Retrofit or C&I Small Business programs.
- ***Small Business*.** For the small business electric measure group, DNV GL identified electric records in the RI SBS tracking data. For the MA comparison group, we identified records in the statewide database we compile annually that were listed as National Grid, C&I, electric, and Small Business. This measure category includes lighting (including prescriptive lighting) and non-lighting electric measures installed under the Small Business Program.
- ***Prescriptive Non-lighting*.** This category includes all electric measures that are not listed as lighting and are not listed as being part of the custom program in the RI database or are specifically listed as being in the prescriptive program in the RI database. Specific measure types include HVAC, compressed air, hot water, food service, refrigeration, and motors/drives. For the MA comparison group, we included electric measures that were listed as prescriptive, were not lighting, and were not in the C&I Multifamily Retrofit or C&I Small Business programs.
- ***Custom Gas.*** For the custom gas measure group, DNV GL identified records in the RI LCI and SBS tracking data that were listed as gas and custom. For the MA comparison group, we identified records in the state-wide database we compile annually that were listed as National Grid, gas, custom, and where end use equaled: "BUILDING SHELL" "COMPREHENSIVE DESIGN" "COMPREHENSIVE DESIGN" "FOOD SERVICE" "HOT WATER" "HVAC" "OTHER" "PROCESS" "FOOD SERVICE". We excluded records from the C&I Multifamily Retrofit or C&I Small Business programs.
- ***Prescriptive Gas.*** For the prescriptive gas measure group, DNV GL identified records in the RI "rebate_projects" data file that were listed as prescriptive and gas. This data included funding years 2016 and 2017. Gross therms were available, but other data such as customer NAICS codes were not. For the MA comparison group, we identified 2016 and 2017 tracking records from our statewide database that were for National Grid and gas. We further filtered the MA records down to prescriptive, retrofit, and not associated with direct install or the small business program. The resulting records contained water heating measures (including pre-rinse spray valves), HVAC (including steam traps), kitchen equipment, and other (including building operator certification and building shell).

Residential Programs

National Grid provided residential tracking database savings summarized by program and major measure type within each program. These data were already summarized by National Grid into major measure types, and DNV GL did not do any additional processing on these data. The programs and major measure types for each are summarized below.

- ***Residential Lighting***. Lighting was the only measure type included in this category.

[4] These are standardized measure categories DNV GL assigns to the MA data.

- ***Residential Behavioral*** programs are comprised mostly of home energy reports.
- ***Residential Home Energy Services (EnergyWise)***. This category included lighting, appliances, envelope, thermostats, and hot water measure types.
- ***Residential Heating and Cooling Equipment*** included HVAC, hot water, and other measure types.
- ***Residential Consumer Products*** included appliances, hot water, and other measure types.
- ***Low-Income Single Family Retrofit*** included lighting, appliances, behavior, envelope, HVAC, hot water, and other measure types.
- ***Residential Multi-Family Retrofit*** and ***Low-Income Multi-Family Retrofit*** included lighting, appliances, envelope, HVAC, hot water, and other measure types.
- ***Residential New Construction*** included lighting, HVAC, hot water, appliances, and other measure types.
- ***Demand Response*** programs include billing options and some WiFi thermostats.

3.2 Compare National Grid Billing and Program Tracking Databases

When using one population as a proxy for another, it is best practice to confirm that the two populations are similar on dimensions that affect the metric in question (generally gross savings realization rates for this study). Characteristics such as measure mix, size (consumption) of participating customers, industry sector of participating customers, and the size of participating buildings are recorded in the tracking data and can have a substantial effect on gross savings.

DNV GL had access to National Grid billing and tracking data for C&I and residential customers in MA through the MA customer profile database, maintained by DNV GL. We also had access to the RI C&I program tracking data through previous evaluation work completed for National Grid. We issued a data request in July 2018 for RI C&I billing, residential billing, and residential tracking data. National Grid provided the RI C&I and residential billing data in September 2018. We received savings by measure categories for each of the residential programs in August 2019.

DNV GL divided RI C&I participation into the seven measure categories listed in the previous section: Custom Electric Non-lighting, Custom Electric Lighting, Upstream Lighting, Prescriptive Lighting, Small Business Electric, Prescriptive Non-lighting, and Custom Gas. We determined the measure types within each of these categories for RI and matched them to similar measure types in the MA tracking data. To compare the MA and RI participant populations, DNV GL aggregated the following metrics within each state's respective billing and tracking data by measure group:

- Distribution of savings by measure type
- Annual consumption of participants
- Distribution of participating accounts by NAICS code
- Distribution of participating accounts by building sizes[5]

3.3 Compile and Compare Demographic/Firmographic Information

DNV GL compared the percent distribution of various demographic and firmographic characteristics for the two states from the American Community Survey (ACS) for residential characteristics and the Bureau of Labor Statistics (BLS) for employment trends by industry sector. These analyses also helped establish the

[5] NAICS codes and building sizes were missing for approximately 30% of the data.

similarities or differences of the underlying residential and business populations in each state. The specific characteristics compared for the residential and C&I populations were originally presented in the September 2018 interim report.

3.4 Interviews with National Grid Staff

DNV GL conducted three interviews with National Grid staff in MA and RI. The interviews sought to gather information on topics that help to determine if MA results are relevant and theoretically applicable to RI:

- State policy similarities and differences
- Programs available in each state
- Designs of programs that are available in each state (measures, incentive levels)
- Evaluation practices
- Ex-ante savings calculations employed
- TRM differences (baselines, algorithms)
- Staffing and subcontractor overlaps, particularly engineers developing savings estimates

3.5 Meta-analysis of Existing RI Studies

DNV GL compiled and analyzed the results of recent evaluations that included both RI and MA customers to better understand when and where previous evaluation results differ. Appendix 7.2 lists the studies we reviewed, year of publication, and states covered by each report. We conducted the meta-analysis to determine how similar or different previous evaluation results were between the two states. As part of the meta-analysis, we also compared the similarities and differences of evaluation methods used in each state as described below:

1. DNV GL completed a high-level review of most of the studies documenting the study type (e.g., impact evaluation, market characterization, baseline), sector, measures covered, and measure program year(s) for each study.
2. DNV GL verified the states included in the study and determined whether results for MA and RI were listed separately or combined for those studies that included results for both states.
3. DNV GL conducted a more detailed review of each study and recorded which key metrics were listed in each report (e.g., tracking savings, evaluated savings, realization rate, net-to-gross ratio, etc.).
4. Following this detailed review, DNV GL again reviewed our complete list of studies to determine whether a given study's results could be combined with another study's results.
5. DNV GL flagged those studies that cover the same measures and use similar metrics to report results.

The past evaluation studies were grouped according to one of the following approaches for determining the recommended piggybacking approach for a particular measure category:

1. For studies with complete and comparable evaluation data for both MA and RI, we compared the aggregate RI to MA evaluation results reported for each respective state. This comparison required that the studies pertained to similar measures in each state and that the studies listed both an evaluation outcome and some form of statistical precision or variance estimate. Statistical difference testing used the same confidence levels used in the original report for any specific metric or finding. This category

consisted mostly of C&I studies; many of the residential studies did not provide necessary precision or variance statistics.

2. For the studies that did not have complete and comparable evaluation data for both MA and RI, but where DNV GL conducted the evaluations, we retrieved and analyzed the raw analysis files. DNV GL had raw data for most of the C&I studies. Because RI plans for future evaluations to consider broader measure groups (lighting and non-lighting), we also pooled measures that were evaluated separately in the previous studies. We were then able to compare these pooled metrics between MA and RI.
3. For those studies with results that could not be combined with other studies, but included separate results for MA and RI, we analyzed differences and similarities in measure-level results for RI and MA. We also looked closely at methodological similarities and differences for studies in this category. Most residential studies fell into this group.

The next two sections present the findings. First we present the findings for C&I, starting with the results of our interviews with National Grid staff, then moving to economic trends, then in-depth review of measure category differences and comparisons of results of previous C&I studies. Next, we present residential findings. These include interviews with National Grid staff, demographic differences, and comparisons of the results of past residential studies.

4 FINDINGS - C&I

4.1 Program Design and Policy Context

DNV GL conducted in-person interviews with C&I program and evaluation staff to identify similarities and differences between RI and MA that may impact the relevance of piggybacking approaches. Overall, the interview findings imply that evaluators should exercise caution when using piggybacking methods that do not involve an independent RI sample. However, similarities in program designs increase the validity of leveraging techniques first established in MA. Table 4-1 provides a summary of the interview results and highlights for C&I.

Table 4-1. Summary of Program Design and Policy Interviews: C&I

Research topic	Finding	Implication
Codes/ baselines	The PAs report codes are one of the biggest ways MA and RI differ. In the past the codes were more similar, but now MA code is more than one cycle ahead of RI. Many baseline codes are different: MA is ahead in terms of their code dictated baselines by one cycle. RI is operating under 2012 IECC, while MA is operating under IECC 2015. MA will be adopting IECC 2018 baseline, while RI will be moving to IECC 2015 in 2018. Note that code only applies to new construction, major renovation or end of useful life. MA has adopted amendments to strengthen codes relative to IECC standards, while RI has adopted weakening amendments. MA also has a stretch code established by the Green Community Act, which RI does not have. Many buildings adopt the more efficient stretch code. The MA PAs still offer incentives for code as opposed to stretch code, so this does not impact the baseline, but receive additional credit if customers adopt the stretch code.	Baseline differences make it difficult to leverage MA evaluation results for RI for programs based on code dependent measures such as new construction. This suggests that leveraging the MA evaluation approach but conducting a separate RI evaluation are more appropriate approaches to piggybacking than direct use of MA evaluation results for RI evaluations. For instances in which RI leverages MA evaluation results for measures that exist in MA but are new to RI, results should be adjusted to reflect differences in code.
Savings computations	The algebra for gross savings is similar, but the baselines are different. MA has a dual baseline and is one cycle ahead of RI in terms of the baseline level for measures dependent on code compliance.	Dual baselines does not affect first year savings, which is what previous evaluations have reported.
Net savings	The states have different net-to-gross (NTG) survey cycles causing the net savings to be different. The last NTG survey in RI was in 2016 and is run approximately every 3 years. NTG results are used only prospectively in RI and in MA. MA can apply new evaluation results retrospectively, provided they are not NTG (i.e. if results come in during the planning cycle).	Previous impact evaluations have not reported on net savings. For future net savings piggybacking considerations, retrospective results from MA should not be applied to RI prospectively. Evaluators need to consider the timing of NTG studies to determine whether they can be leveraged prospectively.

Research topic	Finding	Implication
Planning cycle	MA files plans every 3 years, while RI files 3-year plans and annual plans. Annual plans might provide RI with more flexibility than MA to change programs which may impact the comparability of programs and measures.	Measure mixes for the same programs could vary substantially. When measure mixes differ, they can be adjusted for in sampling and/or post weighting when using pooled samples approaches. Measure mix differences based on tracking data are reported for each individual C&I measure type in the subsections of 4.3. This is one factor that may impact the measure mix in an evaluation and the ability to leverage results directly or pool samples from MA evaluations. Substantial year over year changes to the measure mix in RI will dilute the relevance of MA evaluation study design for RI.
Savings goals	MA uses lifetime savings for goals, while RI uses annual savings. RI may be switching to lifetime savings in the future.	The different savings goals can impact the measures installed in each jurisdiction. Implementers are incentivized based on annual savings in RI allowing them to focus on higher annual savings measures that might not result in greater lifetime savings. MA implementors focus on lifetime savings. If there are large differences in the measure installation mix, it can substantially limit the relevance of MA evaluation results for RI. Differences in measure mix should be taken into account when pooling samples.
Programs and measures	The programs themselves and measures covered are nearly identical. Both states have the same upstream, retrofit, small business, and custom programs as well as the same appliance and equipment standards. They also use the same approach for determining end of useful life. They also use the same screening tool for custom measures but do have differing assumptions due to differences in BCR test benefit streams planning cycle, baselines, and goals.	This improves the ability to use MA study design for RI evaluations. Depending upon whether other conditions regarding measure mix, codes, and planning cycle are met, will determine whether pooling samples from MA evaluations or independent evaluations that leverage MA techniques are appropriate.
Service territories	The similarities and differences in customer base depend on the region of each state. For example, according to one interviewee, "in Worcester, where National Grid is the electric utility, the customers are more similar to RI than in Boston where National Grid is the gas provider."	Regional differences should be taken into account when deciding to pool samples or not.

Research topic	Finding	Implication
Economic Benefits / incentives	RI uses a ratio of 0.57*spend to estimate additional economic benefits from measure installation, making it much easier for projects to meet cost-effectiveness tests than in MA.	Use of economic benefits for cost-effectiveness tests could impact the measure mix within a program.
Custom studies	Custom projects will depend on how well the savings calculation vendors perform. There should not be much difference since they are mostly the same vendors.	No impact.
TRM	The MA TRM is more detailed. There are differences in the numbers reflected in the state specific evaluations, but the use of a different TRM is not an important difference, given many of the measures are the same and the basic algorithms are similar.	No impact.

4.2 Economic Trends

Population-level firmographic comparisons between RI and MA are more difficult to obtain than residential demographic differences.[6] In lieu of such population-level firmographics, DNV GL analyzed differences in economic trends in each state. To the extent that such economic trends affect program participation, these trends could reflect differences between the two states that would cause MA to be a poor representative of RI.

This section focuses on economic growth. When the economy or a business is growing, it might have different priorities than when it is shrinking. A shrinking economy means businesses are not expanding and therefore probably not investing in new construction. Participation in new construction efficiency programs would be expected to decrease during such times. Likewise, in a shrinking economy, businesses probably have less cash flow available to invest in capital improvements and thus might be less likely to invest in retrofit efficiency measures as well.[7] In contrast, in a growing economy more new construction can be expected, and cash flow probably allows for the consideration of capital improvement projects.

This section summarizes economic trend data reported by the Federal Bureau of Labor Statistics (BLS). These data include unemployment rate, gross state product, and job growth trends by key industry sectors. Employment rates are easy to obtain and generally considered to correlate with economic growth. The industry sectors reported by the BLS are similar to NAICS codes but are not exactly the same.

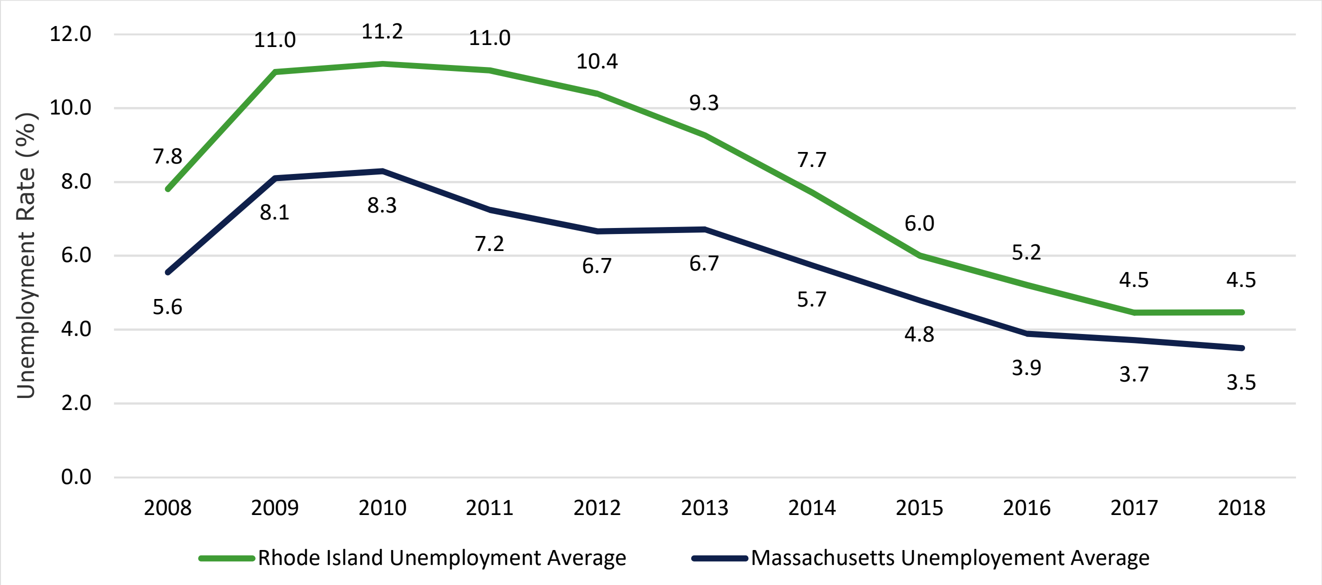
In general, the MA economy has grown faster over the past 10 years than the RI economy, but the overall year-to-year trendlines are parallel. This growth is not universal, however – there are some industries where RI growth is greater than MA and where the year-to-year trendlines are substantially different. The industries where trendlines are substantially different are the ones where evaluators should exercise the most caution when pooling MA and RI samples or using MA results as a proxy for RI results.

[6] Where National Grid billing or tracking data contained such information such as NAICS code or total annual usage, we factored it into the measure group comparisons presented in section 9.3.

[7] On the other hand, in business sectors where energy is a major cost, they might be more interested in retrofit programs as a means to drive down their costs.

For the past 10 years, unemployment trends have been very similar in each state, although overall unemployment rates are higher in RI than in MA (Figure 4-1). Both states have been at or near "full employment" since 2016.

Figure 4-1. Unemployment Rate Comparison



Despite the parallel unemployment trends, MA has experienced more rapid economic growth since 2010 (Figure 4-2). MA gross state product (GSP) has increased by an average of 2.1% per year since 2010 while RI's GSP has increased by an average of 0.8% per year.

Figure 4-2. Gross State Product Comparison

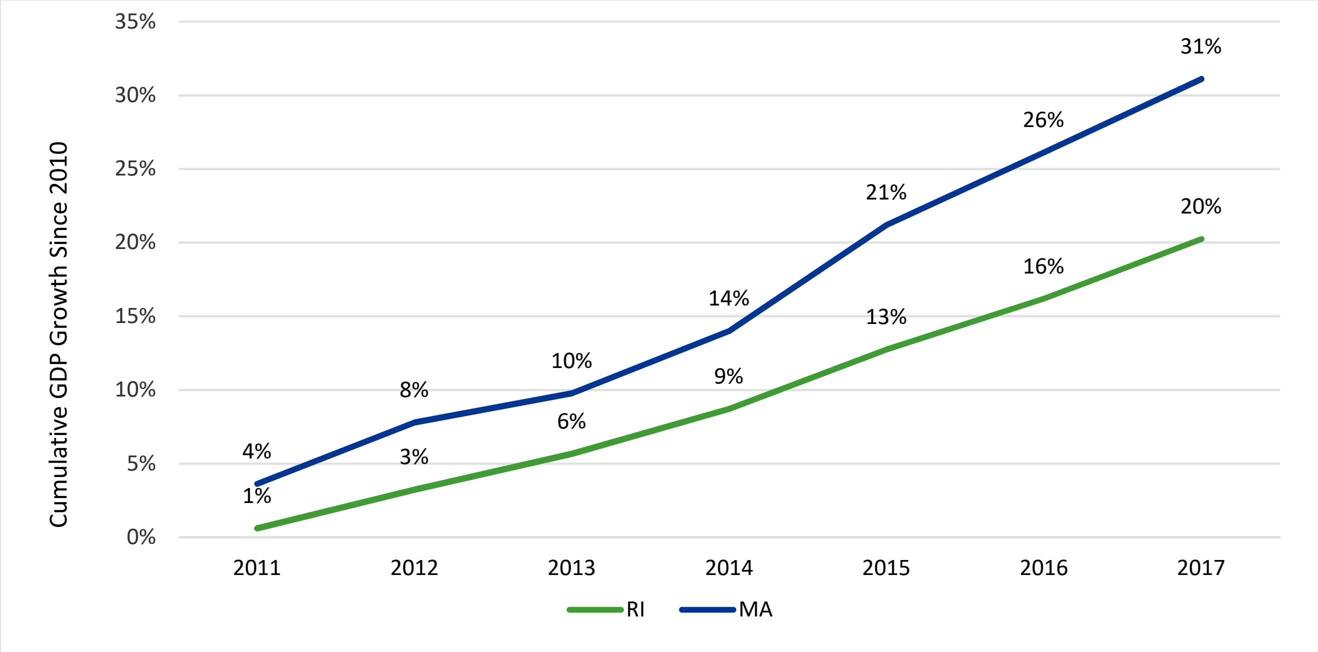
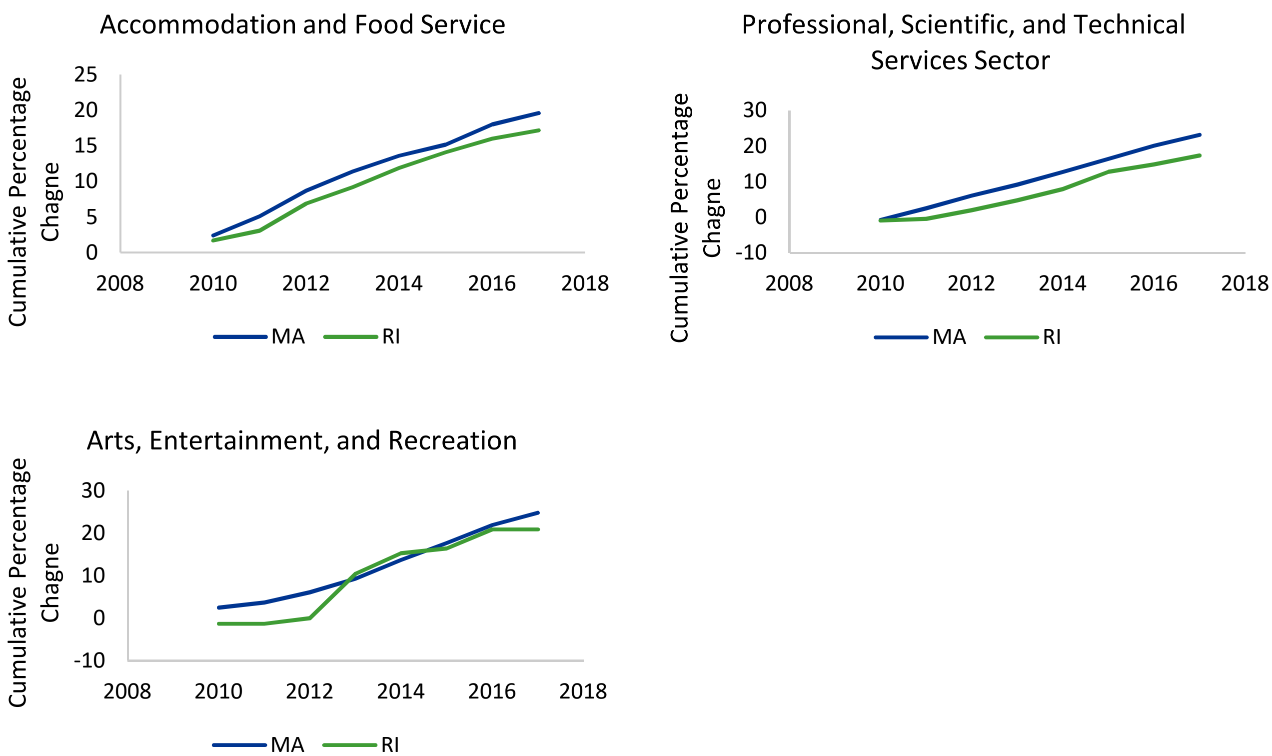


Figure 4-3, Figure 4-4, and Figure 4-5 show the job growth trends in RI and MA reported by the Bureau of Labor Statistics (retrieved Feb 04, 2019) for 2010 through 2017 for the most commonly occurring NAICS codes for participants in National Grid's efficiency programs. The industries shown are based on the two-digit super-categories provide by the BLS. They approximate two-digit NAICS codes. The trends are shown as cumulative annual change since 2009. The growth trend comparisons fall into three categories:

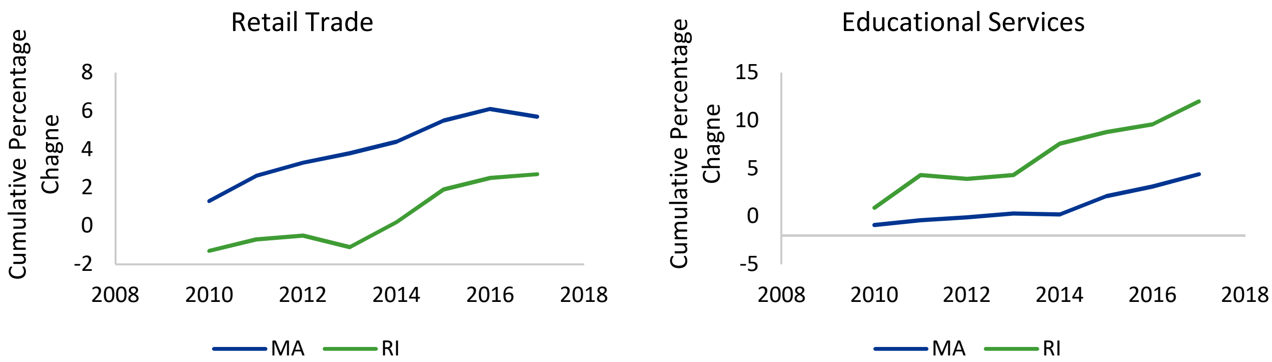
Industries where the trends are very similar between the states (Figure 4-3). These include Accommodation and Food Service; Professional, Scientific, and Technical Services; and Arts, Entertainment, and Recreation.

Figure 4-3. Industries with Similar Growth Trends



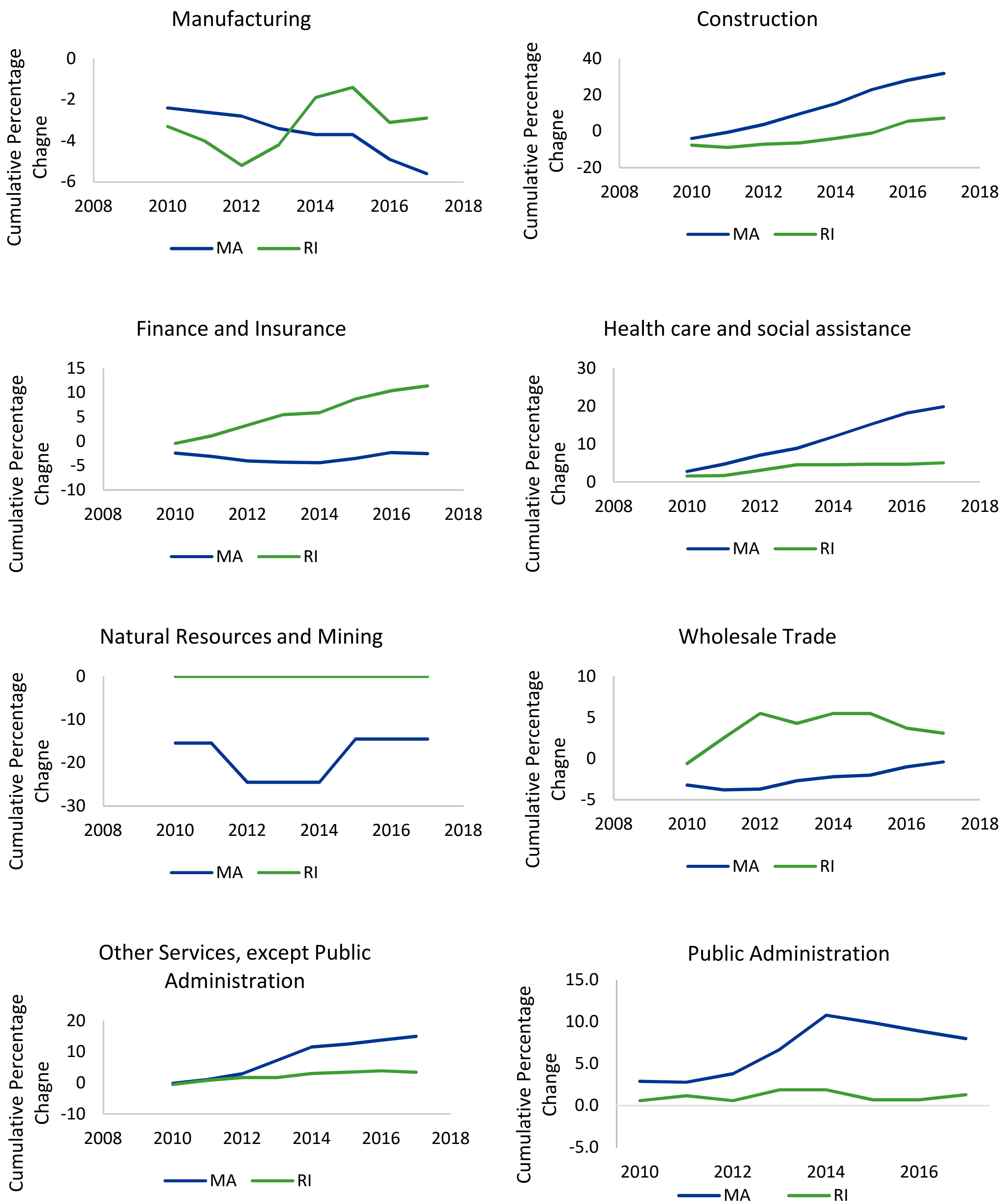
Industries where the trends go in a similar direction, but one state has substantially greater/less growth than the other (Figure 4-4)**.** These include Retail Trade and Educational Services. Note that RI is growing more quickly in the education services sector.

Figure 4-4. Industries with Similar Growth Trends, but Different Magnitude



Industries where the trends diverge and do not look similar (Figure 4-5)**.** These include Manufacturing, Construction, Finance and Insurance, Health Care and Social Assistance, Natural Resources and Mining, Wholesale Trade, and Other Services and Public Administration. The odd shape for Natural Resources and Mining is due to small sample sizes. There is very little resource extraction happening in either state. Wholesale Trade represents sales to retailers and distributors rather than directly to end-consumers.

Figure 4-5. Industries with Divergent Growth Trends



4.3 Comparisons by Measure Category

Table 4-2 presents the total proportion of kWh savings accounted by C&I measure categories for National Grid in RI and MA for 2015-2017. RI savings are slightly more concentrated in prescriptive and upstream lighting than in MA. However, a chi-square test indicates that the variation in distribution of total kWh savings across measure groups was not statistically different between both states. For gas programs, approximately 72% of 2015-2018 therm savings in RI came through the custom program. The other 28% came through prescriptive.

Table 4-2. Proportion of Total National Grid Electric Savings by C&I Measure Category

Measure Category	RI % Total kWh Savings	MA % Total kWh Savings
Downstream Prescriptive Lighting	25%	19%
Upstream Lighting	21%[1]	20%
Custom Electric Non-lighting	20%	19%
Custom Electric Lighting	14%	18%
Small Business Electric	13%	15%
Prescriptive Non-lighting	7%	10%
Total	**100%**	**100%**

4.3.1 Downstream Prescriptive Lighting

Recommended Evaluation Approach

DNV GL recommends that future evaluations use Approach 4—Independent Sample to obtain statistically robust results for an independent RI-specific sample. Approach 5 could also be used. This recommendation is based on:

- The program and measures are similar, so Approach 5 (independent studies) is not necessary.
- Previous evaluation results for lighting systems differ, so Approach 1 (direct proxy) and Approach 3 (pooled sample) are not recommended.
- Distributions of participating customers in terms of size and industry differ, which could lead to differences in the parameters such as HOU, ISR, and ∆W that determine lighting savings calculations. Therefore, Approach 2 (shared algorithm) might not result in substantial evaluation cost savings.
- The previous study is from 2011, the lighting market has changed substantially since then and is rapidly evolving, and this program has the greatest proportion of C&I savings. Thus, more conservative and rigorous approaches are justified, so Approach 4 (independent samples) makes sense over Approaches 2 or 3.

Program Comparisons

Figure 4-6 shows how the proportion of prescriptive lighting (reported gross) savings are distributed by measure type across the two states. Both states see the majority of their consumption savings fall under the linear and other LED (not screw-based) measure category. RI is achieving a greater share of program savings than MA from linear and other LED (not screw-based), and a lesser share from screw-based lamps.

A Chi-square test, which tests the relationship between categorical variables, indicates that the measure mix is statistically different at the 90% level.

Figure 4-6. Proportion of Reported Gross Savings by Measure for Prescriptive Lighting

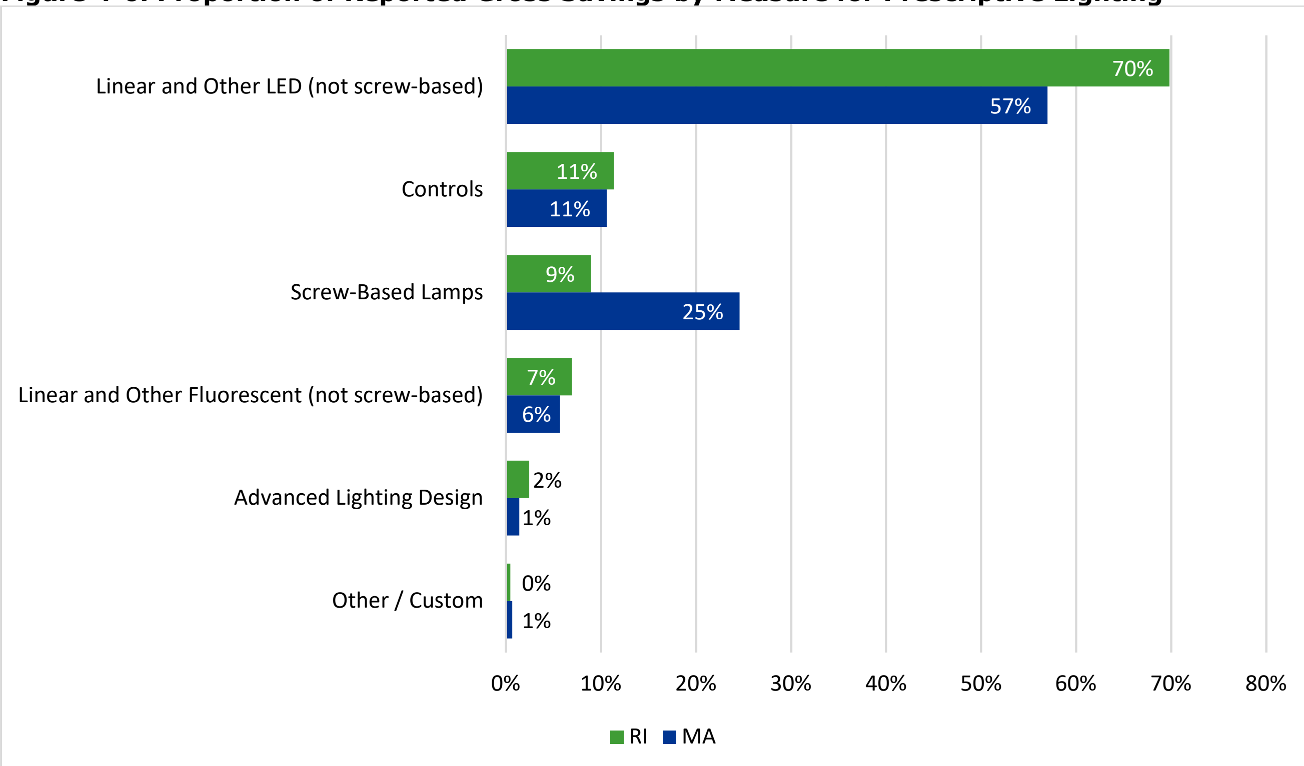


Figure 4-7 shows the median annual consumption of RI participants is consistently greater than that of MA participants between 2012 and 2017. Differences in the medians of the two states are not driven by differences in the largest consumers but rather by a top-heavy distribution of participants in RI relative to MA. This is a key finding for our recommendation.

Figure 4-7. Median Annual Consumption Over 2012-2017 by Participation Year for Prescriptive Lighting

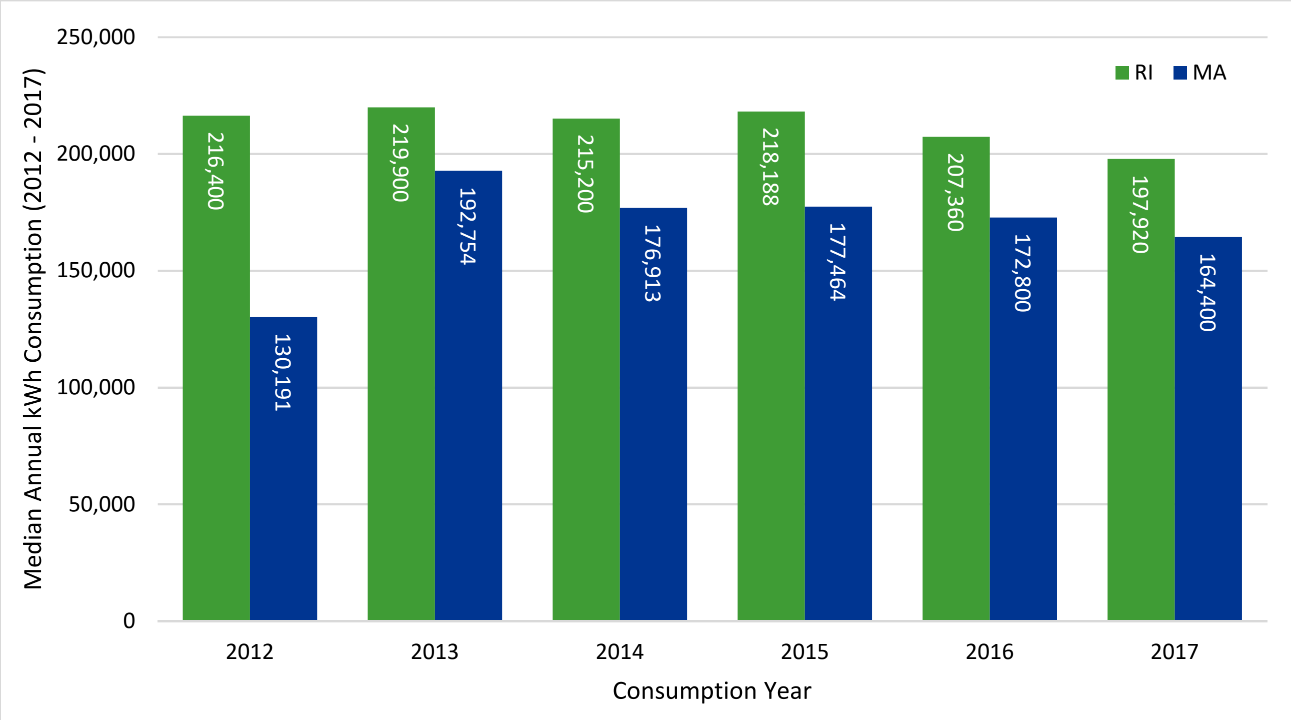


Figure 4-8 shows how the 2014 through 2017 participants are distributed according to NAICS codes. The top thirteen most common codes are shown; the remaining codes are summed into "Other". Across the 2014 to 2017 period, each individual code within "Other" applies to less than 6% of the accounts.

A chi-squared test indicated that the distributions of participants across the different industry categories are statistically significantly different ($p<.01$). RI participants are less likely than MA participants to be Retail Trade, Manufacturing, or Public Administration. However, in general, these differences are small, especially when compared to the proportion of Unknown NAICS codes. These comparisons are limited by the fact that the most common category is unknown.

The NAICS codes that appear in the top seven categories are consistent across the participation years examined. Unknown, Manufacturing, Retail Trade, Education Services, and Health Care and Public Administration are in the top seven each participation year 2014 through 2017.

Based on the distribution of savings, the most important industry sectors for prescriptive lighting in RI are Retail Trade and Educational Services. The BLS trends for those industries (Section 8.2.1) show that the former has followed generally the same direction in both states over the past 10 years, but MA has greater proportional growth than RI. Likewise, the trends for Educational Services also follow the same general direction in both states, but RI has much greater proportional growth in this sector than MA.

Figure 4-8. 2014-2017 Participating Accounts NAICS Codes for Prescriptive Lighting

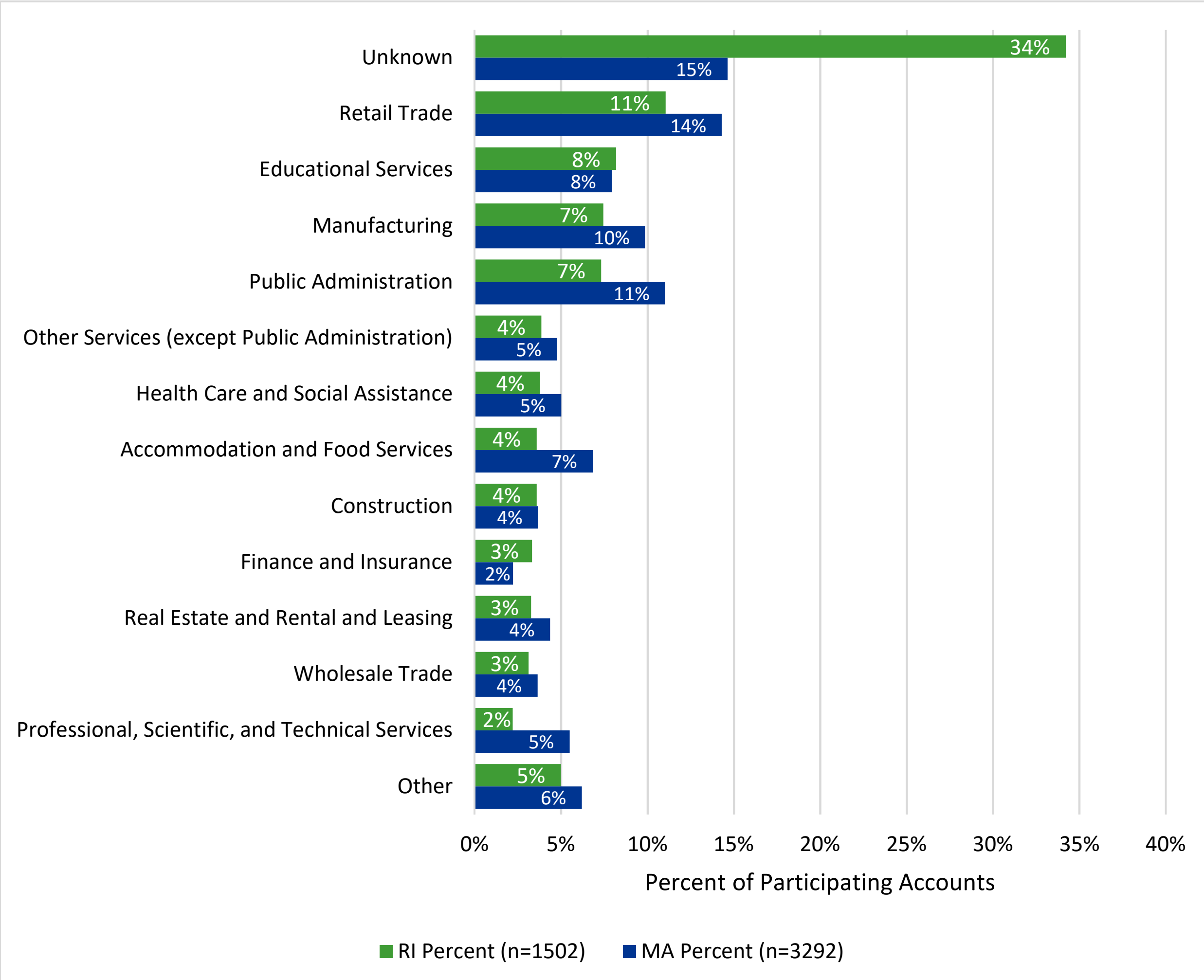
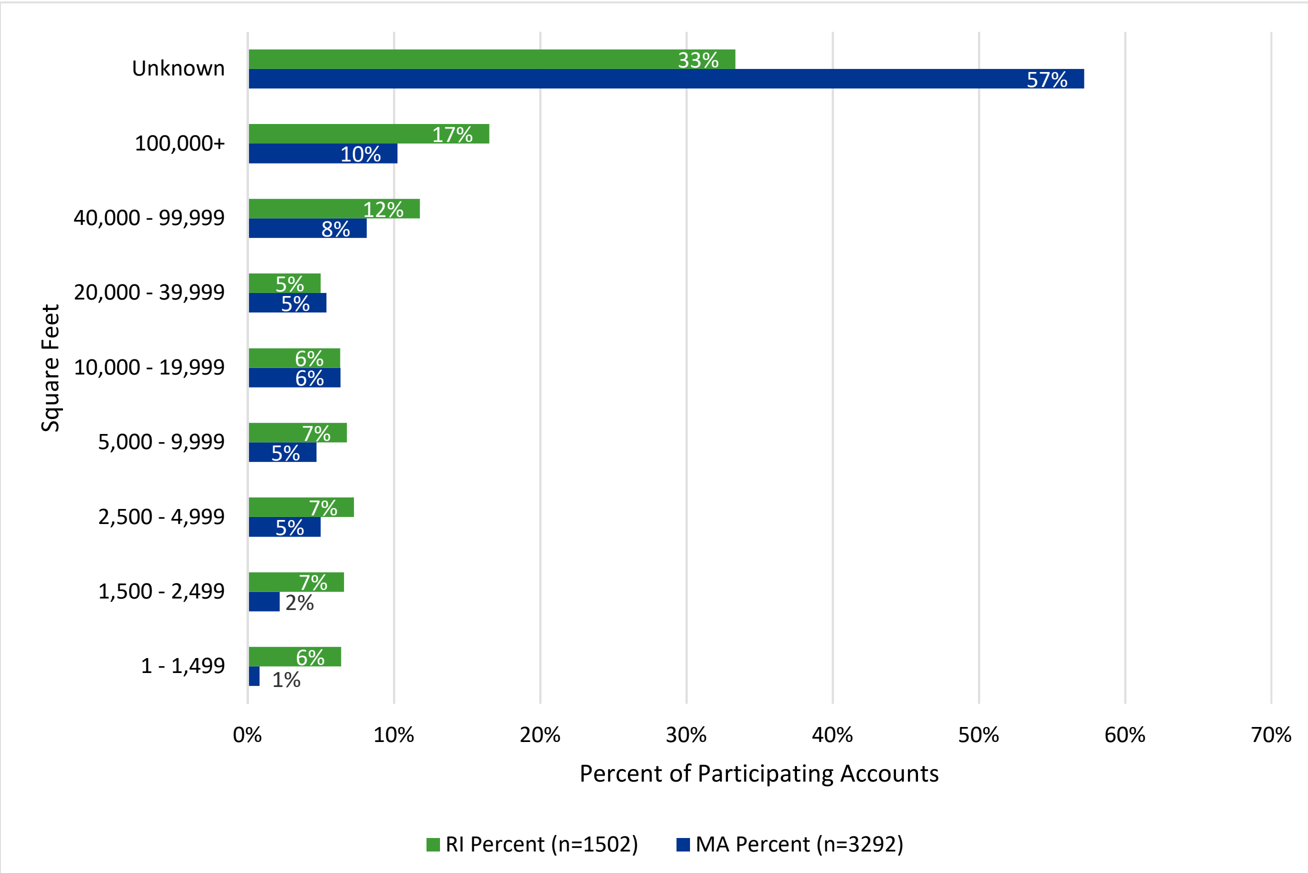


Figure 4-9 shows the distribution of 2014 through 2017 participants by building size categories. A chi-squared test indicated that the distribution by building size is significantly different (p<.01) between RI and MA. The chi-squared test remains statistically significant (p<.01) even if the unknown category is removed.[8]

[8] Future evaluators are likely to have the same level of information available here, including the high rate of unknown NAICS codes. If they factor industry sector into their evaluation plans, they will have to consider the unknown category as one of the categories. Thus, these distributions are best considered with the unknown category remaining.

Figure 4-9. Percent of 2014-2017 Participants by Building Size for Prescriptive Lighting



Previous Evaluation Comparisons

One previous evaluation applies to these participants.

1. Impact Evaluation of 2011 RI Prescriptive Retrofit Lighting Installations (RI).

The primary data collection method was site visits with HOU metering. This study used an independent RI sample. Because DNV GL conducted this and a similar MA study, we had access to raw MA data for a sister study and used it to test interstate differences in major evaluation metrics (Table 4-3). Differences in realization rates and hours of use for lighting systems were statistically significant. Differences in realization rate for controls were not significant, although they were a similar magnitude as the systems differences.

Table 4-3. Summary of Previous Evaluation Comparisons for Prescriptive Lighting

Evaluation	Metric	RI	MA	Statistically Different?
2011 Prescriptive Retrofit Lighting Installations	Population (N)	241	1330	N/A
	Systems sample (n)	18	27	N/A
	Systems Realization rate: kWh savings	89%	103%	**
	Systems Average per project MWh savings	71	175	N/A
	Controls sample (n)	10	20	N/A
	Controls Realization rate: kWh savings	68%	82%	n.s
	Controls Average per project MWh savings1[9]	33	41	N/A
	Verified Average Hours of Use (Systems)	3244	4676	**
	Verified Average Hours of Use (Controls)	1180	1551	n.s.

n.s. not significantly different
**: difference statistically significant at 90% confidence level
1 Average savings *per* controls project. All controls projects also had systems, but not vice versa.

4.3.2 Upstream Lighting

Recommended Evaluation Approach

DNV GL recommends that future evaluations use Approach 4—Independent Sample to obtain statistically robust results for an independent RI-specific sample. This recommendation is based on:

- The programs and measures offered are similar, so Approach 5 is not necessary.
- Tracked gross savings estimates differ, so Approaches 1 and 3 are not recommended.
- In the previous (2015) evaluation, many metrics had statistically significant differences between RI and MA. Metrics where the differences were not statistically significant still differed by substantial amounts, and the lack of statistical significance is most likely due to small sample sizes. These differences apply to underlying parameters such as HOU, which would limit the evaluation cost savings from Approach 2. This difference would also lead away from Approaches 1 and 3.
- Lighting is a rapidly changing market and the second largest C&I program in terms of savings. This suggests that more conservative/rigorous methods are justified, which would lead to Approach 4 over Approach 2.

Program Comparisons

According to program staff, baseline wattage assumptions are consistent across RI and MA. One exception is C&I new construction A-lines, which differ because RI code has lagged MA updates. Differences in planning cycles, evaluation results, and the application of evaluation results has led to differences in the calculated

9

tracked gross savings of upstream LEDs, despite very similar baseline wattages. The RI[10] and MA[11] TRMs did not clearly indicate the annual kWh savings for interior C&I upstream LED lighting. DNV GL consulted with National Grid staff who recommended the following savings baselines for upstream C&I LED bulbs (Table 4-4).

Table 4-4. Upstream LED Annual kWh Savings: C&I

Bulb type	RI	MA
A-line (75/100w)	47.11	30.50
A-line (40/60w)	33.53	21.70

When realization rates are calculated as evaluated savings divided by tracked gross savings, differences in tracked gross savings need to be accounted for in the piggybacking approach. Consider an evaluation that finds the exact same evaluated savings in MA and RI of 30 kWh per lamp. The realization rate for a C&I 75W A-line in MA will be 30/30.5 or 98%. The realization rate for that measure in RI will be 30/47.11 = 64%. In other words, because the MA tracked gross savings are lower, the realization rates for the exact same evaluated savings will be biased upwards relative to RI. The implications for piggybacking are:

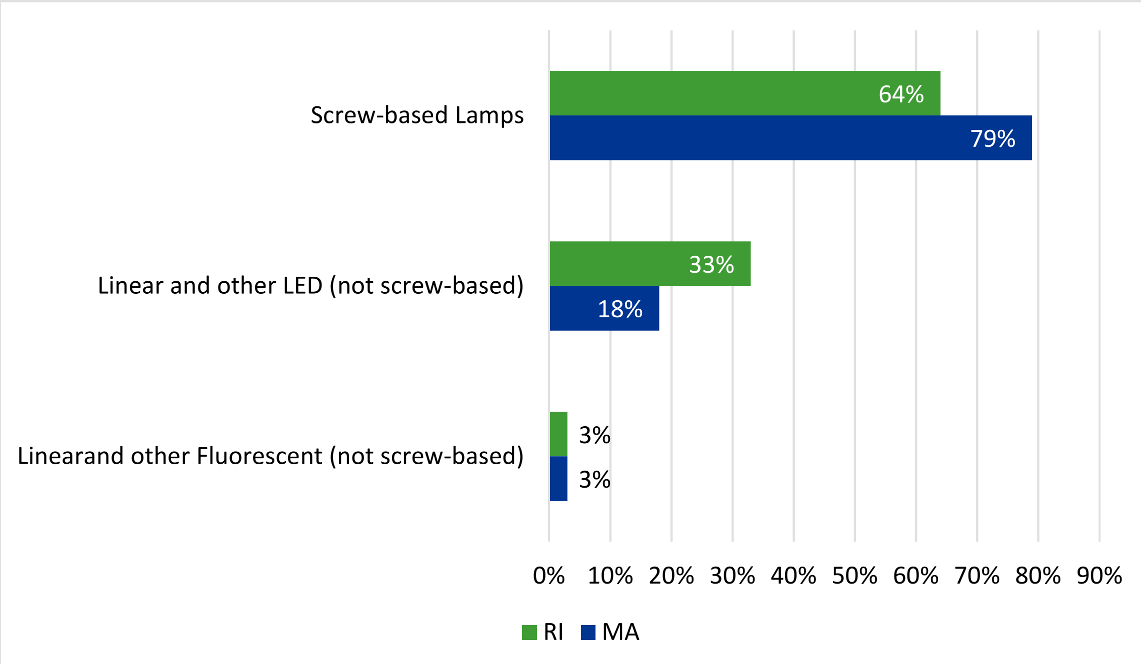
- Direct proxy (Approach 1) is not recommended because the MA results can be expected to have bias.
- Approach 2 could be used if evaluators were careful to parse out and account for the differences in the underlying variables that go into the tracked gross annual kWh calculation.
- Approach 3 should not be used unless evaluators also parse out those underlying differences, and use them to calculate *new* RI-centric realization rates for the MA sites before combining them with the RI evaluation results. This would still allow the RI evaluations to save on field data collection costs, but it is not the way Approach 3 has generally been executed in the past. It is more of a blend of Approach 3 and Approach 2.
- Approach 4 and Approach 5 could be used without modification because the RI realization rate would be based only on RI evaluated savings and RI tracked savings.

Figure 4-10 shows how the proportion of upstream lighting (reported gross) savings are distributed across specific measure types in each state from 2014-2017. Both states see the majority of their consumption savings fall under the screw-based LED lamps measure category, although a lesser proportion of RI savings is in this category. In contrast, RI achieves a greater proportion of savings from Linear and Other LEDs. A chi-squared test indicated a statistically significant difference across the measure type distributions between the two states ($p<.01$).

[10] National Grid Rhode Island Technical Reference Manual 2019 Program Year (November 2018). This version lists 6 annual kWh for all C&I prescriptive internal LED lamps as well as 6 kW. It does not seem like both values can be accurate. Follow-up conversations with National Grid staff produced the numbers shown in the table.

[11] http://ma-eeac.org/wordpress/wp-content/uploads/2016-2018-Plan-1.pdf

Figure 4-10. Proportion of Reported Gross Savings by Measure for Upstream Lighting



National Grid did not track the individual accounts that participated in the upstream lighting program in the program years analyzed for this study. Thus, individual, participant level comparisons were not available for this measure group.

Previous Evaluation Comparisons

One previous evaluation applies to this measure group:

1. Impact Evaluation of PY2015 RI Commercial & Industrial Upstream Lighting Initiative (MA and RI)

The primary data collection method for this study was site visits. This evaluation originally utilized a pooled sample of both RI and MA sites (Approach 3).[12] DNV GL compared the RI and MA results to provide an analytic analysis. Table 4-5 shows evaluation metrics split by RI and MA. Statistical difference testing was based on the confidence level used in the original report for that metric.

Overall realization rates for kWh savings differed by approximately 40%, although the difference did not reach statistical significance. Differences in realization rates for specific technologies ranged from 15% to 75%. Most of these differences were statistically significant. Differences in HOU for all types of specific technology groups were statistically significant. It should be noted that the small sample sizes reduce statistical power particularly for testing involving sub-samples. This results in some large differences in results failing to achieve statistical significance. These are key findings for our recommendation.

[12] This study utilized data from another evaluation done previously: The Impact Evaluation of PY2015 Massachusetts Commercial & Industrial Upstream Lighting initiative, which used sites from all primary administrators (PAs). The MA sites used in this evaluation is a subset of that data from National Grid only. The RI sites were collected separately and the sites of the two states were pooled for analysis.

Table 4-5. Summary of Previous Evaluation Comparisons for Upstream Lighting

Evaluation	Metric	RI	MA (National Grid only)	Statistically Different?
Impact Evaluation of PY2015 RI Commercial & Industrial Upstream Lighting Initiative (account level)	Population (N)	3547	8131	N/A
	Sample(n)	29	73	N/A
	Realization rate: kWh savings (overall)	84%	47%	n.s.
	Annual MWh Realization Rate (TLEDs)	163%	198%	n.s.
	Annual MWh Realization Rate (Stairwell Kits)	83%	8%	**
	Annual MWh Realization Rate (Retrofit Kits)	61%	48%	n.s.
	Annual MWh Realization Rate (A-forms and Decoratives)	87%	34%	*
	Annual MWh Realization Rate (G24s)	152%	120%	n.s.
	In-service rate RR (TLEDs)	70%	92%	n.s.
	In-service rate RR (Stairwell kits)	84%	58%	n.s.
	In-service rate RR (Retrofit kits)	55%	69%	n.s.
	In-service rate RR (A-lines and Decoratives)	67%	65%	n.s.
	In-service rate RR (G24s)	65%	69%	n.s.
	Hours of Use RR (TLEDs)	102%	125%	*
	Hours of Use RR (Stairwell Kits)	97%	26%	**
	Hours of Use RR (Retrofit Kits)	128%	77%	**
	Hours of Use RR (A-lines and Decoratives)	96%	66%	**
	Hours of Use RR (G24s)	155%	132%	**

n.s. not statistically significant
* different at 80% confidence level
** different at 90% confidence level

4.3.3 Custom Electric Non-lighting

Recommended Evaluation Approach

DNV GL recommends that future evaluations use Approach 4—Independent Sample to obtain statistically robust results for an independent RI-specific sample. This recommendation is based on:

- Programs are similar so Approach 5 is not necessary.
- As a custom program, Approach 2 is not applicable.
- Previous evaluation results differ, so we would not recommend Approaches 1 or 3.
- National Grid uses similar engineering firms and methods in both states; this would make Approach 3 a possibility if previous evaluation results were similar.

Even though there is a high amount of overlap in the engineering firms used in RI and MA, this program makes up a large percent of annual savings. In addition, measure mixes differ, customer characteristics differ, and past evaluation results differed. Therefore, we suggest that evaluations move towards an independent RI sample that can leverage site data collection tools (Approach 4) from MA. It is our understanding that Approach 4 is already being used in the next evaluation. While we do not recommend using Approach 3, if evaluators choose to do so in the future, then we recommend taking steps to correct for differences in measure mix and customer types when selecting which MA sample points to include.

Program Comparisons

Figure 4-11 shows how the proportion of custom electric (reported gross) savings are distributed across the two states. RI is achieving a greater share of custom electric non-lighting program savings from compressed air, refrigeration, and other, and a lesser share from HVAC and process than MA.

Figure 4-11. Proportion of Reported Gross Savings by Measure (custom electric non-lighting, 2013-2017)

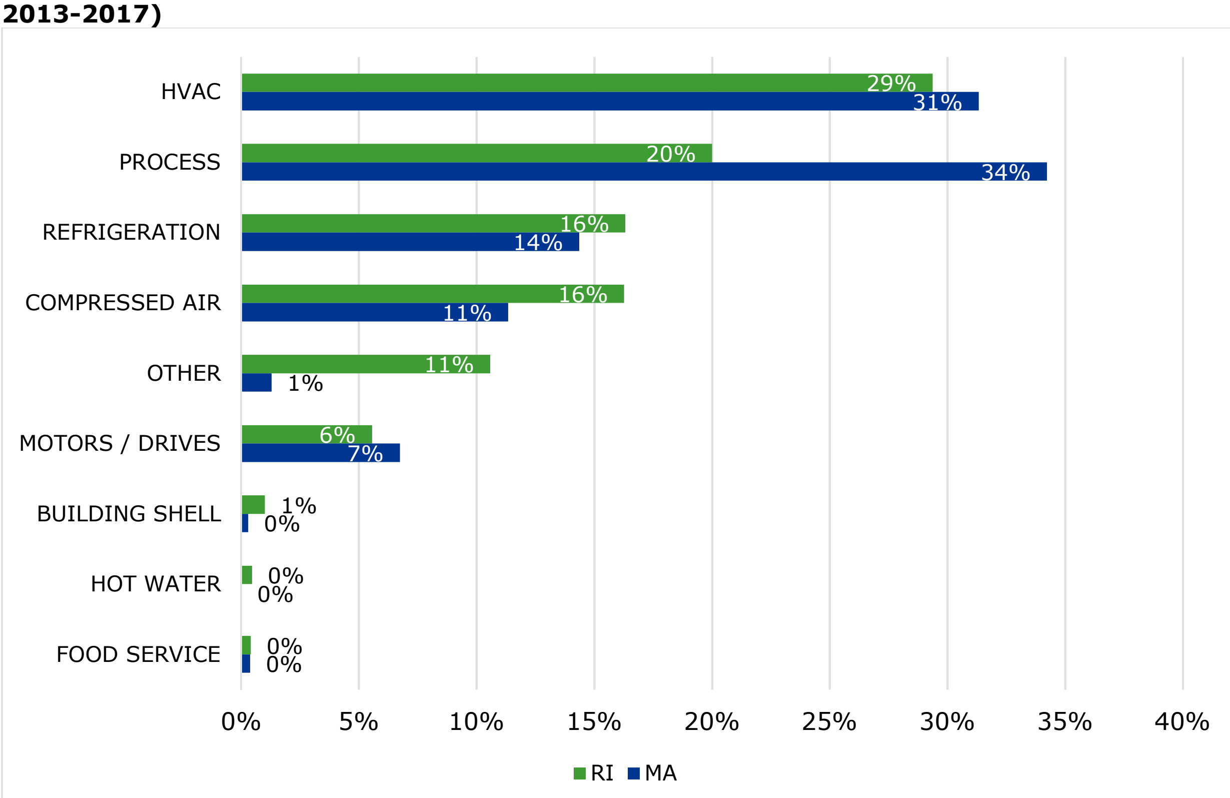


Figure 4-12 shows that the median annual consumption (calculated over 2012-2017) of RI custom non-lighting participants was less than MA participants.

Figure 4-12. Participant Median Annual Consumption (custom electric non-lighting, 2012-2017)

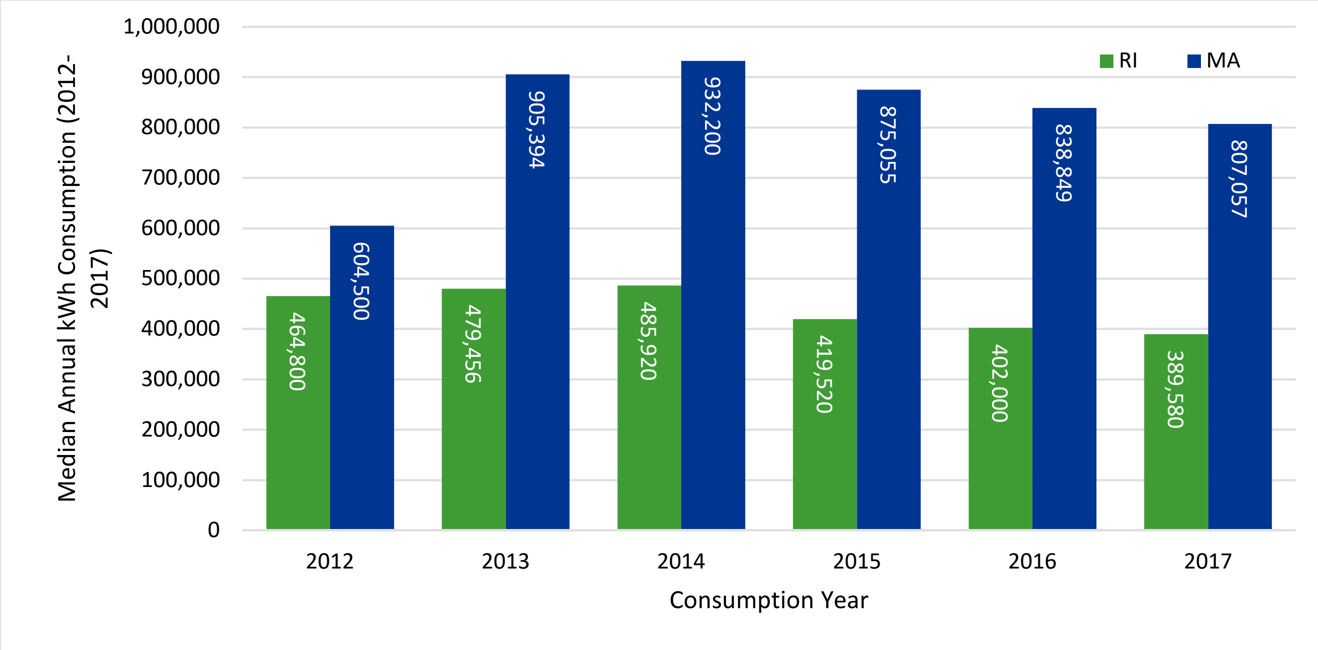


Figure 4-13 shows how the 2014 through 2017 participants are distributed according to NAICS codes. The top seven most common codes are shown; the remaining codes are summed into "Other". A chi-square test indicates that the difference in distribution of participating accounts by NAICS code in MA and RI were statistically significant from each other ($p < 0.1$). This analysis is somewhat limited by the high proportion of unknown NAICS codes. However, these distributions remain statistically different when the unknown category is removed.

Of the four most important sectors, Manufacturing shows the greatest difference in growth trends between the two states (Section 8.2.1). The slopes for Education Services and Retail Trade are similar for both states, but the magnitude of growth is significantly different for each. Accommodation and Food Services has similar growth trends across both states.

Figure 4-13. 2014-2017 Participating Accounts by NAICS Codes for Custom Electric Non-lighting

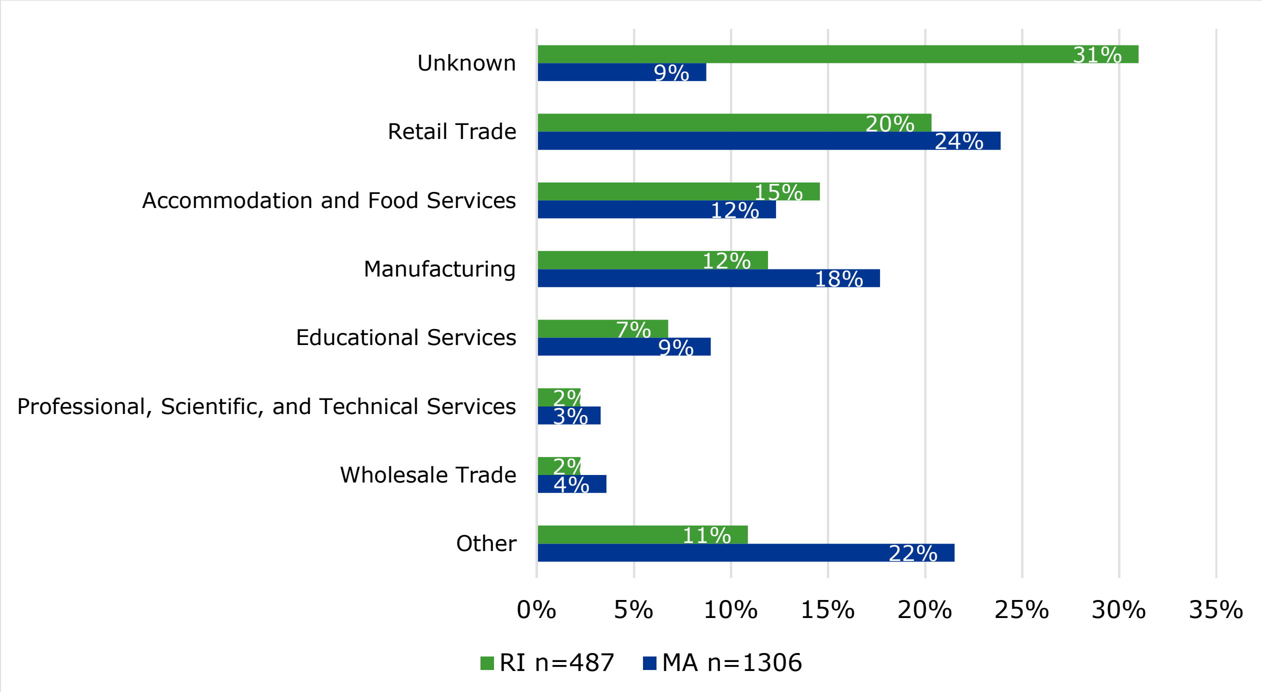
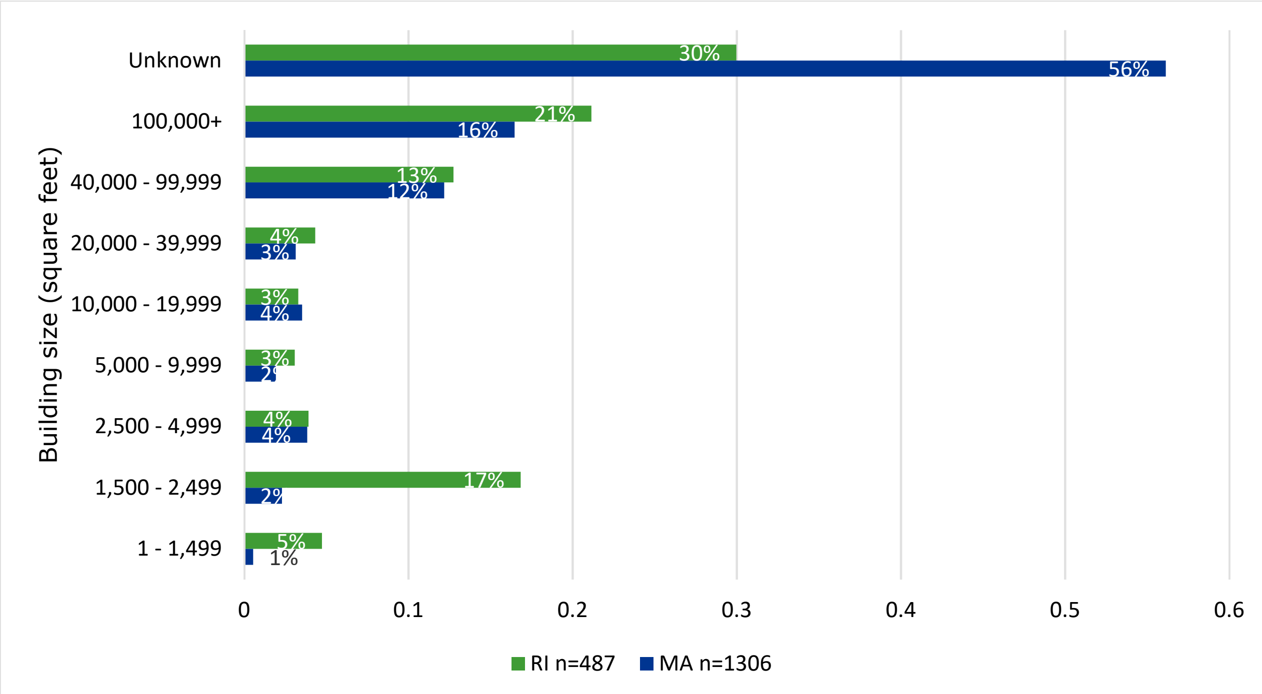


Figure 4-14 shows how the 2014 through 2017 participants break down according to building size categories. There are some differences in the customer types reached by each program. The most substantial categorical difference is the proportion of unknowns in MA. A chi-square test indicated the difference in distribution of RI and MA accounts by building size was statistically significant ($p < .01$). This comparison is limited by the fact that the most common category is unknown.

Figure 4-14. Percent of 2014-2017 Participants by Building Size for Custom Electric Non-lighting



Previous Evaluation Comparisons

Four previous evaluations apply to these participants:

1. Impact Evaluation of 2014 Custom HVAC Installations (MA and RI).
2. 2014 RI Custom Process Impact Evaluation (MA and RI).
3. Impact Evaluation of National Grid Rhode Island's Custom Refrigeration, Motor and Other Installations (MA and RI; 2014).
4. RI Commercial and Industrial Impact Evaluation of 2013-2015 Custom CDA Installations (MA and RI).

These evaluations originally utilized a pooled sample approach (Approach 3). DNV GL separated and compared the RI and MA results for each study such that each result represents the findings from that state only. We then re-pooled the state-specific results for both studies to provide a meta-analytic analysis. The choice of confidence levels was based on the confidence levels reported in the original studies. Table 4-6 shows where RI and MA participants had statistically significantly different results in evaluations 1 to 3. We report the Comprehensive Design differences in a separate table because they are not included in the pooled results in Table 4-6.

Realization rates for kWh savings varied significantly between the states in both studies and the pooled sample. Additionally, differences in average project size, both overall and within specific strata are apparent. In particular, MA projects tend to be one category (stratum) larger than the RI projects. Removing the projects in the largest MA stratum does not change the results of the statistical difference tests. These are key findings for our recommendation.

Table 4-6. Summary of Previous Evaluation Comparisons for Custom Electric Non-lighting

Evaluation	Metric	RI	MA	Statistically Different?
Impact Evaluation of 2014 Custom HVAC Installations	Population (N)	31	57	N/A
	Sample(n)	6	23	N/A
	Realization rate: kWh savings	91%	75%	**
	Realization rate: Summer on-peak kW	67%	70%	n.s.
	Realization rate: Winter on-peak kW	98%	67%	*
	Realization rate: % On-peak	84%	105%	**
	Average project MWh savings (overall)	98	305	**
	Average project MWh savings (stratum 1)	28	71	**
	Average project MWh savings (stratum 2)	117	276	**
	Average project MWh savings (stratum 3)	272	560	**
	Average project MWh savings (stratum 4)	694	1,599	**
2014 RI Custom Process Impact Evaluation	Population (N)	11	58	N/A
	Sample(n)	4	20	N/A
	Realization rate: kWh savings	111%	68%	**
	Realization rate: Summer on-peak kW	80%	65%	n.s.
	Realization rate: Winter on-peak kW	46%	75%	*
	Realization rate: % On-peak	105%	92%	n.s.
	Average project MWh savings (overall)	187	183	n.s.
	Average project MWh savings (stratum 1)	85	92	n.s.
	Average project MWh savings (stratum 2)	459	350	**
	Average project MWh savings (stratum 3)	-	782	N/A
Impact Evaluation of National Grid Rhode Island's Custom Refrigeration, Motor and Other Installations	Population (N)	21	169	N/A
	Sample (n)	6	24	N/A
	Overall realization rate: kWh savings	100%	82%	**
	Realization rate: Summer on-peak kW	114%	88%	**
	Realization rate: Winter on-peak kW	117%	86%	**
	Realization rate: % On-peak	139%	109%	**
	Average project MWh savings (overall)	145	103	N/A
	Average project MWh savings (stratum 1)	84	27	**
	Average project MWh savings (stratum 2)	446	134	**
	Average project MWh savings (stratum 3)	-	703	N/A
Pooled	Population (N)	80	276	N/A
	Sample(n)	16	69	N/A
	Realization rate: kWh savings	98%	63%	**
	Realization rate: Summer on-peak kW[1]	81%	74%	n.s.
	Realization rate: Winter on-peak kW[1]	89%	69%	*
	Realization rate: % On-peak[1]	51%	50%	n.s.
	Average project MWh savings (overall)	245	448	**

n.s. not significantly different
* different at 80% confidence level
** different at 90% confidence level
1 sample size for metric: RI n=18, MA n=64

Table 4-7. Summary of Previous Evaluation Comparisons for Custom Electric Non-lighting

Evaluation	Metric	RI	MA	Statistically Different?
RI Commercial and Industrial Impact Evaluation of 2013-2015 Custom CDA Installations	Population (N)	5	19	N/A
	Sample (n)	2	4	N/A
	Overall realization rate: kWh savings	67%	45%	**
	Realization rate: Summer on-peak kW	62%	46%	n.s.
	Realization rate: Winter on-peak kW	71%	22%	n.s.
	Realization rate: % On-peak	71%	91%	n.s.
	Average project MWh savings (overall)	156	531	N/A

n.s. not significantly different
** different at 90% confidence level

4.3.4 Custom Electric Lighting

Recommended Evaluation Approach

As for custom non-lighting, we suggest using Approach 4 (independent samples) in future evaluations of this program. This recommendation is based on:

- Programs are similar so Approach 5 is not necessary.
- As a custom program, Approach 2 is not applicable.
- Previous evaluation results differ, so we would not recommend Approaches 1 or 3.

Despite similar measure mixes, because past evaluation results differed, this measure group has a relatively large amount of savings, and the fact that lighting is a rapidly evolving market we recommend Approach 4 (independent samples). We understand the current evaluation of this program is already implementing Approach 4.

Program Comparisons

Participation data for custom electric lighting by measure types more specific than “Lighting” was not available.

Figure 4-15 shows that the median consumption for RI custom lighting participants was less than MA participants in all participation years.

Figure 4-15. Median Annual Participant Consumption (custom electric lighting, 2012-2017)

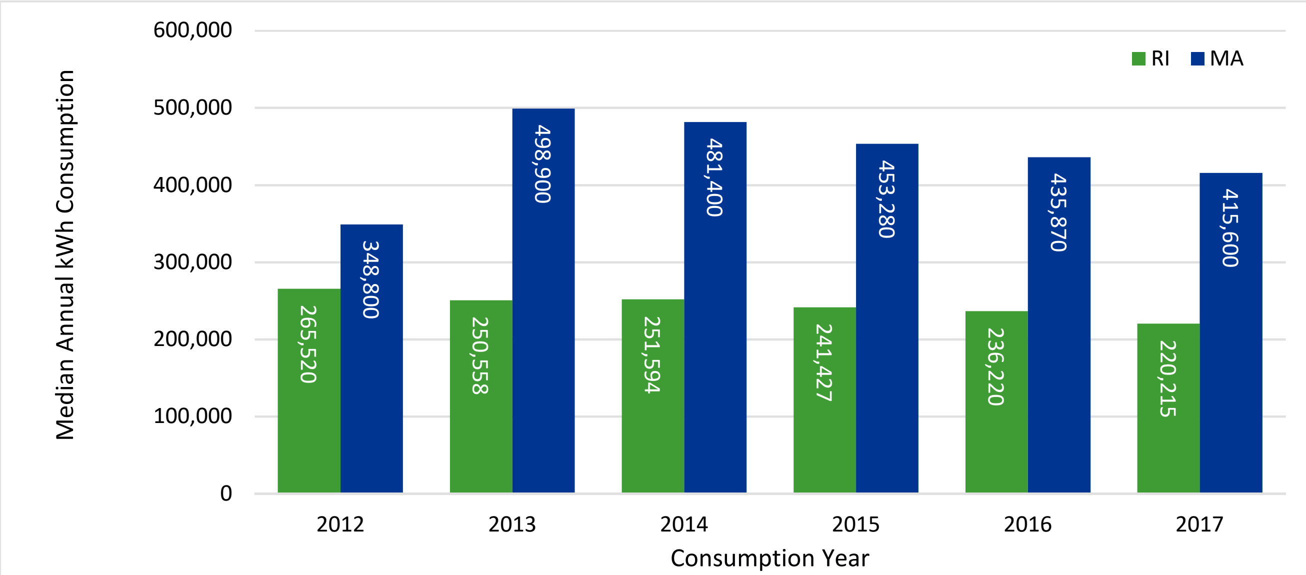


Figure 4-16 shows how the 2014 through 2017 participants are distributed by NAICS code. The top seven most common codes are shown; the remaining codes are summed into "Other".

A chi-squared test indicated that the distributions of participants across the different industry categories are statistically significantly different ($p<.01$). RI participants are more likely than MA participants to come from the Accommodation and Food Services sector and less likely to come from Retail Trade or Manufacturing. However, these comparisons are limited by the fact that the most common category is "Unknown".

Based on the distribution of savings, the industry sectors with the most custom electric lighting savings in RI are Accommodation and Food Services and Educational Services. The BLS trends for those industries show that the former has followed generally the same trend in both states over the past 10 years (Section 4.2). The trends for Educational Services also follow the same general direction in both states, but MA has much greater growth in this sector than RI.

Figure 4-16. 2014-2017 Participating Accounts NAICS Codes for Custom Electric Lighting

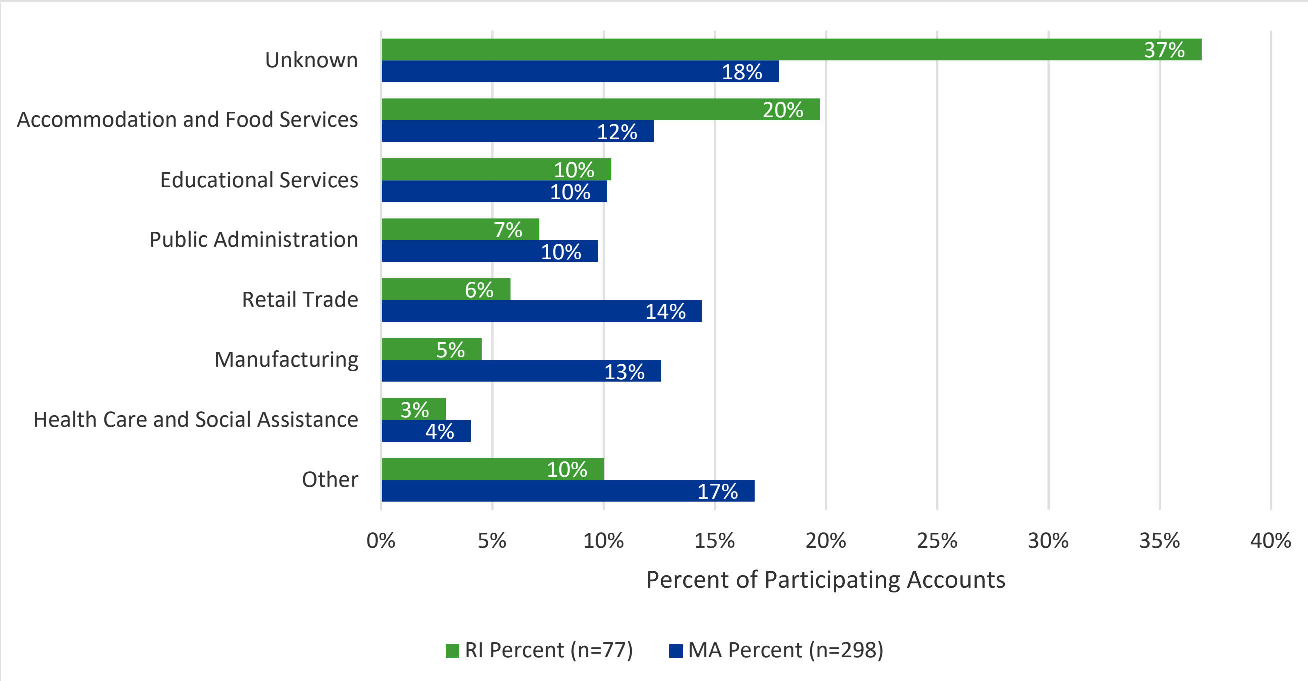
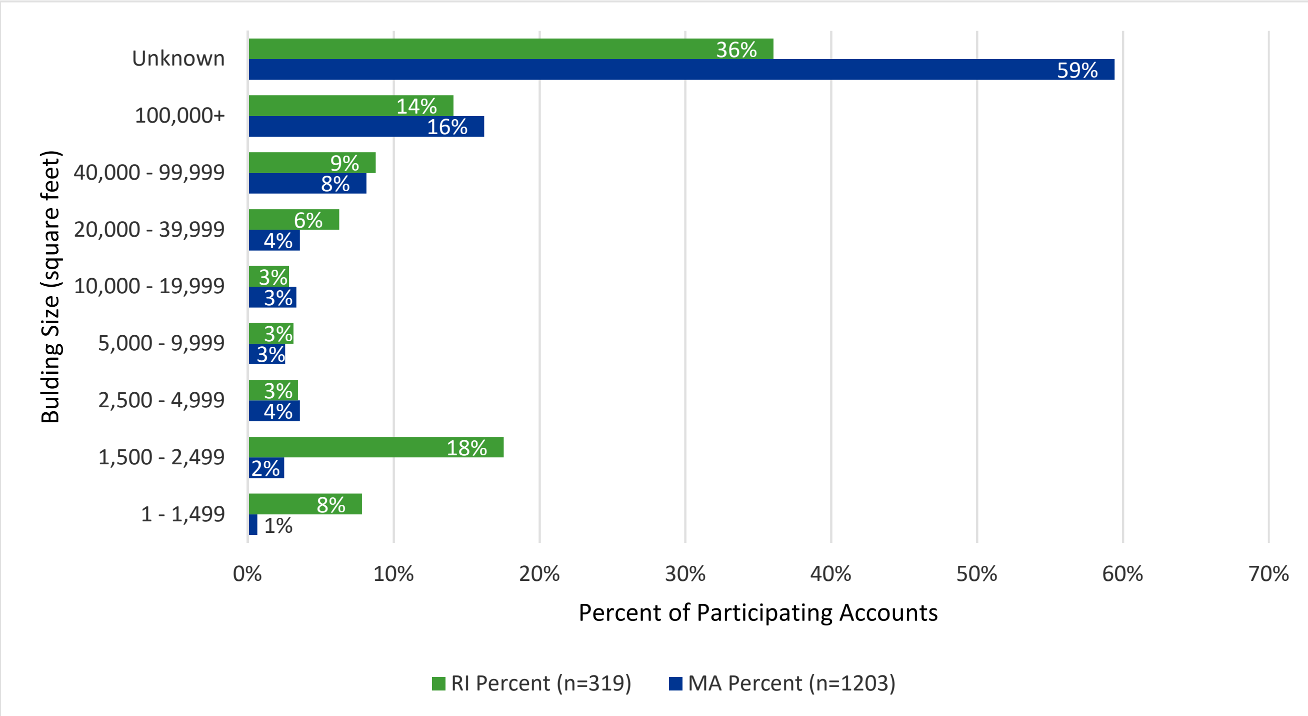


Figure 4-17 shows the distribution of 2014 through 2017 participants by building size categories. As for the industry-sector distribution, a chi-squared test indicated that the distribution by building size is significantly different ($p<.01$) for RI and MA. RI participants are more likely than MA participants to be in the smallest two size categories, as well as in the 20,000 – 39,999 square foot size category. This comparison is limited by the fact that the most common category is "Unknown".

Figure 4-17. Percent of 2014-2017 Participants by Building Size for Custom Electric Lighting



Previous Evaluation Comparisons

One previous evaluation applied to this measure type:

1. Impact Evaluation of 2011 RI Custom Lighting Installations (MA and RI).

The data collection method used in this study was site visits with metering. This evaluation utilized a pooled sample (Approach 3). DNV GL separated and compared the RI and MA results for each study such that each result represents the findings from that state only. The choice of confidence levels was based on the confidence levels reported in the original studies. Table 4-8 shows where RI and MA participants had statistically significantly different results in this evaluation. Realization rates for kWh savings and winter on-peak kW varied significantly between the states. Differences between Summer on-peak kW were not significant.

Table 4-8. Summary of Previous Evaluation Comparisons for Custom Electric Lighting

Metric	RI	MA	Statistically Different?
Population (N)	45	84	N/A
Sample (n)	4	14	N/A
Realization rate: kWh savings	80%	98%	**
Realization rate: Summer on-peak kW	75%	116%	n.s.
Realization rate: Winter on-peak kW	64%	85%	*

n.s. not significantly different
* different at 80% confidence level
** different at 90% confidence level

4.3.5 Small Business Electric

Recommended Evaluation Approach

DNV GL recommends that future evaluations can use pooled samples (Approach 3), but with steps taken to adjust MA results to be more representative of RI customer characteristics. Approach 2 could also be justified due to a lack of any information that would definitely eliminate it and the fact this is a relatively small program. Lighting savings constitute approximately 90% of the program savings, so if those are removed, the remaining savings would be approximately 1% of statewide C&I electric savings in which case Approach 1 (direct proxy) could be justified. These recommendations are based on:

- Programs are similar so Approach 5 is not necessary.
- Most of the previous evaluation results did not differ between states, so Approach 3 is possible.
- The distribution of customers by industry segment differs, which might affect the values of savings parameters such as HOU and ISR, so evaluation cost savings for Approach 2 might be limited. This at least points to the need for adjustments to pooled samples in Approach 3.
- This program accounts for a relatively small amount of savings, especially if lighting savings are removed from the evaluation, in which case Approach 2 or even Approach 1 is justified.

Program Comparisons

Figure 4-18 shows how the proportion of small business electric (reported gross) measure savings are distributed across the two states. Measures representing less than 1% of the mix have been omitted from this graph. For both RI and MA, lighting accounts for about 90% of the overall savings, with refrigeration and HVAC comprising most of the rest. Both states show a similar distribution of savings across these three measures. A Chi-square test did not indicate statistically different distributions of savings.

Figure 4-18. Proportion of Reported Gross Savings by Measure for Small Business Electric

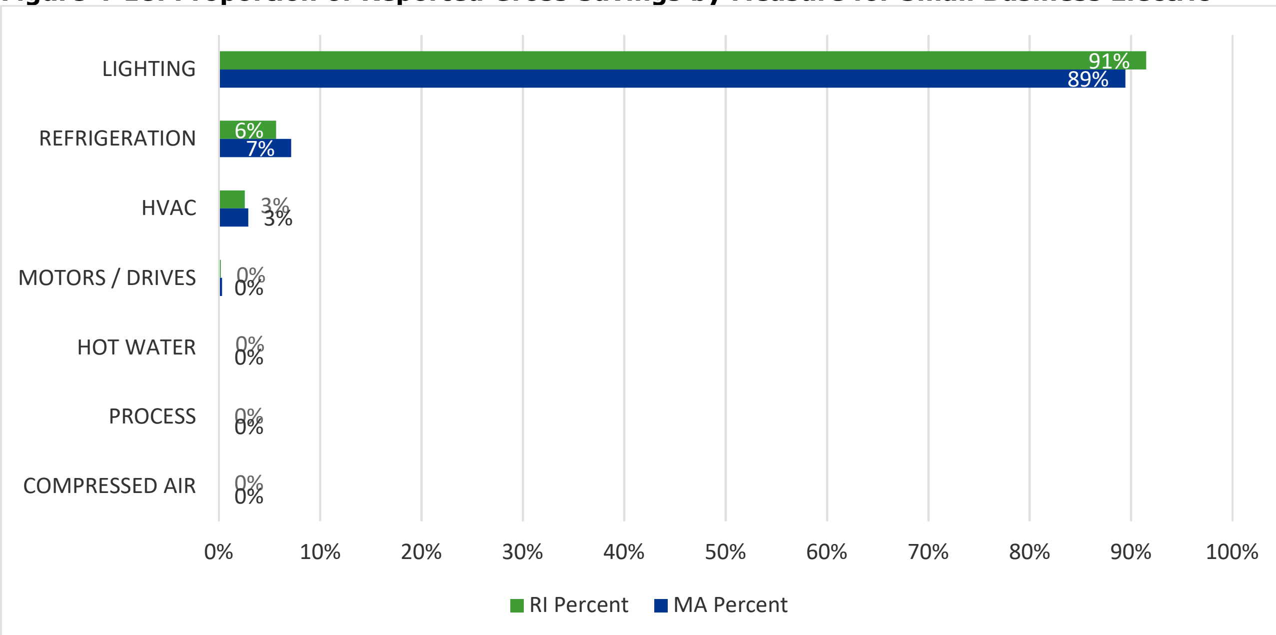


Figure 4-19 shows that the median consumption for RI participants was similar to MA participants in all participation years except 2012. Median consumption in RI was significantly greater than MA in 2012.

Figure 4-19. Median Annual Consumption Over 2012-2017 by Participation Year for Small Business Electric

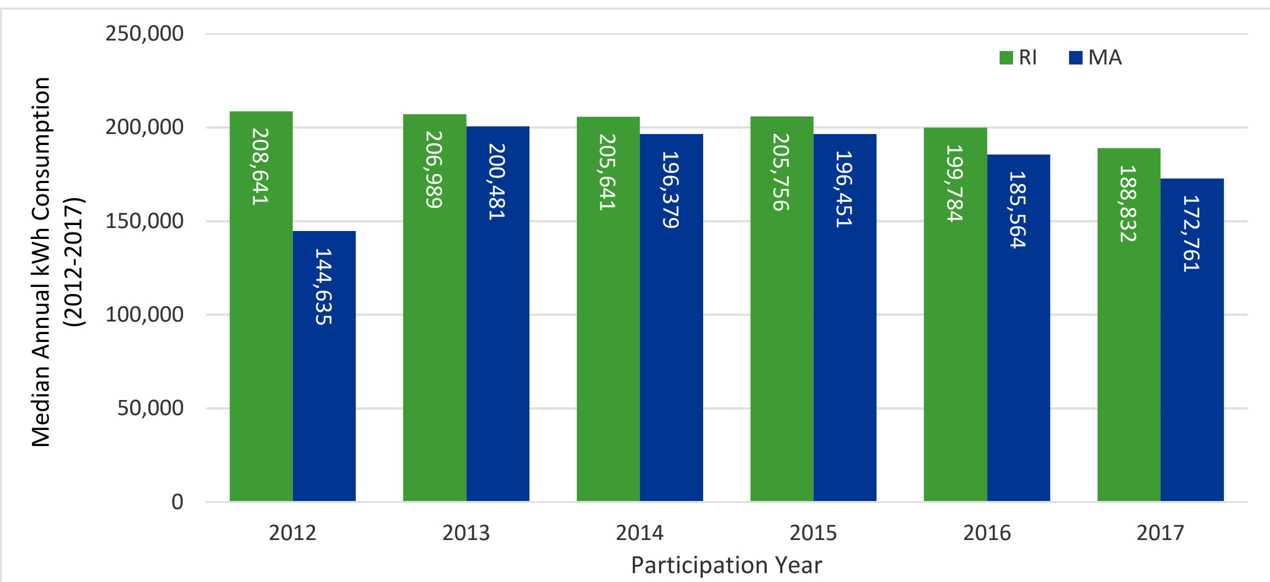


Figure 4-20 shows how the 2014 through 2017 participants are distributed by NAICS code. The top seven most common codes are shown; the remaining codes are summed into "Other". Across the 2014 to 2017 period, each specific measure type within "Other" applies to less than 5% of the accounts.

A chi-squared test indicated that the distributions of participants across the different industry categories are statistically significantly different ($p<.01$). RI participants are less likely than MA participants to come from the Retail Trade, Accommodation and Food Services, and Professional, Scientific, And Technical Services, and slightly less likely to participate in Other Services (except Public Administration) and Health Care and Social Assistance. MA participants are slightly more likely than RI participants to come from Manufacturing. However, these comparisons are limited by the fact that the most common RI category is unknown.

The most important industry sectors for small business electric in RI are Retail Trade and Other Services (except Public Administration). The BLS trends for those industries (Section 8.2.1) show that the former has not followed the same trend in both states over the past 10 years. The trends for Other Services (except Public Administration) follow the same general direction between the states, but MA has much greater proportional growth in this sector than RI.

Figure 4-20. 2014-2017 Participating Accounts NAICS Codes for Small Business Electric

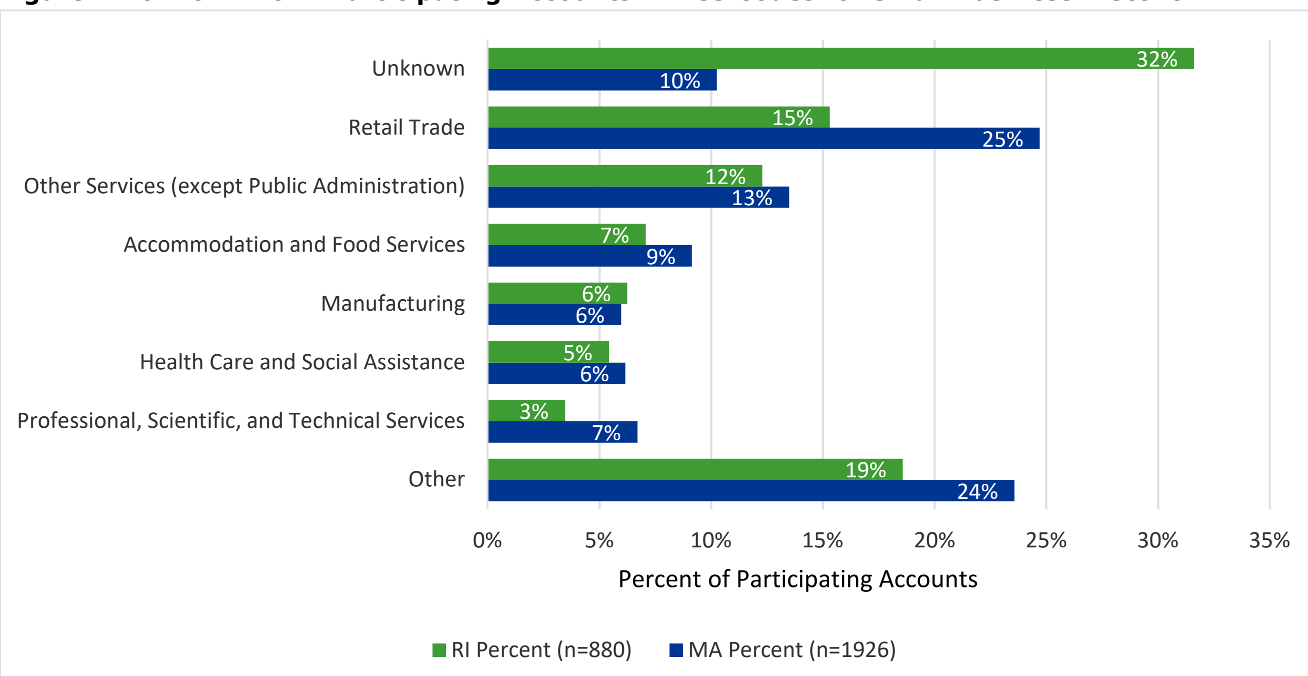
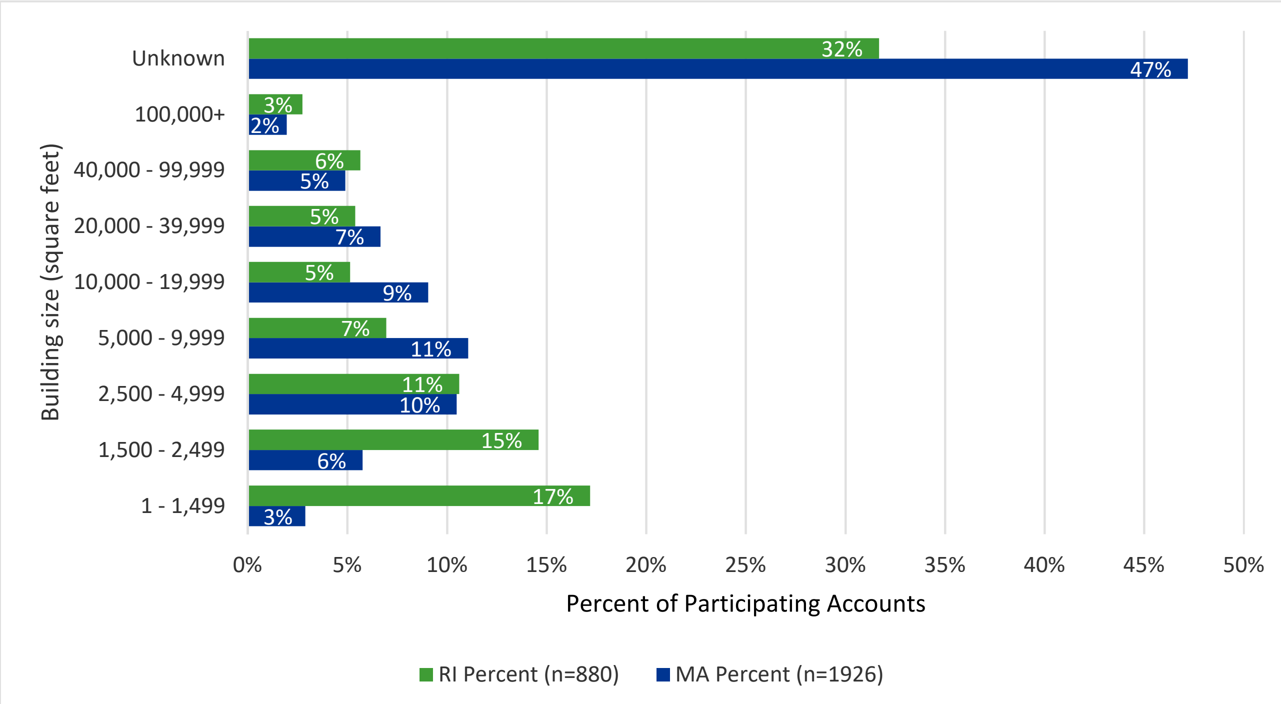


Figure 4-21 shows how the distribution of 2014 through 2017 participants by building size categories. As for the industry-sector distribution, a chi-squared test indicated that the distribution by building size is significantly different ($p<.01$) for RI and MA. RI participants are more likely than MA participants to be in the medium size categories, as well as in the smallest size categories. This comparison is limited by the fact that the most common category is unknown.

Figure 4-21. Percent of 2014-2017 Participants by Building Size for Small Business Electric



Previous Evaluation Comparisons

One previous evaluation applied to this program:

1. Impact Evaluation of PY2016 RI Commercial and Industrial Small Business Initiative (MA and RI).

This study, which covered only lighting projects, used site visits for data collection. This evaluation utilized a pooled sample (Approach 3). DNV GL separated and compared the RI and MA results for each study such that each result represents the findings from that state only. The choice of confidence levels was based on the confidence levels reported in the original studies. Table 4-9 shows where RI and MA participants had statistically significantly different results in this evaluation only for winter peak kW. Realization rates for kWh and Summer peak kW were not significantly different.

Table 4-9. Summary of Previous Evaluation Comparisons for Small Business Electric

Metric	RI	MA	Statistically Different?
Population (N)	787	1506	N/A
Sample (n)	30	55	N/A
Realization rate: kWh savings	107%	104%	n.s.
Realization rate: Summer on-peak kW	83%	94%	n.s.
Realization rate: Winter on-peak kW	126%	93%	*

n.s. not significantly different
* different at 80% confidence level

4.3.6 Prescriptive Electric Non-lighting

DNV GL recommends that future evaluations use independent samples (Approach 4). However, because of the relatively small size of this program, Approaches 2 or 3 could be justified. If the evaluations focus on individual, specific measures as they have tended to do in the past, then the amount of savings for each evaluation would be further reduced. This would increase justification to use Approaches 2 or 3 rather than 4. This recommendation is based on:

- Programs are similar so Approach 5 is not necessary.
- While overall realization rates in the previous evaluations were not significantly different, the magnitude of the difference was large and failed to achieve statistical significance because of small sample sizes. Therefore, we cannot completely eliminate, but would not recommend Approach 1 or 3.
- Distributions of participants in terms of consumption was similar, but distributions by industry type and measure mixes differed. This suggests that the parameters in Approach 2 could vary, limiting the evaluation cost savings of that approach.
- This is the smallest measure group in terms of C&I savings, so less expensive evaluation methods can be justified.

Program Comparisons

Figure 4-22 shows how the proportion of prescriptive non-lighting (reported gross) savings are distributed across the two states. RI is achieving a greater share of program savings from compressed air, hot water, and other measures. RI also sees a lesser share from HVAC, motors/drives, refrigeration, and motors/drives than MA.

Figure 4-22 Proportion of Reported Gross Savings by Measure for Prescriptive Non-lighting

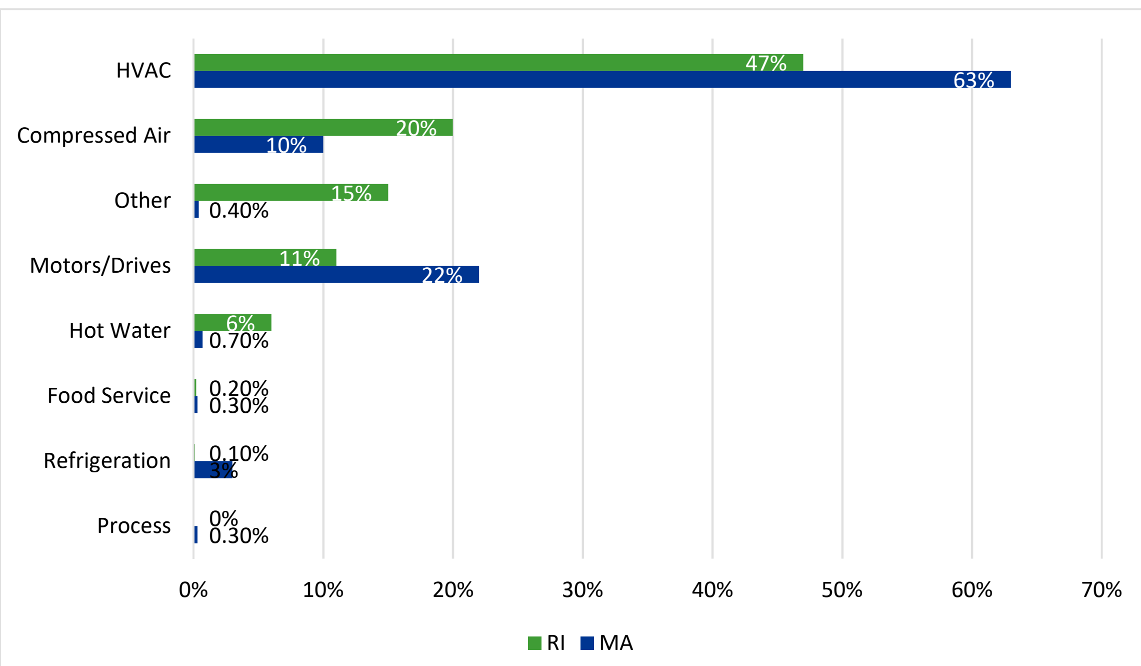


Figure 4-23 shows that the median annual consumption (averaged over 2012 to 2017) of RI participants was near equal to MA participants, except for organizations that participated in 2012. This is a key finding for our recommendation.

Figure 4-23 Median Annual Consumption Over 2012-2017 by Participation Year Prescriptive Non-lighting

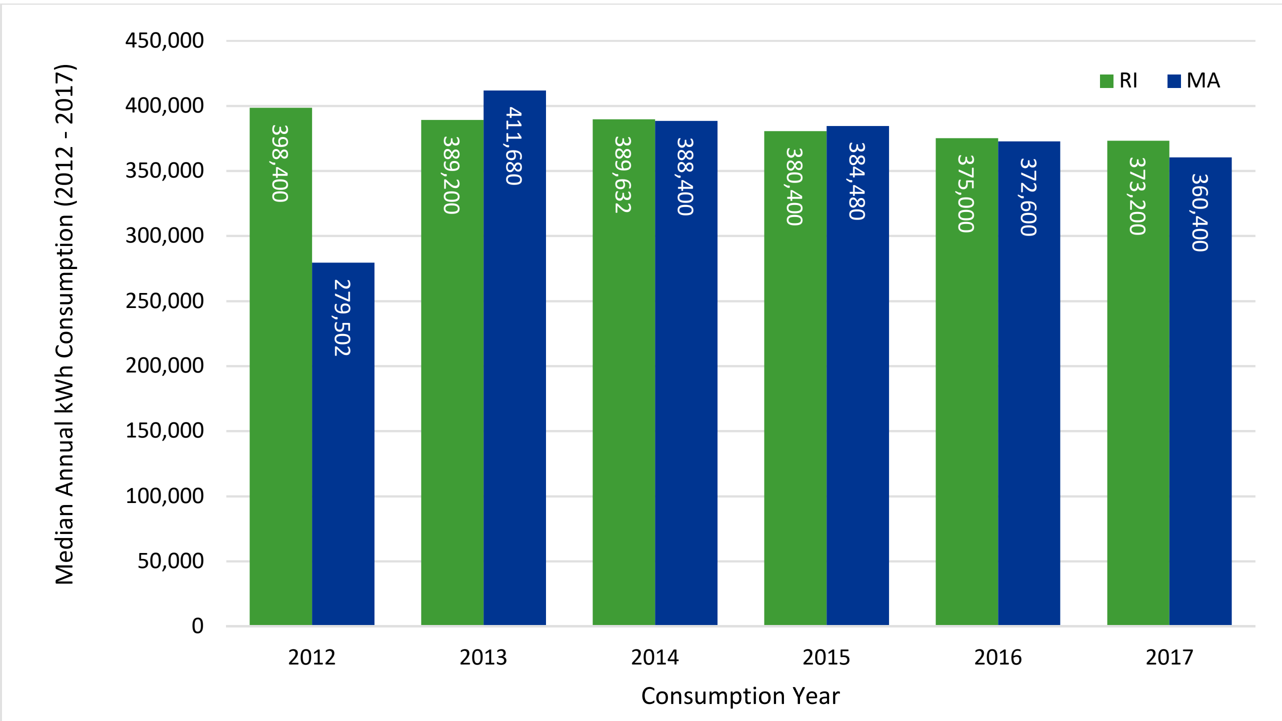


Figure 4-24 shows how the cumulative 2014 through 2017 participants are distributed according to NAICS codes for RI and MA. The top seven most common codes are shown. Across the 2014 to 2017 period, the "Other" category applies to less than 4% of the accounts. For the industry-sector distribution, a chi-squared test indicated that the distribution of participants by NAICS code was not statistically different between RI and MA.

Of the four most important sectors, Manufacturing shows the greatest difference in growth trends between the two states. The slopes for Education Services and Retail Trade are similar for both states, but the magnitude of growth is significantly different for each. Accommodation and Food Services has similar growth trends across both states.

Figure 4-24 2014-2017 Participating Accounts NAICS Codes for Prescriptive Non-lighting

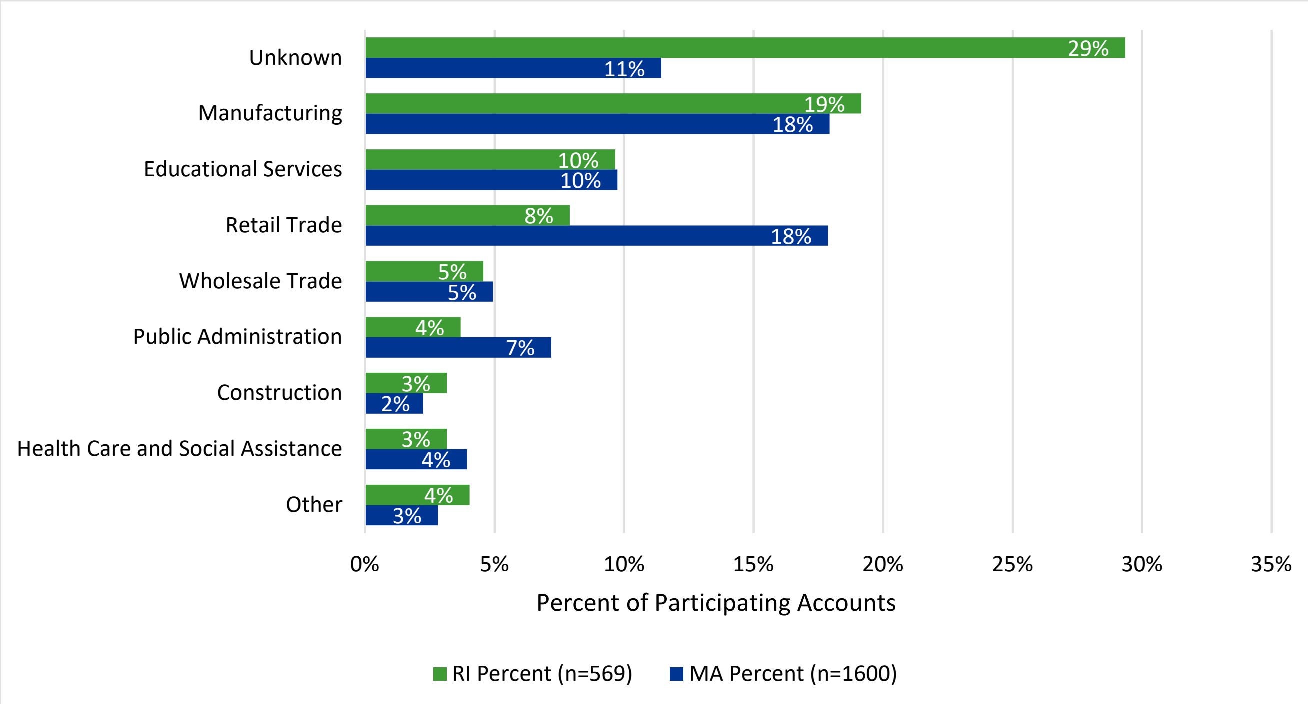
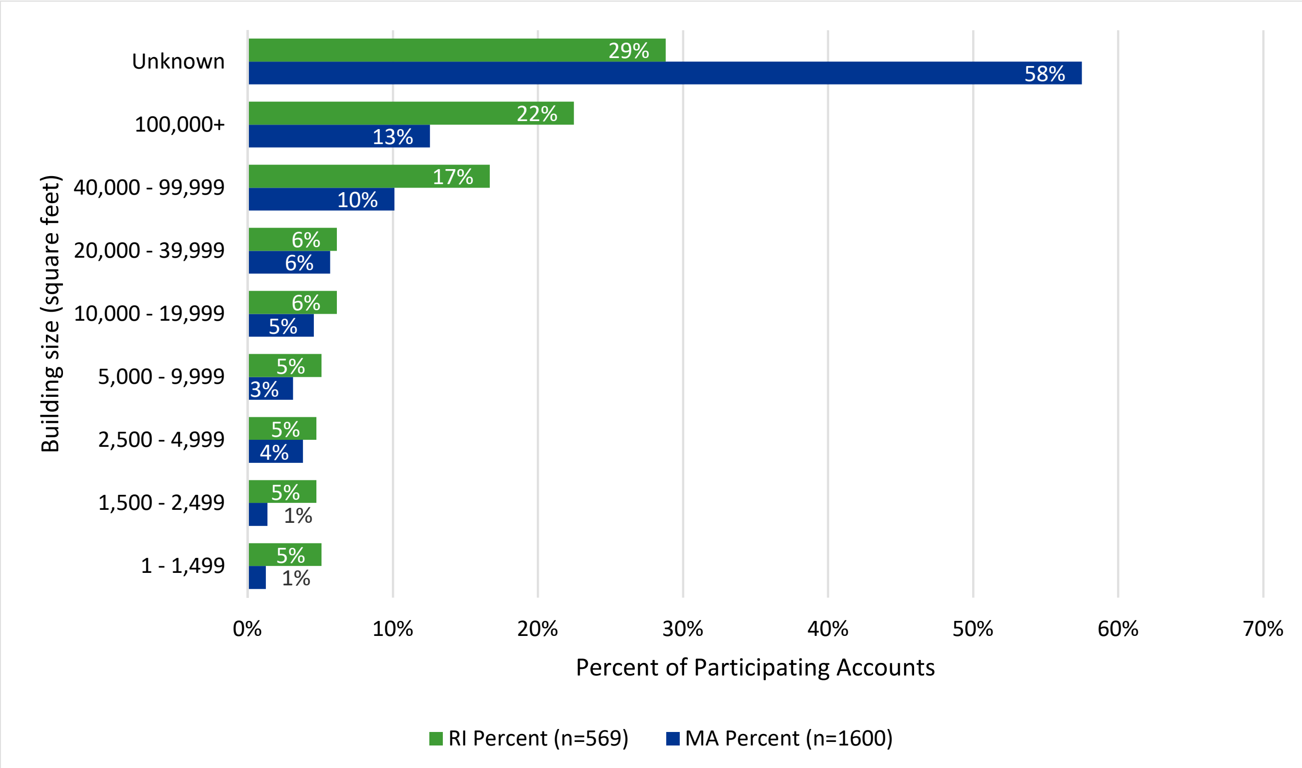


Figure 4-25 shows how the participants between 2014 through 2017 break down according to building size categories. A chi-square test indicated that there was a statistically significant difference in the distribution of participants across building types. RI participants are less likely to be categorized as "unknown". However, even when the unknown category is removed the chi-squared test is still statistically significant at $p<.01$.

Figure 4-25 Percent of 2014-2017 Participants by Building Size for Prescriptive Non-lighting



Previous Evaluation Comparisons

DNV GL completed only one impact evaluation for prescriptive non-lighting in 2014:

1. Impact Evaluation of 2014 RI Prescriptive Compressed Air Installations (MA and RI).

This evaluation originally utilized a pooled sample (Approach 3). Separate results by state are shown in Table 4-10. The overall realization rates were not significantly different. However, the error band around the RI results was very wide considering only four sites were included in that sample. Realization rates for two of the strata were significantly different.

Table 4-10 Summary of Previous Evaluation Comparisons for Prescriptive Non-lighting

Evaluation	Metric	RI	MA	Statistically Different?
Impact Evaluation of 2014 RI Prescriptive Compressed Air Installations	Population (N)	35	104	N/A
	Sample(n)	4	32	N/A
	Realization rate: kWh savings	97%	123%	n.s.
	Total end-use population kWh savings (overall)	1,023,085	4,471,422	N/A
	Average state realization rate (stratum 1)	-	12%	
	Average state realization rate (stratum 2)	-	141%	
	Average state realization rate (stratum 3)	108%	175%	**
	Average state realization rate (stratum 4)	79%	106%	**
	Average state realization rate (stratum 5)	-	132%	N/A
	Average state realization rate (stratum 6)	-	168%	N/A
	Average state realization rate (stratum 7)	-	92%	N/A
	Average state realization rate (stratum 8)	-	70%	N/A

n.s. not significantly different
** different at 90% confidence level

4.3.7 Custom Gas

Recommended Evaluation Approach

DNV GL recommends Approach 4 (independent samples) for future evaluations of this measure type. This recommendation is based on:

- Programs are similar so Approach 5 is not necessary.
- As a custom measure group, Approach 2 is not applicable. Even if it were, the differences in customer characteristics and measure mixes could limit the usefulness of Approach 2.
- Previous evaluation results differ significantly, so we do not recommend Approaches 1 and 3.
- This measure group accounts for approximately 78% of gas savings, so high rigor methods are justified. This favors Approach 4.

Program Comparisons

Figure 4-26 shows how the proportion of custom gas (reported gross) savings are distributed across end-use for the two states. RI is achieving a greater share of program savings from HVAC, a relatively equal share from other and building shares, and a lesser share from comprehensive design, process, and hot water than MA. A chi-squared test showed that the distribution across measure types was statistically significant. This distribution of savings across the two states are a key finding for our recommendation.

Figure 4-26. Proportion of Reported Gross Savings by Measure for Custom Gas

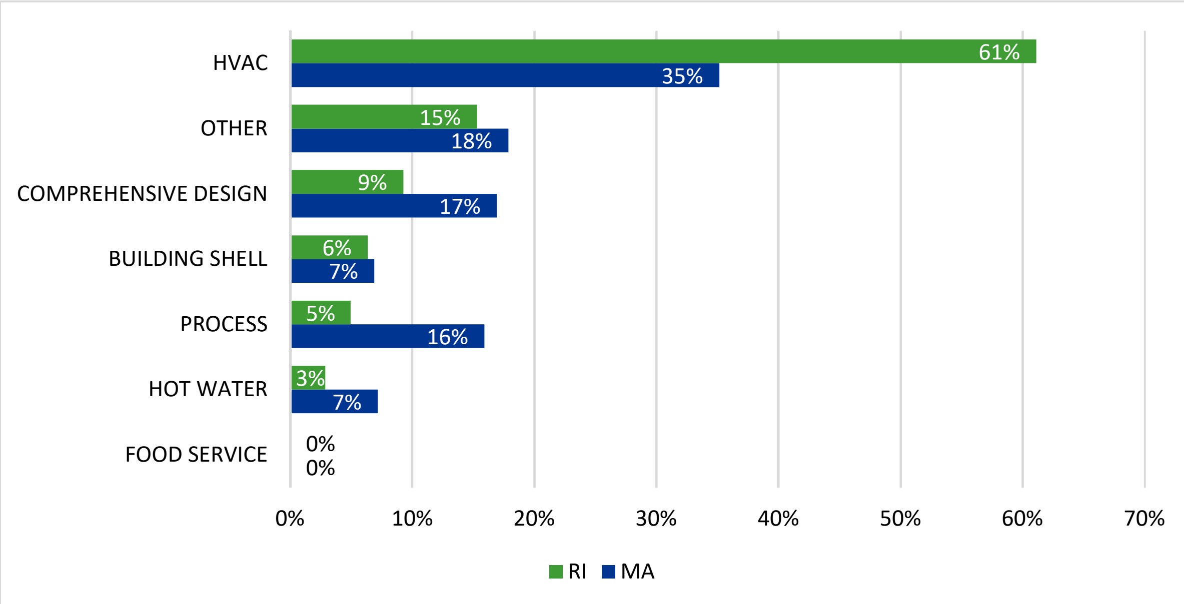


Figure 4-27 shows that the median annual consumption (averaged over 2012 to 2017) of RI participants was greater than MA participants, particularly for accounts that participated in 2012. This is a key finding for our recommendation.

Figure 4-27. Median Annual Consumption Over 2012-2017 by Participation Year for Custom Gas

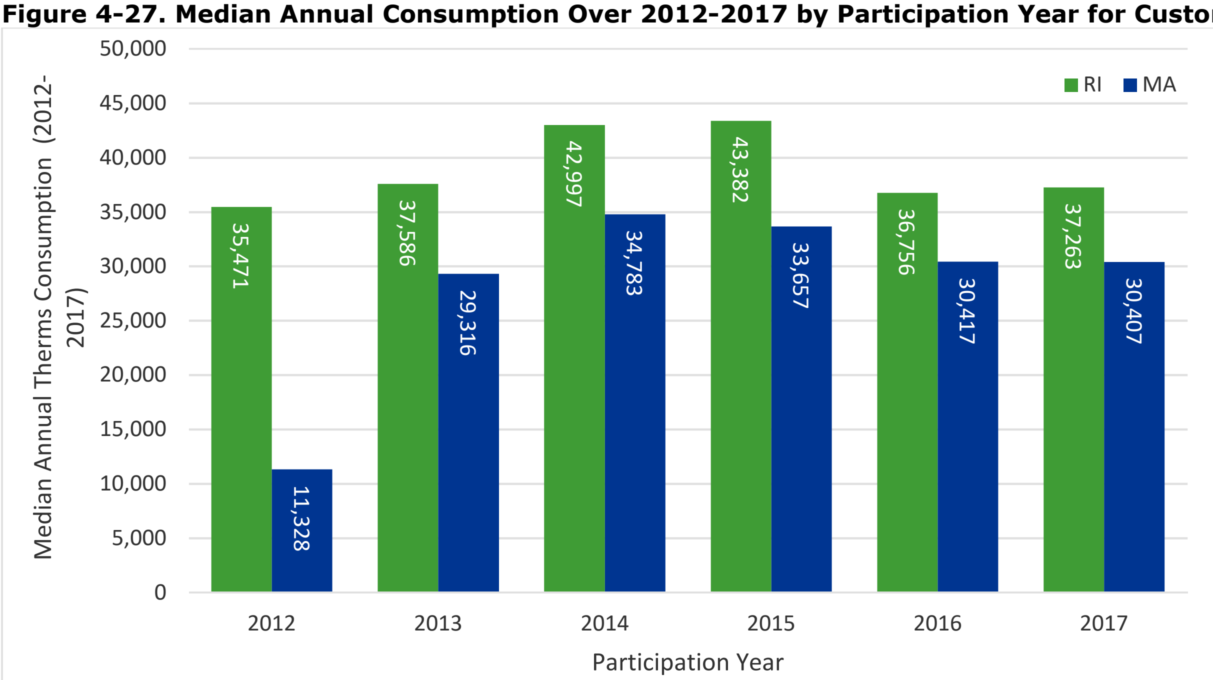


Figure 4-28 shows how the cumulative 2014 through 2017 participants are distributed according to NAICS codes. The top seven most common codes are shown; the remaining codes are summed into "Other". Across the 2014 to 2017 period, each individual code within "Other" applies to less than 4% of the accounts in MA and 3% in RI. A chi-squared test indicated the distributions were significantly different (p<.01). MA participants are more likely than RI participants to be classified within Educational Services, Accommodation and Food Services, Health Care and Social Assistance, and more likely to be classified as Unknown. This comparison is limited by the fact that the most common category in RI is unknown.

Figure 4-28. 2014-2017 Participating Accounts NAICS Codes for Custom Gas

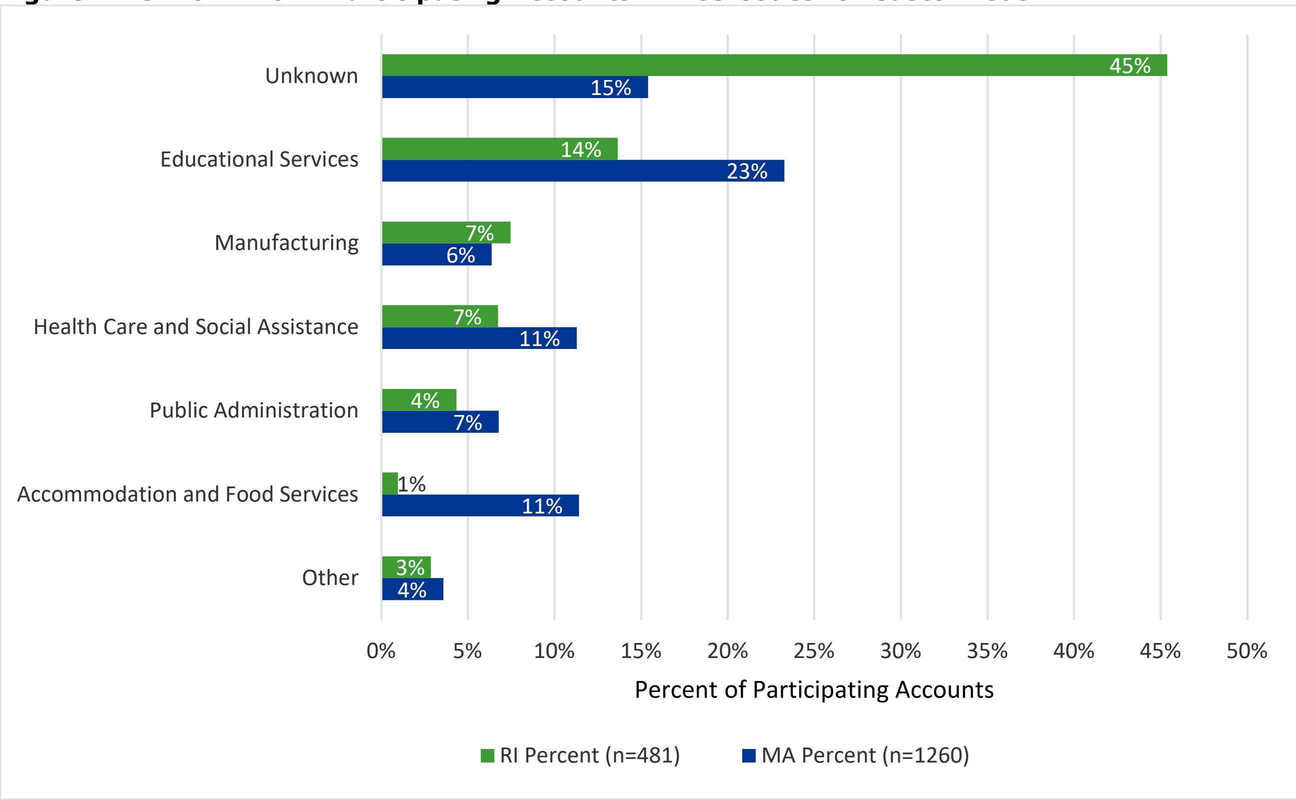
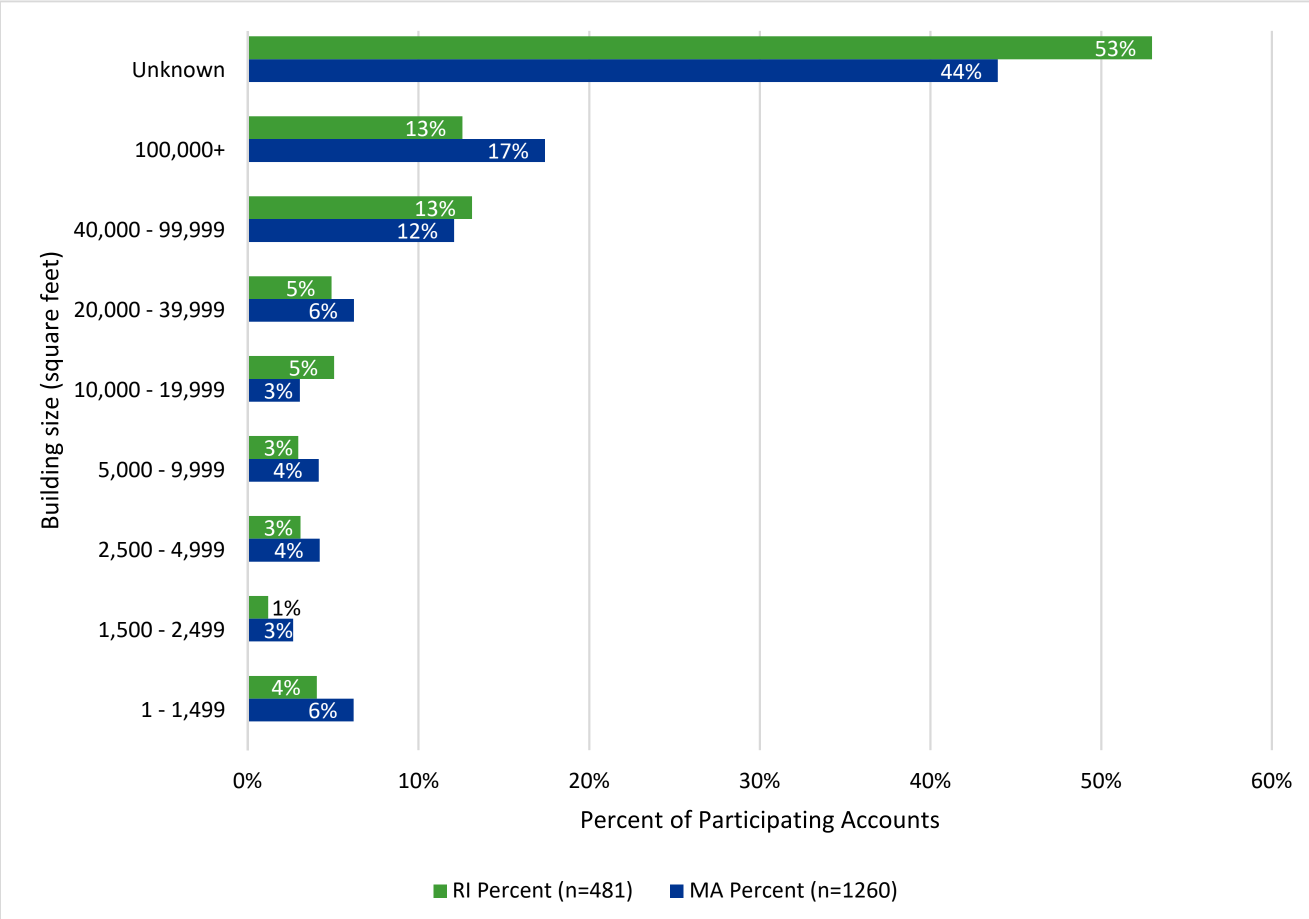


Figure 4-29 shows how the 2014 through 2017 participants break down according to building size categories. The distributions are statistically different according to a chi-square test.

Figure 4-29. Percent of 2014-2017 Participants by Building Size for Custom Gas



Previous Evaluation Comparisons

There was only one previous evaluation that applied to these participants:

1. Impact Evaluation of 2014 Custom Gas Installations in RI (MA and RI).
2. Impact Evaluation of PY2016 Custom Gas Installations in RI (MA and RI).

These evaluations were focused on presenting final realization rates for custom gas energy efficiency measures installed in RI in 2014 and 2016. Both studies used a pooled sample approach and aggregated specific site results to determine realization rates separately for National Grid's custom gas program in RI and MA (Approach 3). To determine statistical difference in overall realization rates, the choice of confidence levels was based at 20%. Overall realization rates for therms savings in both studies were significantly different (Table 4-11).

Table 4-11. Summary of Previous Evaluation Comparisons for Custom Gas

Evaluation	Metric	RI	MA	Statistically Different?
2014	Population (N)	83	111	N/A
	Sample (n)	7	14	N/A
	Realization rate: therms savings	98%	79%	*
	Population average savings per customer (therms)	26,848	16,866	N/A
	Total savings (annual therms)	2,228,376	1,872,148	N/A
2016	Population (N)	87	301	N/A
	Sample (n)	8	21	N/A
	Realization rate: therm savings	71%	88%	*
	Population average savings per customer (tracked therms)	12,813	17,081	N/A
	Total savings (annual tracked therms)	1,114,770	5,141,434	N/A

* different at 80% confidence level
n.s. difference not statistically significant

4.3.8 Prescriptive Gas

Recommended Evaluation Approach

There is insufficient information to make a strong recommendation for prescriptive gas evaluation approaches in the future. The past evaluation practices have focused on specific measure types, such as steam traps or pre-rinse spray valves, and used a combination of Approach 1 (direct proxy) and Approach 3 (pooled samples). DNV GL recommends not using Approach 1 for the measure category as a whole because it represents approximately 25% of annual gas savings. We would recommend an approach that includes at least some RI sample, but that would include Approaches 2, 3, and 4. However, if evaluators follow past approaches of evaluating very specific measure types (e.g. pre-rinse spray valves), Approach 1 could be justified for measures that represent very low proportions of savings. This recommendation is based on:

- Similar program designs and evaluation goals, so Approach 5 is not necessary.
- Savings distribution by measure type differs, so we recommend against Approach 1 if the category is evaluated as a whole.
- Previous evaluation results did not differ, but the relevance of those results is limited.
- This measure category accounts for approximately 25% of annual gas savings, so we would not recommend Approach 1 for the measure category as a whole. For specific measure types within the category that have very low participation volume (e.g. pre-rinse spray valves in 2016 and 2017), Approach 1 could be justified.

Program Comparisons

Figure 4-30 shows how the proportion of prescriptive gas reported gross savings for 2016 and 2017 are distributed across measure types for the two states. A chi-squared test showed that the distribution across measure types was statistically significant. RI is achieving a greater share of program savings from HVAC, and less from hot water and the “other” category. The other category includes codes and standards, building operator certification, and building shell measures. The majority (54%) of RI savings recorded as prescriptive gas savings are from steam traps (which appear in the HVAC category). In contrast, 8% of the MA savings are from steam traps. Even if these savings are removed, the measure mixes between the two states differ (Figure 4-31). In these program years, RI achieved less than 1% of savings from pre-rinse spray valves compared to 8% in MA. This distribution of savings across the two states are a key finding for our recommendation.

DNV GL had limited data for the prescriptive gas program. We did not have sufficient data to make comparisons of customer firmographics.

Figure 4-30. Proportion of Reported Gross Savings by Measure for Prescriptive Gas

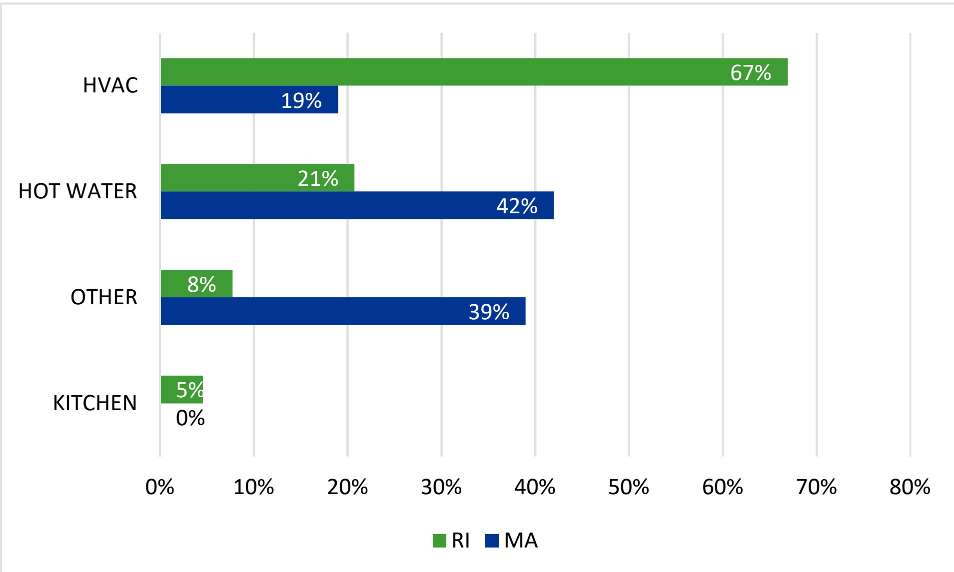
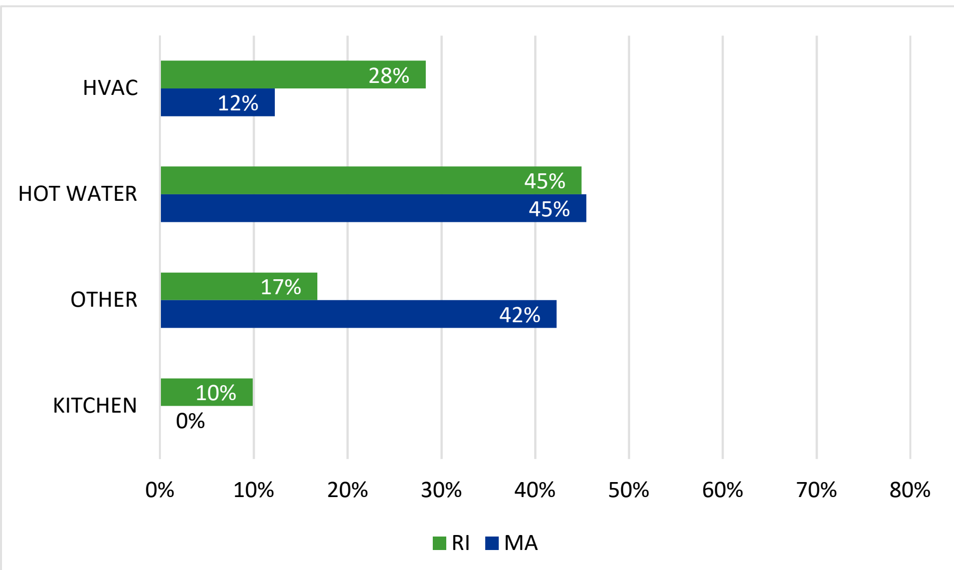


Figure 4-31. Proportion of Reported Gross Savings by Measure for Prescriptive Gas; Steam Traps Removed



Previous Evaluation Comparisons

There are 2 recent studies posted by the RI EERMC relevant to C&I prescriptive gas measures:

1. Steam Trap Evaluation Phase 2 (2017; MA)
2. Impact Evaluation of National Grid Rhode Island C&I Prescriptive Gas Pre-Rinse Spray Valve Measure (2014; RI + MA).

Study 1 is a report for Massachusetts only. 'Rhode Island' does not appear in the document. Thus, this study represents the direct proxy approach. Study 2 used the pooled sample approach. In study 2, the savings per spray valve were not significantly different between RI and MA. However, according to study 2, at the time (program year 2012), pre-rinse spray valves represented 68% of prescriptive gas savings for RI. They now (program years 2016 and 2017) account for approximately 1%. Thus, spray valves are not nearly as relevant for prescriptive gas savings now as they were.

5 FINDINGS - RESIDENTIAL

5.1 Program Design and Policy Context

DNV GL conducted in-person interviews with Residential program and evaluation staff to identify similarities and differences between RI and MA that may impact the relevance of piggybacking approaches. Overall, the interview findings imply that evaluators should exercise caution when using piggybacking methods that do not involve an independent RI sample. However, similarities in program designs increase the validity of leveraging techniques first established in MA. Table 5-1 summarizes the interview results for residential programs.

Table 5-1. Summary of Program Design and Policy Interviews: Residential

Research topic	Finding	Implication
Codes/ Baselines	How the PAs take into account codes are one of the biggest ways MA and RI differ. In the past the codes were more similar, but now MA code is more than one cycle ahead of RI. Many baseline codes are different: MA is ahead in terms of their code dictated baselines by one cycle. RI is operating under 2012 IECC, while MA is operating under IECC 2015. MA will be adopting IECC 2018 baseline, while RI will be moving to IECC 2015 in 2018. Note that code only applies to new construction, major renovation or end of useful life. MA has adopted amendments to strengthen codes relative to IECC standards, while RI has adopted weakening amendments. MA also has a stretch code established by the Green Community Act, which RI does not have. Many buildings adopt the more efficient stretch code. The MA PAs still offer incentives for code as opposed to stretch code, so this does not impact the baseline, but receive additional credit if customers adopt the stretch code.	Baseline differences make it difficult to leverage MA evaluation results for RI for programs based on code dependent measures such as new construction. This suggests that leveraging the MA evaluation approach but conducting a separate RI evaluation are more appropriate approaches to piggybacking than direct use of MA evaluation results for RI evaluations. For instances in which RI leverages MA evaluation results for measures that exist in MA but are new to RI, results should be adjusted to reflect differences in code.
Savings calculations	In MA, energy savings is modeled for Ex Ante savings for weatherization (air seal and duct sealing). RI uses deemed savings. RI also uses a different blower door test than MA.	Differences in the specific savings algorithms can limit the use of Approach 2 (shared algorithms) and Approach 3 (pooled samples).
Net savings	The states have different net-to-gross (NTG) survey cycles causing the net savings to be different. According to the interviwees, the last NTG survey in RI was in 2016 and is run approximately every 3 years. NTG results are used only prospectively in RI and in MA. MA can apply new evaluation results retrospectively, provided they are not NTG (i.e. if results come in during the planning cycle).	Previous impact evaluations have not reported on net savings. For future net savings piggybacking considerations, evaluators need to consider the timing of NTG studies to determine whether they can be leveraged prospectively.

Research topic	Finding	Implication
Planning cycle	MA files plans every 3 years, while RI files 3-year plans and annual plans. Annual plans provide RI with more flexibility than MA to change programs which may impact the comparability of programs and measures.	Measure mixes for the same programs could vary substantially. When measure mixes differ, they can be adjusted for in sampling and/or post weighting when using pooled samples approaches. Measure mix differences based on tracking data are reported for each Residential program in the subsections of 5.3. This is one factor that may impact the measure mix in an evaluation and the ability to leverage results directly or pool samples from MA evaluations. Substantial year over year changes to the measure mix in RI will dilute the relevance of MA evaluation study design for RI.
Savings goals	MA uses lifetime savings for goals, while RI uses annual savings. RI may be switching to lifetime savings in the future.	The different savings goals can impact the measures installed in each jurisdiction. Implementers are incentivized based on annual savings in RI allowing them to focus on higher annual savings measures that might not result in greater lifetime savings. MA implementors focus on lifetime savings. If there are large differences in the measure installation mix, it can substantially limit the relevance of MA evaluation results for RI. Differences in measure mix should be taken into account when pooling samples.
Program design	MA is changing the way they identify and count participants from number of units to type of building. In MA they used to count single family (SF) and multi-family (MF) by number of units in a building. According to the interviewees MAis moving to Low rise/High rise (Building type). This means they will combine SF/MF and not look at units. RI will continue to count number of units.	This will have a major impact on the ability to leverage evaluation MA results as a proxy or pool samples going forward. Once the basic unit of measure changes, regardless of how savings are calculated, it will not be possible to add sample from MA evaluations without a separate sample plan and study design.

Research topic	Finding	Implication
Measures	Both states use most of the same measures. MA sometimes introduces new measures before RI. This is particularly the case in products and appliances. Gas heating rebates in RI are half that of MA. There are other slight differences in measures. New construction has the most differences in measures where the baseline and code are different.	Differences in measures will limit the relevance of MA evaluation results for RI. If RI studies include some sample from MA studies, measure differences should be taken into account and may limit the relevance of this alternative if measure differences lead to inconsistent sample designs. However, piggybacking can be particularly useful when MA introduces a new measure. Evaluation results in MA for new measures can serve as a good estimate or proxy in RI while the measures gain sufficient market penetration to allow RI-only sampling for evaluation.
Service territories	Territories are similar.	Evaluations should account for demographic differences when leveraging results directly or pooling sample with MA evaluations.
Economic Benefits / incentives	RI's cost effectiveness tests include substantially greater economic benefits.	Use of economic benefits for screening could have an impact on the measure mix within a program.
TRM	Savings calculations in the residential TRMs are similar, but baselines can differ.	Baseline differences can limit the direct applicability of MA results to RI.

5.2 Demographic Comparisons

DNV GL obtained demographic information relevant to each state from the U.S. Census. These statistics include population and income, educational attainment, home occupancy, occupied homes by number of units in structure, number of bedrooms per home, year of construction, tenure in home, home heating fuel, and presence of a home business. The major differences and implications for program design are summarized in Table 5-2. Full statistics are reported in appendix Section 7.1The implications for evaluation are indirect and based on an assumption that program statewide demographics are characteristic of participants. Because of the extra uncertainty this introduces, we do not factor in these implications as strongly for residential as we did direct program participant differences for C&I.

In general, the demographic differences between MA and RI suggest the possibility of differences in underlying consumption and participation rates. At a minimum, evaluators should measure and attempt to control for such differences during sampling and/or post-weighting when using shared algorithm (Approach 2) or pooled samples (Approach 3).

Table 5-2. Major Demographic Differences and Implications for Program Design

Difference	Evaluation Implications
Incomes and educational attainment are higher in MA	Income is likely correlated with larger homes, which to some extent correlates with higher usage. Education might correlate with higher likelihood to participate in programs, but it is impossible to determine whether program participants have different education levels in each state.
Based on presence of children, elderly, and home businesses, homes in MA are more likely to have someone home in the middle of the day on weekdays	This could affect responses to demand response (DR) programs. Homes with people present during the day might respond less to DR signals.
People in MA are more likely to live in apartments in large buildings	This could affect the ability of MA residents to participate, for example, if the building owns the heating system. This affect could increase or decrease participation depending on how PAs address such situations.
Homes in RI are smaller	This difference likely overlaps with income differences. Smaller houses probably correlate with lower usage.
The proportion of pre-1940's construction is slightly higher in MA	A concurrent study in MA finds that homes built before 1940 are less likely to participate in efficiency programs, than homes built more recently. Thus, with slightly fewer homes in this age category, RI might expect slightly higher participation rates, all else being equal.
RI has more heating oil and less electric heat	RI homes might have lower gas and electric use than MA homes.

5.3 Review of Residential Programs

Table 5-3 presents the total proportion of savings by residential program for National Grid in RI and MA for 2015-2018. A chi-square test indicates that the variation in distribution both kWh and gas savings across programs was not statistically significant between both states.

Table 5-3 Proportion of Total National Grid Savings by Residential Program

Program	RI % Total kWh Savings	MA % Total kWh savings	RI % Total gas Savings	MA % Total gas Savings
Residential Lighting	50%	55%	-	-
Behavioral	23%	20%	38%	42%
Residential Home Energy Services	12%	13%	24%	20%
Residential Heating and Cooling Equipment	-	-	13%	18%
Residential Consumer Products	4%	2%	-	-
Low-Income Single Family Retrofit	3%	1%	6%	3%
Residential Multi-Family Retrofit	3%	2%	6%	3%
Low-Income Multi-Family Retrofit	3%	3%	8%	9%
Residential New Construction	1%	1%	5%	5%
Total	100%	100%	100%	100%

DNV GL reviewed 36 studies covering the residential sector in RI and/or MA. Many of the residential studies did not report statistics such as confidence intervals or standard errors, so meta-analytic techniques to compare results were often not possible even when by-state results were available. Unlike the C&I programs, DNV GL did not have access to raw evaluation results because other firms conducted the original evaluations.

5.3.1 Lighting

Recommended Evaluation Approach

DNV GL recommends that future evaluations utilize Approach 2 (shared algorithm) or 4 (independent samples). The key consideration is that future evaluations use an individual RI sample. Evaluations can leverage evaluation approach, data collection instruments, and if timing of efforts coincides, management of data collection efforts from MA. Depending on the specific evaluation goals (particularly if data collection related to individual homes is not planned), evaluators might be able to apply specific MA values for metrics such as delta watts (by replaced bulb type) and HOU (by room type), applied to the specific distributions of replaced bulbs and rooms representative of RI. This recommendation is based on:

- Similar program designs and evaluation goals so Approach 5 (independent studies) not necessary.
- This is a large enough program that Approach 1 (direct proxy) is not justified.
- There is mixed evidence of differences in the lighting markets in RI and MA. Such differences would likely lead to differences in ISR and ΔW. These differences are not sufficient to completely eliminate Approach 2 (shared algorithm), but do suggest the need to make adjustments to how MA parameters are used.
- Smaller homes (RI) might have fewer fixtures and thus lower savings. This is additional rationale to avoid Approach 1. It also suggests the need for adjustments in Approaches 2 or 3.
- RI effectively sets tracked gross savings directly from evaluation results. Considering the demographic and lighting market differences between RI and MA, we do not recommend approaches 1 or 3.

Study Comparisons

We identified the following five studies as having lighting measures:

1. Northeast Residential Lighting Hours-of-Use Study (2014; MA, RI, NY).
2. RI2311 National Grid Rhode Island Lighting Market Assessment (2018; MA, RI).
3. RLPNC 16-7: 2016-2017 Lighting Market Assessment Consumer Survey and On-site Saturation Study (2017; MA).
4. 2017 MA Saturation and Characterization Results (2018; MA; presentation).
5. Rhode Island 2017 Lighting Sales Data Analysis (2019; MA, RI).
6. 2018 Rhode Island Shelf Stocking Study (2019, MA, RI).

Studies 1, 2, and 3 were components of the same multi-state study conducted by NMR. These studies appear to have used Approach 4 with combined data collection, but separate samples collected for each state in the study. Study 4 presented results only and did not describe methods; results covered only MA. All four studies focused on market assessment of lighting (and sometimes other) measures. As such, they all used similar methods. Those methods included surveys, site visits with loggers, and regression modelling.

Finding Comparisons

Studies 2 and 3 provided findings that could be compared across states including bulb type saturation rates, penetration rates by room, stored bulbs, location bulbs obtained, and satisfaction with LEDs.

According to Study 2, the LED saturation rate in RI is 33%, compared to 27% in MA. In addition, the ENERGY STAR LED saturation rate is higher in RI (24%) than MA (17%). Figure 5-1 shows percent penetration of LED bulbs by room type for MA and RI. RI has a systematically higher proportion of LED bulbs in all rooms with the most pronounced differences appearing in the office and dining room spaces. Chi-squared tests revealed significant differences between the two states ($p<0.01$). The penetration data for RI originated from Study 2, which is a 2018 evaluation, where the MA originated from Study 3 which is a 2017 evaluation. Over this time the LED adoption curves for both states is quite steep, where LED saturation in MA went up by approximately 10%.[13] Study 2 ultimately concluded through a modelling approach that the overall saturation rate in MA in 2018 is likely to be equivalent to the 33% overall saturation rate found for RI.

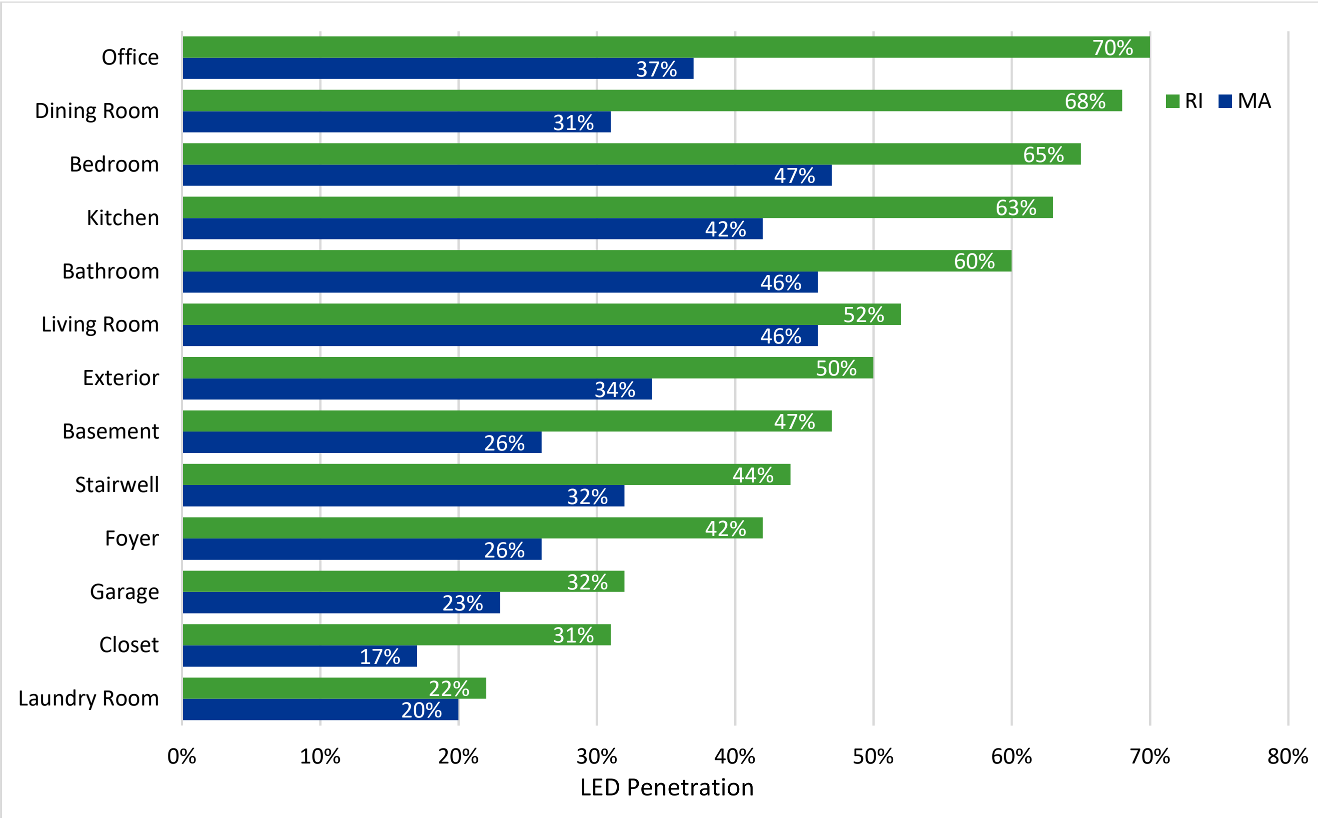
Studies 5 and 6 contain more recent data that the lighting markets are still substantially different in each state. According to study 5[14],RI had a 55% LED market share in 2017, compared to 49% in MA. According to Study 6[15], the distribution of retail shelf space dedicated to LEDs differed significantly between RI and MA.

13 NMR Group, Inc. (2018). RI2311 National Grid Rhode Island Lighting Market Assessment. Submitted to National Grid Rhode Island. *Figure 11, pg 34*

14 Figure 1 on p. 4

15 Figure 6 on p. 16

Figure 5-1. LED Penetration by Room Type[16]



Satisfaction with LED bulbs was similar in each state. Almost all the respondents in each state (RI 93%; MA: 93%) reported being "Very satisfied" or "Somewhat satisfied" with their LED bulbs.

Likewise, storage statistics in each state were nearly identical, with both RI and MA respondents indicating an average of 2.7 LEDs in storage compared to 2.3 in MA.

While HOU varies by room installation, the differences in penetration rates would suggest that RI and MA should have different overall average HOU. However, study 1 provided a comparison of overall household HOU and HOU by several different room types. MA and RI did not have statistically different HOU at the overall household level or for any room type other than exterior lighting. Therefore, MA HOU by room type, applied to the RI by-room installation rates, could be used to calculate a representative RI overall average hours of use statistic.

Delta watts will depend on the types of bulbs being replaced by LEDs. Considering the different market penetration rates in RI and MA, it is reasonable to assume the mix of replaced bulbs is also likely to differ between the two states. However, again also like hours of use, the difference in wattage between an LED and any particular type of replaced bulb is unlikely to differ between MA and RI. Based on this assumption, evaluators could use MA delta watts by replaced bulb type (e.g. LED vs. CFL), applied to an RI-specific distribution of replaced bulb types to arrive at an RI-specific value for average delta watts for RI.

[16] Note, the MA and RI studies referenced in these figures were conducted one year apart. It is possible that difference in timing accounts for some of the differences apparent in the chart.

5.3.2 Behavioral Programs

Recommended Evaluation Approach

DNV GL recommends that future evaluations can piggyback on overall approach and econometric analyses used in MA, but individual samples should be used for RI data collection and producing results (Approach 4). Approach 5 is also an option. Demographic differences are not applicable for this program because of the random assignment of the participant and control groups. We do not have a strong recommendation related to process evaluations. This recommendation is based on:

- Similar program designs and evaluation goals, so Approach 5 is unnecessary.
- Similar method of analysis, involving comparisons between randomly assigned participant and comparison groups. This makes Approach 2 inapplicable and limits the evaluation cost savings from Approach 3.

Study Comparisons

Four studies were identified as having behavioural measures:

1. RI State-wide Behavioural Evaluation: Savings Persistence Literature and Review (2017; RI).
2. RI Behavioural Program and Pilots Impact Evaluation (2014; RI).
3. Summary for MA Behavioural Program Impact Evaluations (2014; MA).

These studies all utilize econometric analyses to compare savings for randomly assigned treatment and control groups. By their nature, these types of analyses are restricted to the randomly assigned groups. The basic approach of the econometric analyses for these types of programs are usually similar. They utilize billing data to determine before-and-after variances of differences between the treatment and control groups. Because the billing data in MA and RI are similar, analysis code and tools should be transferrable, but individual samples should be used for RI data collection and producing results.

DNV GL does not have a strong recommendation for process evaluation practices. Process evaluations focusing on program design and implementation are likely relevant across states. Conservatively, DNV GL would recommend that National Grid not assume that RI participants respond to the program the same as MA participants. If reactions of MA participants are used as a proxy for RI participants, DNV GL recommends at least post-weighting the responses to match RI demographics. This reflects our standard advice about best practices for pooled samples (Approach 3).

5.3.3 EnergyWise Single Family

Recommended Evaluation Approach

DNV GL recommends the next EnergyWise Single Family evaluation utilize independent samples (Approach 4), primarily because of the substantial differences in previous evaluation results and the use of billing analysis. Approach 5 is also an option. However, because of several caveats associated with those previous evaluation results, we further recommend that if the next evaluation results in similar findings for RI and MA, that subsequent evaluations might be able to utilize pooled samples (Approach 3) if evaluators decide to use methods other than billing analysis. If evaluators pool samples in the future, our standard recommendations regarding sampling and post weighting to ensure that the MA sites represent RI characteristics distributions apply. For example, smaller homes (RI) and apartments (MA) likely have fewer opportunities to participate in this program. These differences may or may not cancel out, but they are demographic differences that could lead to differences in savings. As this is a flagship program for RI, we do not recommend a direct proxy approach (Approach 1).These recommendations are based on:

- Similar program designs so Approach 5 is not necessary.
- Billing analysis methods were used in the previous evaluation. If used in future evaluations, Approach 2 is not applicable, and the evaluation cost savings for Approach 3 are limited.
- Previous evaluation results differed substantially, although with caveats. This leads us to not recommend Approaches 1 or 3, at least for the next evaluation.
- This is a flagship residential program for RI, so higher rigor methods are justified, thus leading us to Approach 4.

Program Comparisons

Figure 5-2 and Figure 5-3 show the distributions of electric and gas savings for the EnergyWise (RI) and Home Energy Services (MA) programs. Chi-squared tests indicated that the electric distributions are not significantly different, but the gas distributions are. MA offers some measures that the RI program does not, such as furnace/boiler replacement and clothes washers.

Figure 5-2. EnergyWise Electric Savings Comparisons

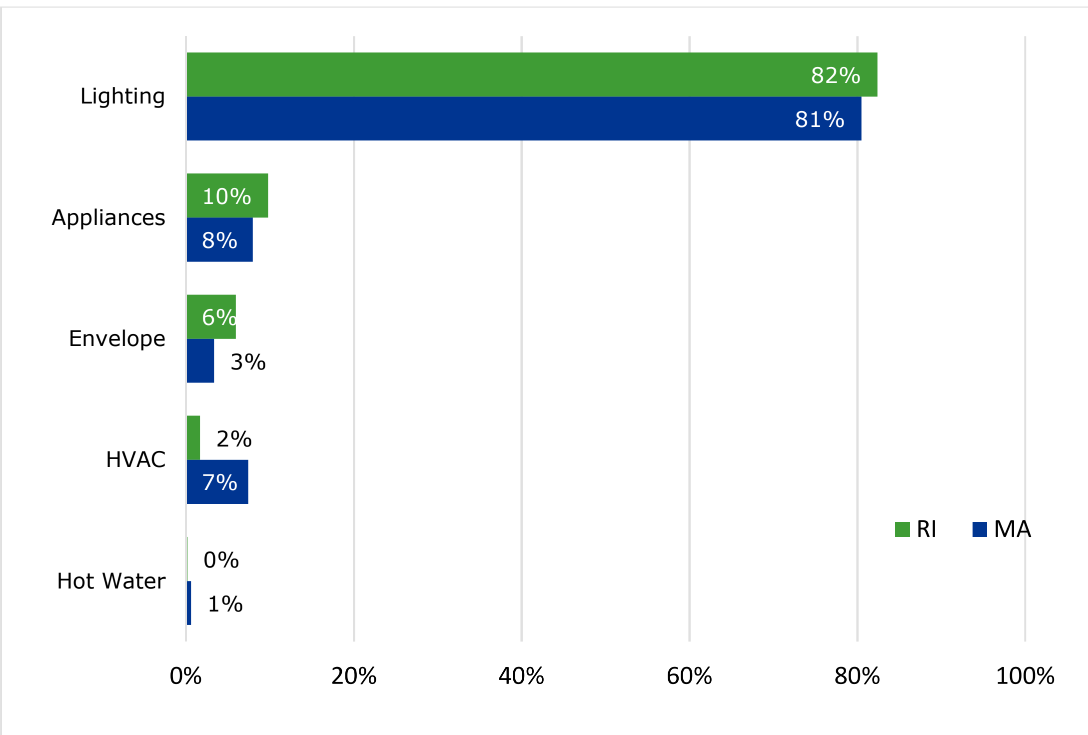
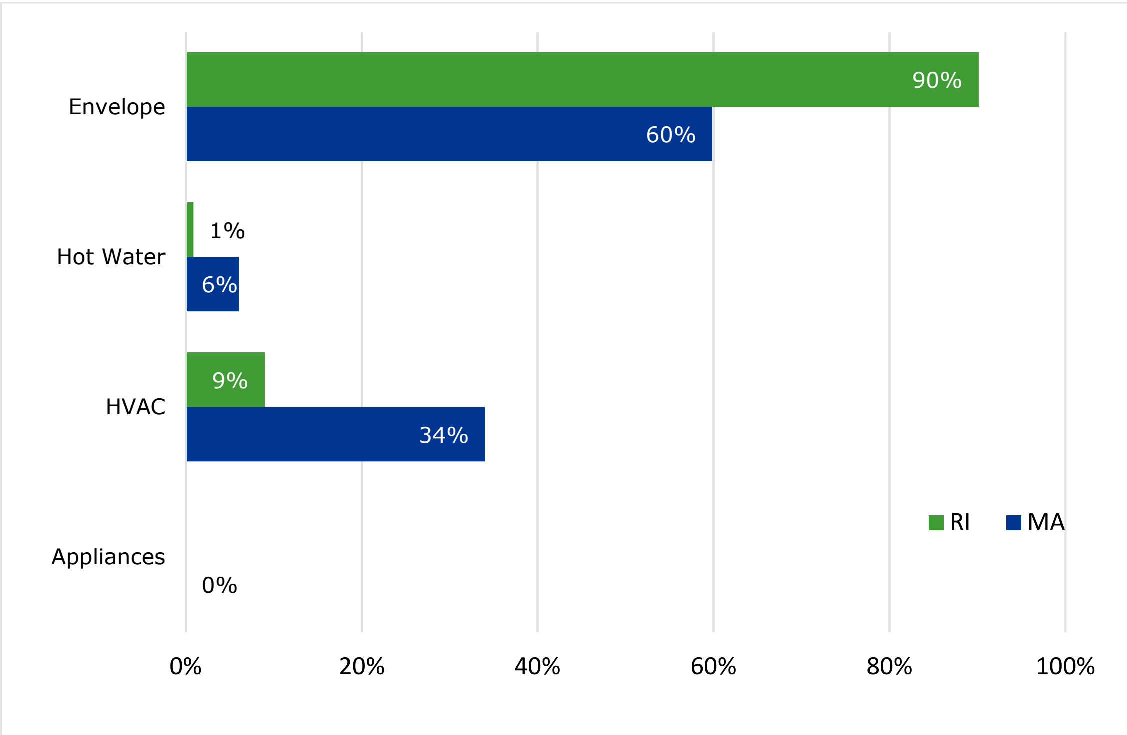


Figure 5-3. EnergyWise Gas Savings Comparisons



Study Comparisons

We identified the following study relevant to the EnergyWise Single Family program:

1. Impact Evaluation of 2014 EnergyWise Single Family Program (RI).
2. Home Energy Services Impact Evaluation (Res 34) (2018; MA)

Study 1 utilized a combination of billing analysis with a matched comparison group and engineering analysis to evaluate the RI program. It utilized an independent RI sample (Approach 4 or 5). According to the report, new methods were utilized, compared to the previous evaluation. It does not reference any similar evaluations conducted in MA. Study 2 also utilized a combination of billing analysis and engineering analysis on an independent MA sample. It additionally utilized building simulation for some analyses . Table 5-4 provides a summary of the comparable metrics documented in these two studies.

The studies contained sufficient information to compute statistical difference tests for the realization rates for weatherization for gas heated homes and for electrically heated homes. The realization rates and absolute evaluated savings for gas-heat weatherization were statistically significantly different while the realization rates and absolute evaluated savings for electric-heat weatherization were not. The realization rate for oil-heated homes was also reported, but without confidence intervals because both studies used engineering analysis to produce the estimates. These realization rates differed by 18%. The studies also provided estimates of annual therm (gas-heated homes) and kWh (electric-heated homes) savings for WiFI and standard programmable thermostats. These metrics also lacked confidence intervals because of the method of estimation. The kWh savings for standard programmable thermostats were similar. The other thermostat savings values were substantially different.

Table 5-4. Summary of Previous Evaluation Comparisons for EnergyWise Program

Metric	RI	MA	Statistically Different?
Study Year	2014	2018	N/A
Realization rate: Weatherization (Gas heating)	33%	73%	*
Evaluated Savings: Weatherization (Gas heating)	130	108	*
Realization rate: Weatherization (Electric heating)	62%	54%	n.s.
Evaluated Savings: Weatherization (Electric heating)	965	1,298	n.s.
Realization rate: Weatherization (Oil heating)	59%	77%	§
Annual Therm Savings (WiFi Thermostat)	Not reported	104	§
Annual Therm Savings (Standard Programmable Thermostat)	16.5	62	§
Annual kWh Savings (WiFi Thermostat)	30	465	§
Annual kWh Savings (Standard Programmable Thermostat)	257	278	§

n.s. not statistically significant
* different at 80% confidence level
§ estimates derived via engineering analysis so studies did not provide confidence intervals

Overall, these findings constitute differences in the previous evaluation results for RI and MA. However, several caveats apply to this conclusion. First, there is a four-year difference in the timing of these evaluations. It is possible that market changes over that period of time account for the differences in results. Furthermore, a limitation included in the RI study was that the tracking data at that time appeared to have missing or incorrect information for baseline insulation levels. The study concluded that this data anomaly could have contributed to the generally low realization rates.

5.3.4 Residential Cooling and Heating

Recommended Evaluation Approach

There was insufficient data available for Residential Cooling and Heating programs/measure for DNV GL to make a strong recommendation for or against any of the piggybacking methods covered in this study. Without the evidence to support a specific recommendation, our general advice about each piggybacking method applies. To support the use Approach 1 (applying MA results directly to RI), the programs should, at a minimum, provide evidence that the participant measure mix between furnaces, boilers, and heat pumps is similar across both states. Ideally, using Approach 2 (applying MA results to a RI-specific sample) would occur after the program had evaluation results for both states and could demonstrate that there are not significant differences on a measure-level. To use Approach 3 (pooled sample), the evaluations should make sure they sample and/or post weight results to ensure that the MA sites are representative of known RI characteristics. For example, smaller homes are likely to have smaller HVAC systems, and oil heating systems would not be eligible or would represent fuel switching. These types of potential demographic differences should be accounted for when selecting the samples in a pooled sample approach.

Program Comparisons

Figure 5-4 and Figure 5-5 show the distribution of Residential Heating and Cooling Savings for electric and gas for RI and MA. Both distributions were significantly different, based on chi-squared tests.

Figure 5-4. Residential Cooling and Heating Electric Savings Comparisons

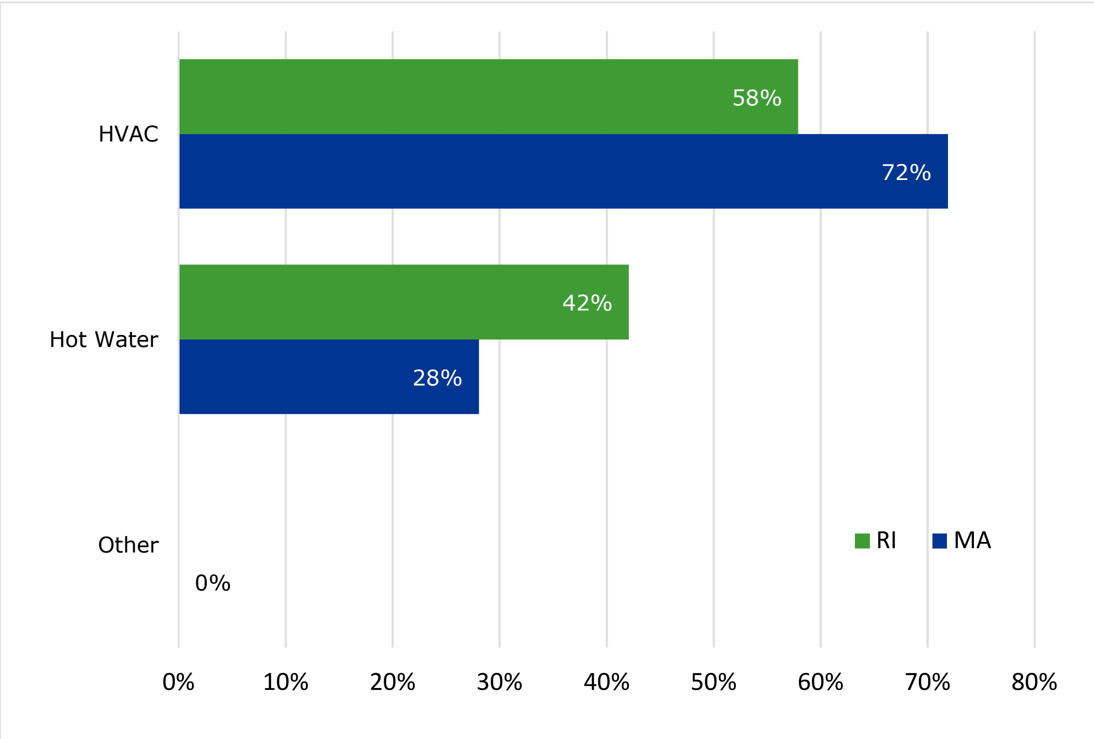
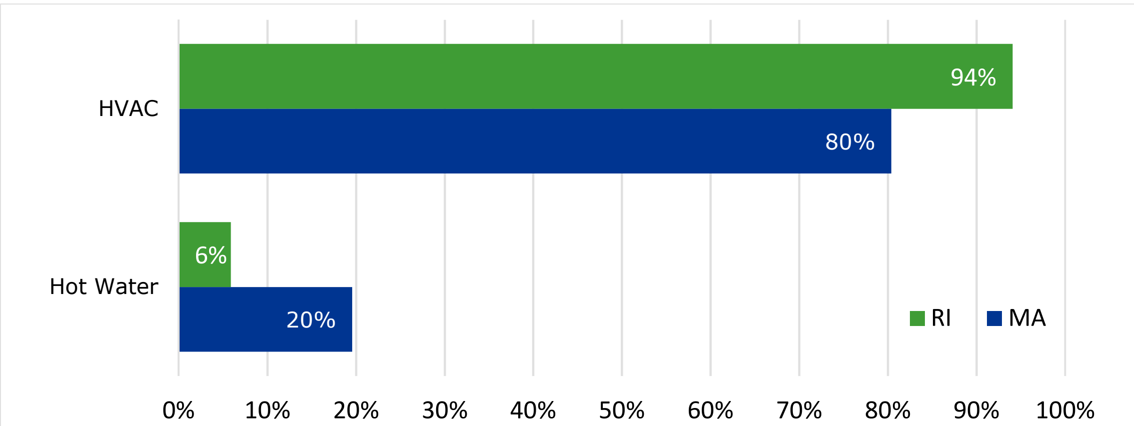


Figure 5-5. Residential Cooling and Heating Gas Savings Comparisons



Available tracking data did not break down HVAC equipment into more discrete types (e.g., furnaces, boilers, heat pumps). Furthermore, the HVAC program evaluation reports did not provide data at a sufficient level to compare participant installation rates between RI and MA. However, DNV GL compared the 2018

Rhode Island Residential Appliance Saturation Survey (RI2311) with the 2017 MA residential baseline saturation and characterization results[17] for single family homes to obtain some comparison of the measure mixes in each state (Table 5-5).[18] Based on a chi-squared test, these distributions are significantly different. Additionally, these distributions are for the general populations, which might not accurately represent program *participants.* Therefore, DNV GL recommends that RI evaluators provide additional data to demonstrate that participant measure mixes are equivalent before utilizing Approach 1 (directly use MA results for RI).

Table 5-5. Heating Systems Present in Single Family Homes

Heating System Type	RI Incidence (n=708)	MA Incidence (n=4012)
Furnace – Natural Gas	21%	22%
Furnace – Fuel Oil	7%	7%
Furnace – Other	2%	1%
Boiler – Natural Gas	35%	34%
Boiler – Fuel Oil	33%	21%
Boiler – Other	1%	2%
Ducted Heat Pump	1%	1%
Ductless Heat Pump	2%	5%

Study Comparisons

We identified three relevant HVAC studies:

1. Ductless Mini-Split Heat Pump (DMSHP) Draft Cooling Season Results (2016; MA, RI).
2. Ductless Mini-Split Heat Pump Impact Evaluation (2016; MA, RI).
3. High Efficiency Heating Equipment Impact Evaluation (2015; MA).

Methodology Comparisons

All three studies used different methods and metrics (Table 5-6).

[17] Prepare by Navigant and presented on April 12, 2018.

[18] These sources listed incidence rates for multifamily homes, but they were not comparable because the MA report broke out shared central heating while the RI report did not.

Table 5-6. Comparison of Methods Used by Previous Residential HVAC Evaluations

Study	Measures	Methods	Metrics
Study 1	Ductless Mini-Split Heat Pump (DMSHP)	Engineering analysis	Efficiency and consumption and savings during cooling season, Seasonal energy efficient ratio SEER
Study 2	Ductless Mini-Split Heat Pump (DMSHP)	Post season survey, usage assessment, Regression analysis (demand), Time series of participation	Operating hours, weighted average savings, population counts, (SEER)
Study 3	High Efficiency Heating Equipment	Survey, On-site visits. Retrofit space heating and combo heater and hot water equipment are analyzed together	Spot measurements of baseline, long term metering of post-retrofit high efficiency equipment, billing analysis, SEER

Findings Comparisons

Study 2 included installation metrics for both RI and MA (Table 5-7). The study did not include sufficient information to conduct statistical testing of the interstate differences. However, anecdotally, these findings suggest that the distribution of types of heat pumps varies between the two states.

Table 5-7. Comparison of Finding of Previous Residential HVAC Evaluations

Metric	Study	RI	MA
% Cold Climate DMSHP Units Installed	Study 2	15%	41%
% Non- Cold Climate DMSHP Units Installed	Study 2	85%	59%
% Single-Head DMSHP Units Installed	Study 2	73%	48%
% Multi-Head DMSHP Units Installed	Study 2	27%	52%

5.3.5 Consumer Products

Recommended Evaluation Approach

DNV GL recommends using the same approach that evaluators used for the 2019 evaluations of this program. This methodology involves multiplying values available from the Uniform Methods Project by characteristics of the participant population in the program tracking database. As such, there is no sampling involved, and pooled samples would not realize any evaluation budget savings. This recommended approach is essentially Approach 2 – applying an algorithm to an independent RI sample. This applies to the appliance recycling measures.

If future evaluators choose to use methods that involve field data collection, DNV GL recommends Approach 3 (pooled sample) for the initial evaluation. This approach should take account of potential demographic differences caused by differences in income and apartment-dwelling when samples are selected. The evaluation should still report RI and MA results separately.

Past RI evaluations have used direct proxy of MA results (Approach 1) for the other measures covered in this program. Those measures make up such a small amount of the residential savings that we recommend continuing to use Approach 1.

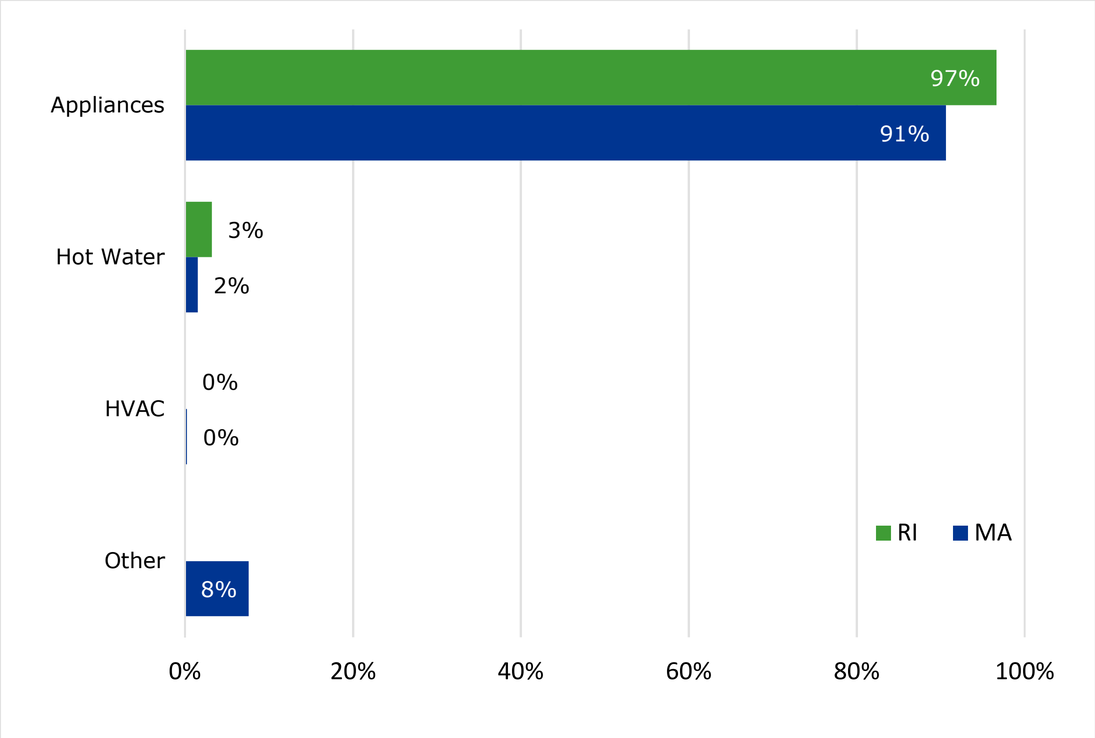
These recommendations are based on:

- Similar program designs so Approach 5 is not necessary.
- The previous evaluation used Approach 2. This method makes sense for this program and would be DNV GL's recommended approach in the future.
- Small differences in previous evaluation results cause us to not recommend Approaches 1 or 3, although we do not completely eliminate them.
- Consumer Products is a relatively small program, so higher cost methods such as Approach 4 might not be practical.
- The measures other than appliance recycling make up a very small portion of RI residential savings, so we recommend continuing to use Approach 1.

Program Comparisons

Figure 5-6 shows the distributions of electric and gas savings for the Consumer Products programs in RI and MA. Chi-squared tests indicated that the savings distribution is significantly different. However, from a practical perspective, these distributions are very similar. Both programs are getting almost all of their savings from appliances (refrigerators and freezers).

Figure 5-6. Consumer Products Electric Savings Comparisons



Study Comparisons

We analysed two recent studies of appliance recycling programs in RI and MA:

1. Appliance Recycling Impact Factor Update (2019; RI).
2. MA19R01-E Appliance Recycling Report (2019; MA).

Both studies used a method of multiplying factors reported in the Uniform Methods Project by information contained in the program tracking databases to obtain evaluated gross savings. Neither study reported precisions, but the methods multiply constants by the entire population in the tracking data, so they could be considered as a census.

Table 5-8 compares the refrigerator and freezer savings for RI and MA. RI's savings values are slightly lower than Massachusetts. Study 1 pointed out the difference for freezers and attributed it to the relatively younger age of freezers in RI. Refrigerators in RI are also slightly younger. Other reported characteristics were similar in each state.

Table 5-8. Savings Comparisons by Measure Type: Consumer Products

Measure		RI	MA
Refrigerators	Gross savings	1,004 kWh	1,027 kWh
	Adjusted Gross savings	883 kWh	904 kWh
Freezers	Gross savings	724 kWh	769 kWh
	Adjusted Gross savings	492 kWh	523 kWh

5.3.6 Income Eligible Single-Family

Recommended Evaluation Approach

DNV GL recommends using an independent sample for RI sites in the next evaluation (Approach 4).[19] Approach 5 is also an option. If that evaluation generates similar results for both states, this program is small enough for later evaluations to use a less costly approach including Approaches 1, 2, or 3. This recommendation is based on:

- Similar program designs, so Approach 5 is not necessary.
- Previous evaluation results differed, so we do not recommend Approaches 1 or 3. However, these evaluations occurred several years apart, which could account for the differences.
- Billing analysis methods were used in the previous evaluation. If used in future evaluations, Approach 2 is not applicable, and the evaluation cost savings for Approach 3 might be limited.
- Differences in the distribution of savings across measures and differences in previous evaluation results within individual measure types also lead us to not recommend Approach 2.

Program Comparisons

Program designs and eligible measures are similar.

[19] Potential demographic differences would not be an issue in independent samples.

Figure 5-7 and Figure 5-8 show the distributions of electric and gas savings for the single family low income retrofit programs in RI and MA. Chi-squared tests indicate that the electric distributions are statistically significantly different, but the gas distributions are not.

Figure 5-7. Income Eligible Single-Family Electric Savings Comparisons

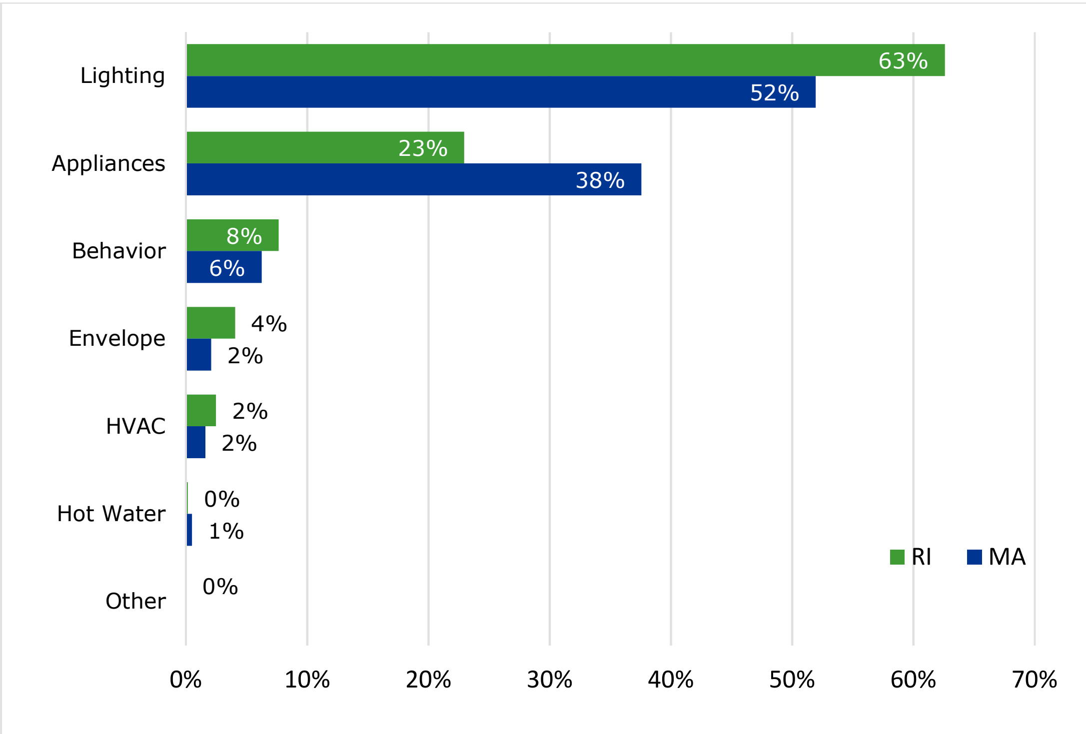
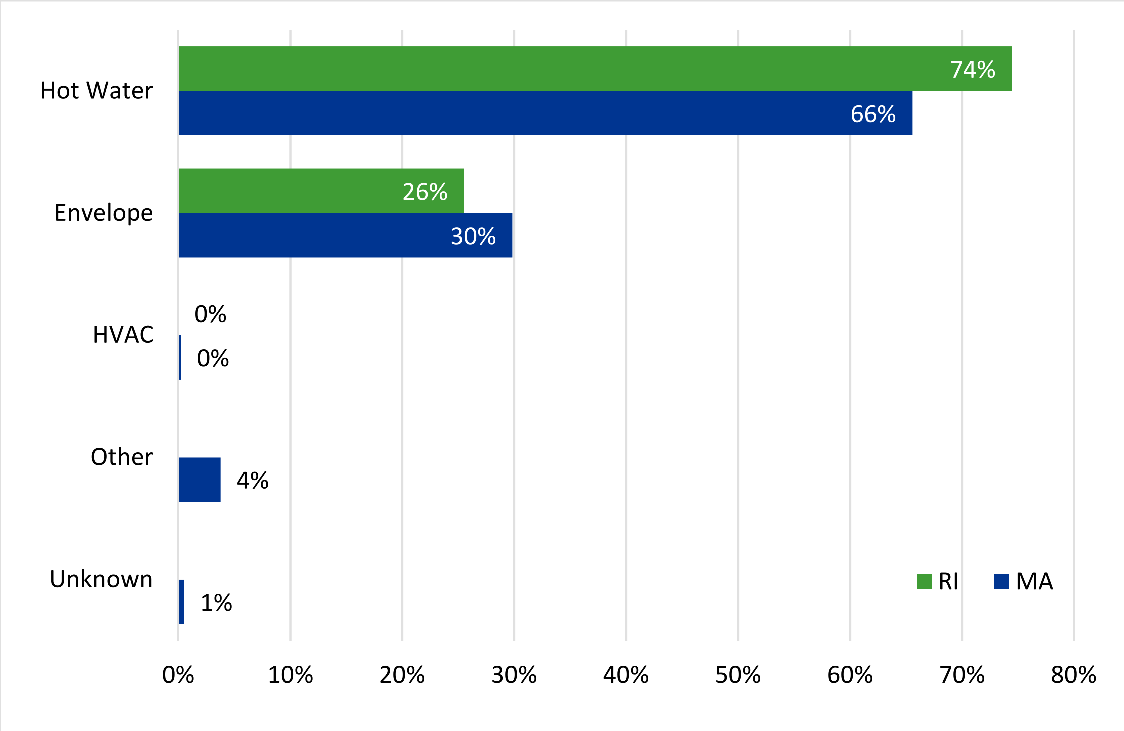


Figure 5-8. Income Eligible Single-Family Gas Savings Comparisons



Study Comparisons

1. Low-Income Single-Family Program Impact Evaluation (2012; MA).
2. Impact Evaluation of the Income Eligible Services Single Family Program (2014; RI).

Both studies utilized a billing analysis and engineering review and reported only for an individual state. Some of the results in study 2 were based directly on those documented in study 1 (e.g. electric savings due to weatherization and heating system replacement). Thus, study 2 used a mix of Approach 4 and Approach 1.

Findings Comparisons

The measures for which study 2 conducted a new billing analysis for RI-specific sample were gas savings for insulation and air and duct sealing, and heating system replacement. The measures that study 2 conducted new billing analyses for electric savings were CFLs and LEDs, refrigerator replacement, freezer replacement, and the catch-all "Other" measure category after all other specific measures were considered. All but the "Other" category had comparable values reported in study 1.

Table 5-9 compares the per measure type savings reported by each study. The evaluated gas savings for insulation, air sealing, and duct sealing were significantly different. The evaluated gas savings for heating systems were not significantly different. There was insufficient information available to conduct statistical testing of the savings differences for the other measures. However, the magnitude of those differences is substantial, and in all cases outside the confidence intervals of the RI estimates.

Table 5-9. Savings Comparisons by Measure Type: Income Eligible Single Family

Measure		RI savings	MA savings
Insulation, air, and duct sealing (gas)	n	162	223
	Savings	16%*	29%*
	Precision (90% confidence)	±21%	±8%
Heating system replacement (gas)	n	29	43
	Savings	18%	23%
	Precision (90% confidence)	±33%	±16%
CFLs[20]	n	1,552	Not reported
	Savings	22 kwh/bulb	45 kWh/bulb
	Precision (90% confidence)	±17%	Not reported
Refrigerator replacement	n	590	597
	Savings	384 kwh	762 kWh
	Precision (90% confidence)	±28%	Not reported
Freezer replacement	n	53	119
	Savings	484 kWh	239 kWh
	Precision (90% confidence)	±65%	Not reported

* Significantly different at 90% confidence level

5.3.7 EnergyWise Multifamily / Income Eligible Multifamily

Recommended Evaluation Approach

DNV GL recommends that future evaluations use Approach 4, or Approach 2 if different evaluation methods are used than in the past.[21] These recommendations are based on:

- Similar program designs and evaluation goals so Approach 5 is not necessary.
- Econometric analysis methods were used in the previous evaluation. If used in future evaluations, Approach 2 is not applicable, and the evaluation cost savings for Approach 3 are limited.
- Past evaluation results differed significantly, so we do not recommend Approaches 1 or 3.
- This is a small program, so lower cost approaches are justified.

Program Comparisons

Figure 5-9 and Figure 5-10 show how the proportion of savings are distributed across the two states for electric and gas measures for the two multifamily programs. Chi-squared tests indicated that the distribution of electric measures for Residential Multi-family Retrofit were not statistically different. The distributions of savings for gas measures for Residential Multi-family Retrofit. The distributions of both the electric and gas measures for Income Eligible Multi-family were statistically different at a 95% or higher confidence level.

[20] Study 2 included LEDs, but Study 1 did not because of age differences. To provide an apples-to-apples comparison, this table uses only the CFL data from Study 2.

[21] Demographic differences would not be an issue with independent samples.

Figure 5-9. Residential Multifamily Retrofit Savings Distributions

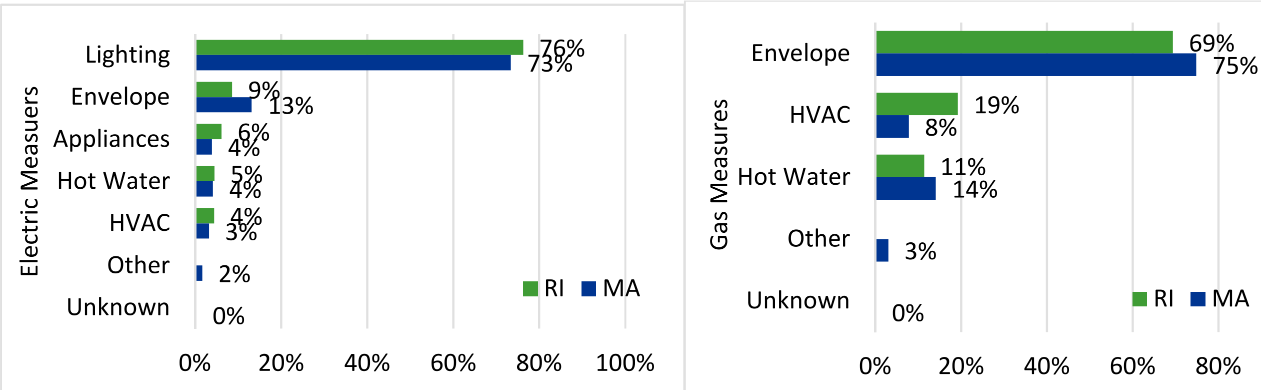
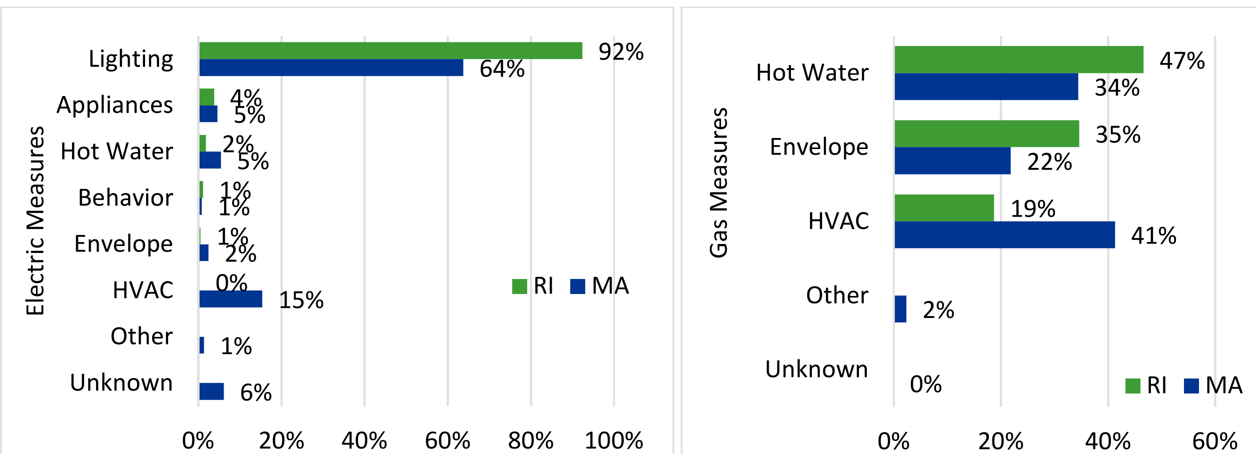


Figure 5-10. Income Eligible Multifamily Savings Distributions



Study Comparisons

Three studies were identified as having behavioural measures:

1. 2013 National Grid Multifamily Program Gas and Electric Impact Study (2016; MA).
2. Multifamily Impact Evaluation National Grid Rhode Island 2016 (2016; RI).
3. Multi-Family Program Impact and Net-to-Gross Evaluation (RES 44) (2017; MA).

Methodology Comparisons

These studies utilized econometric analyses to compare savings for participants and matched comparison groups. By their nature, these types of analyses are restricted to these groups. For these analyses, the matched comparison groups are selected by evaluators to match the characteristics of the participants relevant to the evaluation. These efforts are usually based on billing records, so combining MA and RI samples would not reduce evaluation efforts. Therefore, we do not recommend pooling samples. The basic approach of the econometric analyses for these types of programs are usually similar. They utilize billing data to determine before-and-after differences of differences between the participant and comparison groups. Because the billing data in MA and RI are similar, analysis code and tools should be transferrable.

Results Comparisons

Studies 1 and 2 had overall electric and realization rates reported in a manner that allowed us to compare results across states. Electric realization rates for the multifamily program were statistically significantly different. Gas realization rates were not statistically different, however, the magnitude of the difference was similar to that for electric. Considering these differences, DNV GL would not recommend using MA results as a direct proxy for RI programs (Approach 1).

Table 5-10. EnergyWise Multifamily Realization Rate Comparisons

Metric	RI[1]	MA[2]
Electric population	2,795	31,674
Electric Realization Rate (RR)	57.3%*	24.4%*
Electric RR Precision (90% confidence)	±31%	±49%
Gas Population	516	7,874
Gas RR	52.7%	87.3%
Gas RR Precision (90% confidence)	±31%	±64%

1 Results are from Multifamily Impact Evaluation National Grid Rhode Island 2016
2 Results are from 2013 National Grid Multifamily Program Gas and Electric Impact Study (MA)
* Difference statistically significant at p<.10 level

5.3.8 New Construction, Code Compliance and Building Characteristics

Recommended Evaluation Approach

DNV GL recommends that future evaluations utilize Approach 4. Approach 5 is also an option. This recommendation is based on:

- Code compliance samples must be state-specific. To assess code compliance *in RI*, an independent RI sample is necessary. This indicates the need for Approaches 2 or 4.
- Code differences in MA and RI suggest that using MA parameter values is not always applicable in RI. This reduces the applicability of Approach 2.
- Demographic differences can affect the systems installed in homes, and savings distributions differ which indicates that the programs are achieving savings through different measure mixes. This further reduces the applicability of Approach 2.

Program Comparisons

Figure 5-11 and Figure 5-12 show how the proportion of savings are distributed across the two states for electric and gas measures for the residential new construction programs. Chi-squared tests indicated that both distributions are significantly different at a 95% or higher confidence level.

Figure 5-11. Residential New Construction Electric Savings Distributions

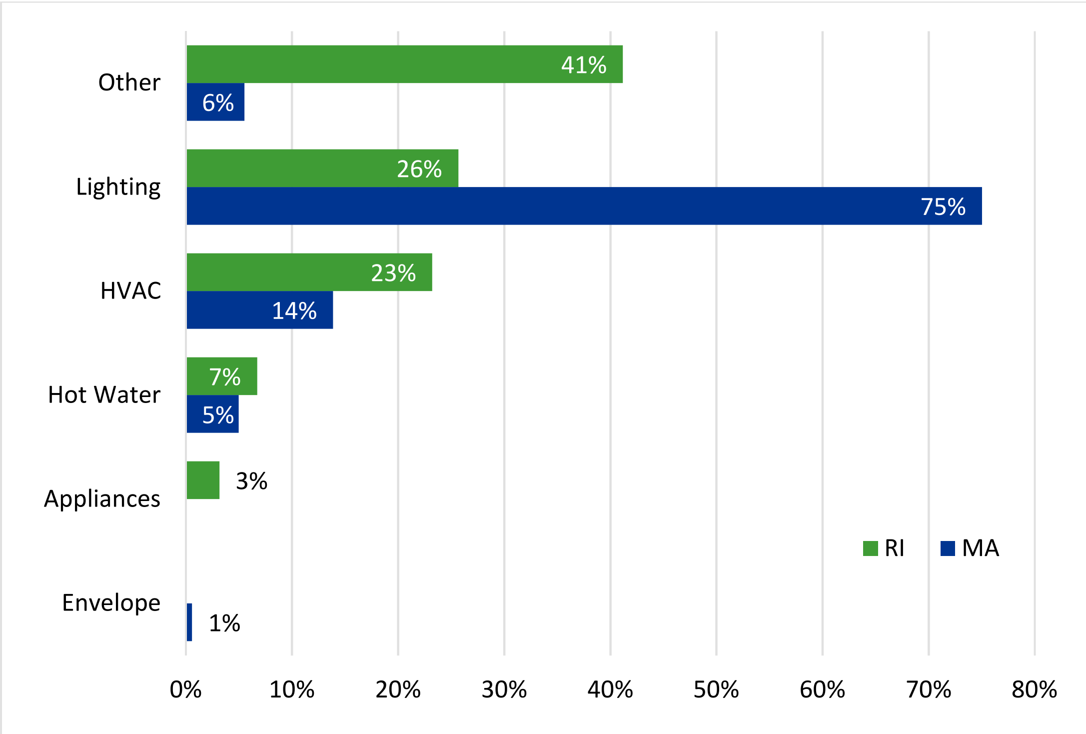
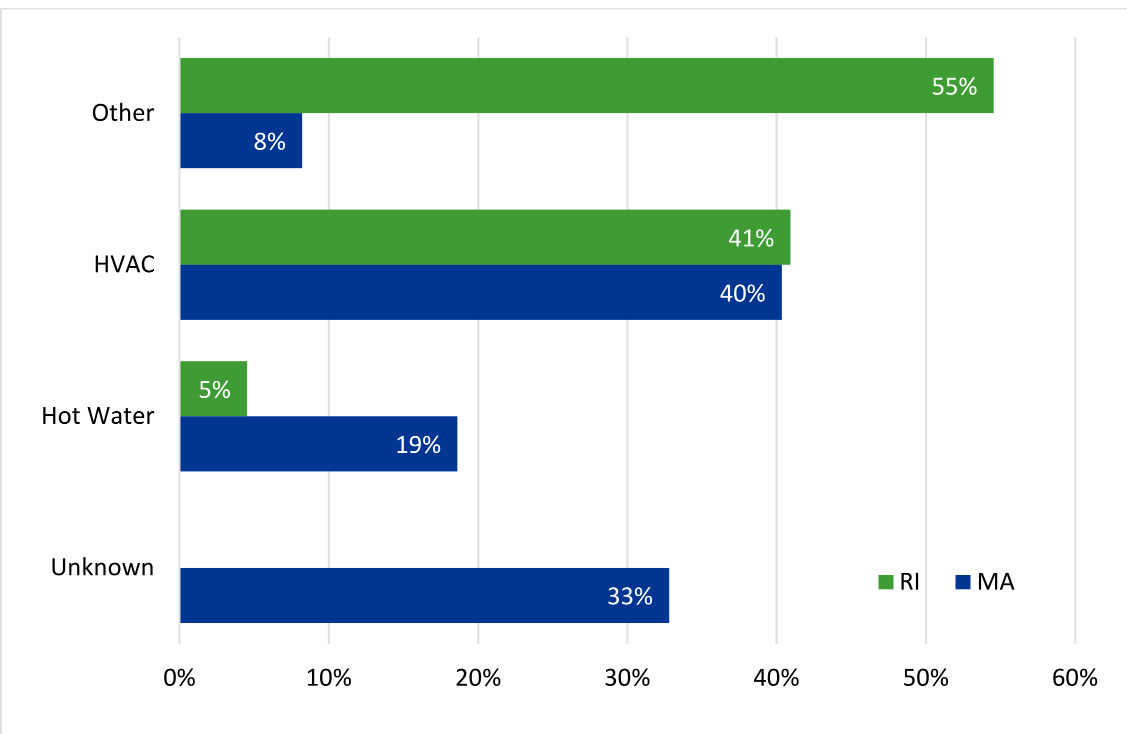


Figure 5-12. Residential New Construction Gas Savings Distributions



Study Comparisons

Four studies addressed this measure category:

1. RI Baseline Study of Single-Family Residential New Construction (2018; RI).
2. 2017 MA Single-Family New Construction Mini-Baseline/Compliance Study (2017; MA).
3. Final 2017 UDRH Inputs for the RI Residential New Construction Program (2017; RI).
4. 2015-2016 MA Single-Family Code Compliance/Baseline Study: Volumes 1 – 5 (2015; MA).

These reports did not provide precisions, confidence intervals, or measures of variance, so we were unable to conduct statistical tests of differences in the values.

Methodology Comparisons

All four studies utilized site visits that collected detailed information about building characteristics. While most of the costs of such site visits would recur in future studies, the actual data collection and analytic tools should be largely reusable.

Findings Comparisons

RI homes score slightly higher than MA home on Home Energy Rating (HER) index scores. They tend to have worse flat ceiling and floor-to-basement insulation than in MA. However, RI homes have higher air infiltration and leakier ducts. RI homes are more often heated by propane and by boilers than those in MA.

Table 5-11 shows a comparison of Home Energy Ratings (HER) Index scores for comparable studies. RI homes scored slightly better than MA homes. This comparison is between study 1 and study 4, which have a three-year difference. It is possible that time difference could account for some of the differences in reported metrics.

Table 5-11. HER Index Scores for Studies in the Building Characteristics Measure Group

HER index score	RI (Study 1)	MA (Study 4)
Number of homes	40	50
Minimum (best)	33	38
Maximum (worst)	100	90
Average	73	70
Median	72	70

Table 5-12 shows a comparison of R-Values for comparable studies. In general, there are differences in R-Value across different metrics between the RI sample in study 1 and the MA sample in study 4. This comparison is between study 1 and study 4, which have a three-year difference. It is possible that time difference could account for some of the differences in reported metrics.

Table 5-12. Average R-Values for Studies in the Building Characteristics Measure Group

Insulation type	RI (Study 1)	MA (Study 4)
Conditioned to Ambient Wall Insulation		
Number of homes	40	50
R-Value (average)	19.8	20.6
Flat ceiling insulation		
Number of homes	32	48
R-Value (average)	36.1	42.4
Vaulted ceiling insulation		
Number of homes	22	31
R-Value (average)	29.4	31.2
Floor insulation over unconditioned basements		
Number of homes (average)	22	44
R-Value	20	31.8

Table 5-13 shows a comparison of duct leakage and air infiltration statistics for comparable studies. The results show substantial differences in total duct leakage between the states. The comparison in Study 1 indicates that the 2012 IECC code in RI established a duct leakage requirement of 8 CFM25, so that MA homes are held to stricter requirements. There is a substantial difference between the states for air infiltration as well. There is a one-year time difference between these two studies. It is possible, but seems unlikely that the differences in reported metrics are partially due to that time difference. These are large differences for only a one-year difference to account for, and the RI study (where leakage and infiltration are worse) is more recent.

Table 5-13. Duct Leakage and Air Infiltration Statistics

Metric	RI (Study 1; n=36)	MA (Study 2; n=98)
Average duct leakage (CFM25/100 sq. ft. CFA)	8.6	3.9
Average air infiltration (ACH50)	5.3	3.6

Table 5-14 shows a comparison of heating equipment statistics for comparable studies. RI has a higher incidence of propane heating (and a lower incidence of natural gas service). RI homes are also more frequently heated by boilers, less often by furnaces. This comparison is between study 1 and study 4, which have a three-year difference. It is possible that time difference could account for some of the differences in reported metrics.

Table 5-14. Heating Equipment Statistics

Metric	RI (Study 1; n=40)	MA (Study 4; n=50)
Primary heating fuel		
Propane	45%	34%
Natural gas	42%	64%
Oil	6%	2%
Electric	7%	-
Heating system type		
Furnace	70%	90%
Boiler	17%	8%
Combined appliance	6%	2%
GSHP	5%	-
ASHP	2%	-

5.3.9 Demand Response Programs

Recommended Evaluation Approach

DNV GL recommends that future evaluations can piggyback on overall approach and econometric analyses used in MA, but individual samples should be used for RI data collection and producing results (Approach 4). If there is insufficient participation volume in RI to produce an independent sample, then pooling samples (Approach 3) is justified. DNV GL does not recommend using MA results as a direct proxy for RI (Approach 1) at this time, because of the differences in results between the two states for the two reports we analysed. This recommendation is based on:

- Similar program designs so Approach 5 is not necessary.
- Evaluations for these programs almost always use billing analyses. Thus, Approach 2 is not applicable and Approach 3 would result in limited evaluation cost savings.
- Previous evaluation results do not differ, making Approaches 1 or 3 possible. However, differences were large enough in absolute terms to suggest caution when using Approaches 1 or 3.
- The demographic difference that RI has more household members home during the day could affect response to DR events. This leads us away from Approaches 1 or 3.
- Thermostat data can be difficult to obtain, which might make Approach 4 impractical.

Program Comparisons

The DR programs are very similar. They are offered at the same time and have the same peak periods.

Study Comparisons

DNV GL analysed two studies on the DR programs for thermostat measures:

1. 2017 Seasonal Savings Evaluation (2018; MA, RI).
2. 2017 Residential Wi-Fi Thermostat DR Evaluation (2018; MA, RI).

Methodology Comparisons

Both studies primarily used logging data provided by the smart thermostats themselves for analysis of participation and savings. Such data is often difficult to obtain because smart thermostat vendors often consider the data proprietary and will not share it. The availability of the thermostat data itself will most likely be the most limiting factor for future evaluations. If there is enough data for an independent RI sample (Approach 4), that would be the most robust approach. But if the available data only allows for pooling (Approach 3), or proxy (Approach 1), then those methods are justifiable in order to utilize the thermostat data.

Study 1 additionally leveraged an experimental design (random encouragement) to facilitate comparisons between an opt-in group and a randomly selected comparison group. This is an excellent method for obtaining comparison groups. Similar to the thermostat data, practical considerations related to setting up this type of study probably override concerns about pooling samples. Approach 4 is the best choice if there is sufficient RI participation to obtain an independent RI sample. If that volume of participation is not available, Approach 3 with pooled samples is justified.

DR programs, in general, often use billing analysis approaches to estimate savings. Pooling samples for those analyses provides minimal evaluation cost savings.

Findings Comparisons

Table 5-15 lists the metrics we found to be comparable across previous studies. Study 1 shows there are no statistically significant differences for average energy savings and average energy savings per device between MA and RI at the 90% confidence level.[22] While statistical tests were not significant, the differences are large enough to suggest caution in applying MA results directly to RI sites. RI had higher savings per device at 15.9 kWh and demand savings per device at 0.03 kW when compared to MA' energy saved per device of12.4 kWh and demand savings per device of 0.02 kW.

[22] This confidence level was based on the confidence levels reported the original studies.

Table 5-15. Summary of Previous Evaluation Comparisons for Thermostat Measures

Metric	Study	RI	MA	Statistically Different?
Average daily savings per device	Study 1	0.49 kWh	0.34 kWh	n.s.
	Study 2	0.47 kWh	0.44 kWh	§
Total savings per device	Study 1	15.9 kWh	12.4 kWh	§
	Study 2	N/A	N/A	N/A
Demand savings per device	Study 1	0.03 kW	0.02 kW	n.s
	Study 2	0.61 kW	0.60 kW	§
Total percent savings	Study 1	N/A	N/A	N/A
	Study 2	74%	78%	§
Increase in overall program savings between 2017 and 2018	Study 1	N/A	N/A	N/A
	Study 2	298%	168%	§

n.s. not statistically significant
** different at 90% confidence level
§ variance estimates unavailable, statistical difference test not possible

6 CONCLUSIONS AND RECOMMENDATIONS

6.1 C&I Recommended Approaches by Measure Group

Our interviews with C&I program staff revealed regulatory environments, program designs, and evaluation goals are similar across RI and MA. The programs offer the same measures and where trade allies are involved, use many of the same trade allies. The C&I custom programs use many of the same trade allies and general methods. Interviewees said there are differences in gross savings baselines, some of which we specifically confirmed by reviewing the technical reference manuals with National Grid staff. Analysis of program tracking and billing databases revealed that most programs had different measure mixes and participant characteristics. Such differences can be accounted for in sampling and post-weighting, and we cite where we found differences for completeness. Most of the past evaluation results differed between states; a few were similar. The past approach, and DNV GL's recommendations for future piggybacking approaches for different C&I measure groups, are listed in Table 6-1 along with the supporting key reasons.

Table 6-1. Recommended Approaches – C&I

Measure Group	Past Approach	Recommended Approach	Key Reasons
Downstream Prescriptive Lighting	**Approach 5** – Independent Study	**Approach 4** – Independent Sample or **Approach 5** – Independent Study	Similar programs Past evaluation results differ Large program Rapidly changing technology
Upstream Lighting	**Approach 3** – Pooled Sample	**Approach 4** – Independent Sample	Similar programs Tracked savings differ Past evaluation results differ Large program Rapidly changing technology
Custom Electric Non-lighting	**Approach 3** – Pooled Sample	**Approach 4** – Independent Sample	Similar programs Custom programs Same engineering firms Past evaluation results differ
Custom Electric Lighting	**Approach 3** – Pooled Sample	**Approach 4** – Independent Sample	Similar programs Custom program Same engineering firms Past evaluation results differ
Small Business Electric	**Approach 3** – Pooled Sample	**Approach 3** – Pooled sample, with adjustments for participants or **Approach 1** – Direct Proxy if limited to non-lighting	Similar programs Past evaluation results same Customer characteristics differ Small proportion of savings
Prescriptive Electric Non-lighting	**Approach 3** – Pooled Sample	**Approach 4** – Independent Sample Or **Approach 3** – Pooled Sample if individual measure types evaluated	Similar programs Past evaluation results differ, though not significant Small proportion of savings
Custom Gas	**Approach 3** – Pooled Sample	**Approach 4** – Independent Sample	Similar program Custom program Past evaluation results differ Contributes 75% of gas savings
Prescriptive Gas	**Approach 1** – Direct Proxy, **Approach 3**- Pooled Sample	Insufficient evidence to make strong recommendation	Insufficient evidence Measure mixes differ Previous evaluations minimally applicable

6.2 Residential Recommended Approaches by Measure Group

Our interviews with residential program staff revealed regulatory environments, program designs, and evaluation goals are similar across RI and MA. Interviewees said there are differences in gross savings baselines (some confirmed via TRM review with National Grid staff) and that MA and RI differ in how they count participation for single-family or multi-family housing. Analysis of program tracking and billing databases revealed that most programs had similar designs and some achieved savings from similar measure mixes. Some past evaluation results were similar in each state, and some were different.

We identified several demographic differences between RI and MA that could cause differences in program savings. These differences can be adjusted for in sampling and post-weighting, and they are listed for completeness. Additionally, these differences are for the entire state populations rather than specifically for program participants, and we do not know how representative they are of program participants.

In many cases, past evaluation approaches for the residential programs relied on billing analyses, for which a pooled sample provides little reduction of evaluation effort or cost. The past approach, and DNV GL's recommendations for piggybacking approaches for different residential programs, are listed in Table 6-2 along with the supporting key reasons.

An overarching recommendation that is primarily applicable to the residential studies reviewed in our meta-analysis is that evaluators should always report precisions or variance statistics (standard error or standard deviation) for final evaluation metrics such as realization rates. Not only do these statistics help place the findings for that study in better context, they facilitate cross-study comparisons in the future.

Table 6-2. Recommended Approaches - Residential

Program	Past Approach	Recommended Approach	Key Reasons
Lighting	**Approach 4** – Independent Samples	**Approach 4** – Independent Samples **or** **Approach 2** – Shared Algorithm (with adjustments)	Similar programs Large program Possibly different lighting markets
Behavioral Programs	**Approach 5** – Independent Studies	**Approach 4** – Independent Samples or **Approach 5** – Independent Studies	Similar programs Billing analysis utilizes independent sample
EnergyWise Single Family	**Approach 5** – Independent Studies	**Approach 4** – Independent Samples **or** **Approach 5** – Independent Studies **or** **Approach 3** – Pooled Sample (if no billing analysis and next evaluation shows similar results)	Similar programs Billing analysis utilizes independent sample Differences in previous evaluation results Flagship residential program for RI
Residential Cooling & Heating	**Approach 4** – Independent Samples	Insufficient evidence to make strong recommendation	Insufficient evidence Small program Minor differences in past evaluation results

Program	Past Approach	Recommended Approach	Key Reasons
Consumer Products	**Approach 1 – Direct** Proxy and **Approach 2** – Shared Algorithm	Appliance Recycling: **Approach 2** – Shared Algorithm **or** **Approach 3** – Pooled Sample (if field data collection used) Other measures: **Approach 1** - Direct Proxy	Similar programs Small program Minor differences in past evaluation results
Income Eligible Single Family	**Approach 5** – Independent Studies	**Approach 4** – Independent Samples or **Approach 5** – Independent Studies for next study; then **Approaches 1, 2, or 3** if next study has similar results for RI and MA	Billing analysis utilizes independent sample Past evaluation results differ but have long time gap
EnergyWise Multi-family	**Approach 4** – Independent Samples	**Approach 4** – Independent Samples or **Approach 2** – Shared Algorithm (if not using billing analysis)	Similar programs Billing analysis utilizes independent sample Past evaluation results differ Small program
New Construction, Code Compliance, and Building Characteristics	**Approach 4** – Independent Samples	**Approach 4** – Independent Samples or **Approach 5** – Independent Studies	Code compliance should be state-specific Code differences
Demand Response Programs	**Approach 4** – Independent Samples	**Approach 4** – Independent Samples **or** **Approach 3** – Pooled Samples if low participation size or constrained data	Similar programs Billing analysis used previously Data might be difficult to obtain

7 APPENDICES

7.1 Demographic Comparisons – Details

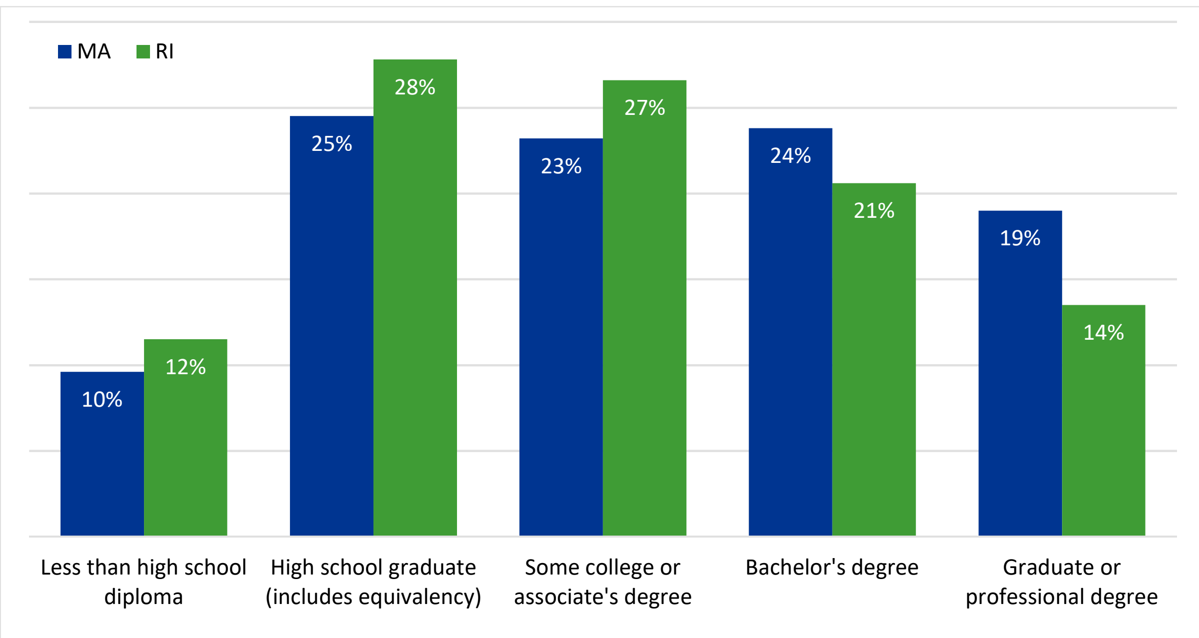
MA is much larger and has higher incomes than RI (Table 7-1) MA has almost seven times as many people, household incomes are approximately 25% higher, and individual income is approximately 20% higher.

Table 7-1. Population and Income

Statistic	RI	MA
Total population	1,056,426	6,811,779
Median household income (dollars)	60,596	75,297
Individuals – Median per capita income (dollars)	33,008	39,771

The population of MA has attained higher levels of education, on average, than RI (Figure 7-1).

Figure 7-1. Educational Attainment (population 25 years and older)



MA has approximately seven times as many occupied homes as RI. Homes in MA are more likely to be owner-occupied than in RI. Family sizes are slightly larger in MA and homes are slightly more likely to have a child present. There were only minimal differences in the percent of homes with a person aged 65 or older present (Table 7-2).

Table 7-2. Home Occupancy

Statistic	RI	MA
Occupied households	408,239	2,579,398
Owner occupied households	58%	62%
Renter-occupied households	42%	38%
Average household size – owner occupied	2.66	2.71
Average household size – renter occupied	2.24	2.26
Homes with children present	47%	51%
Householder 65 years or older	24%	23%

Less than 1% of homes in either state have a home business present. The rate is slightly higher in MA (0.65%) than RI (0.57%).

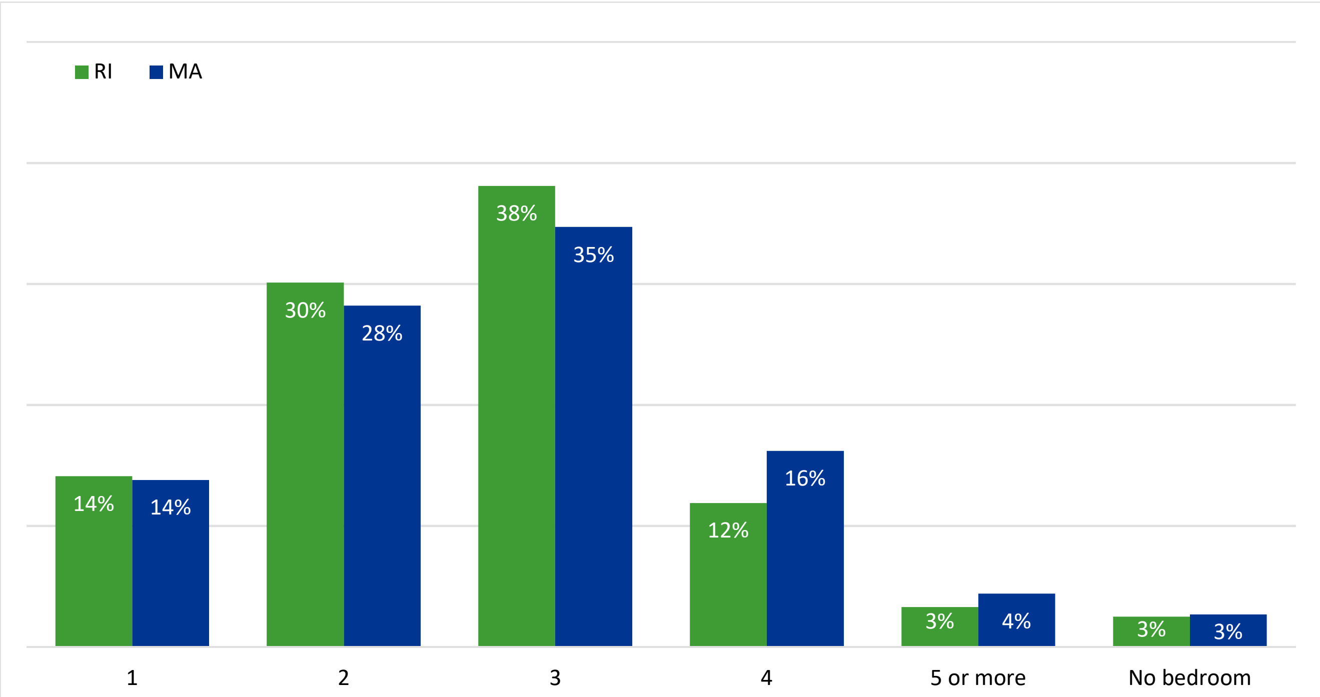
RI homes are more likely to be single-unit, detached, or in duplex or fourplex structures than in MA. In contrast, MA has a greater concentration of buildings with 10 or more units (Table 7-3).

Table 7-3. Units in Structure

Unit type by units in structure	RI	MA
Single unit, detached	55%	52%
2 to 4 units	24%	21%
10 or more units	12%	15%
5 to 9 units	5%	6%
Single unit, attached	3%	5%
Mobile home, boat, RV	1%	1%

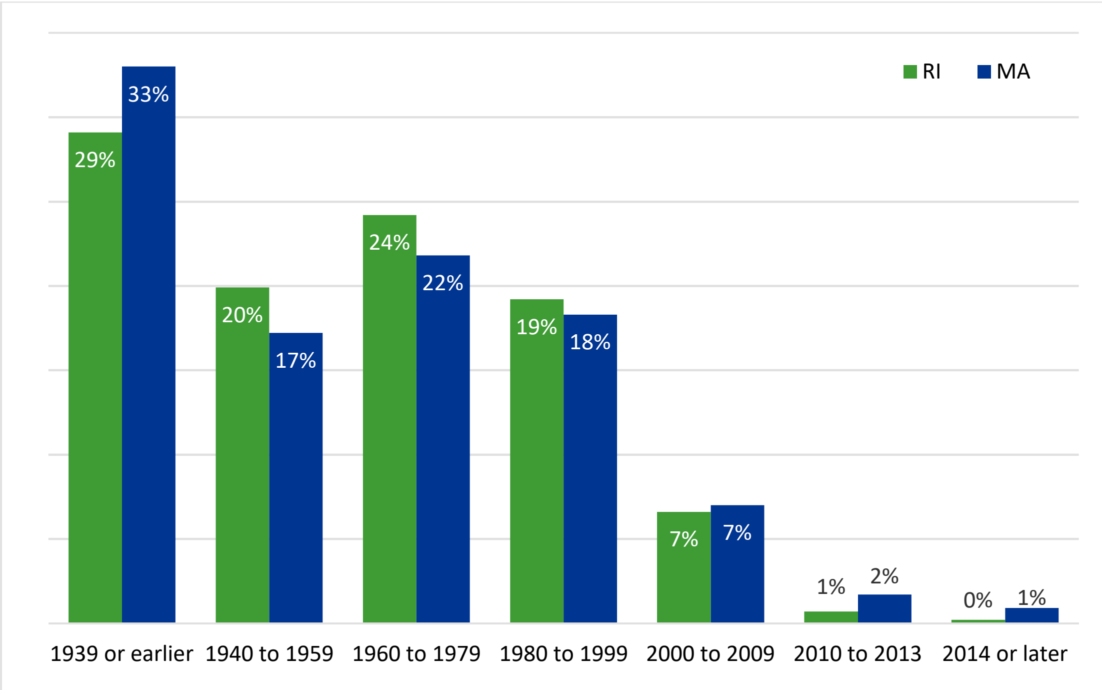
RI homes are more likely to have 2 or 3 bedrooms while MA homes are more likely to have 4 or 5 bedrooms (Figure 7-2). This suggests that MA homes are larger, on average, than RI homes.

Figure 7-2. Number of Bedrooms (occupied units)



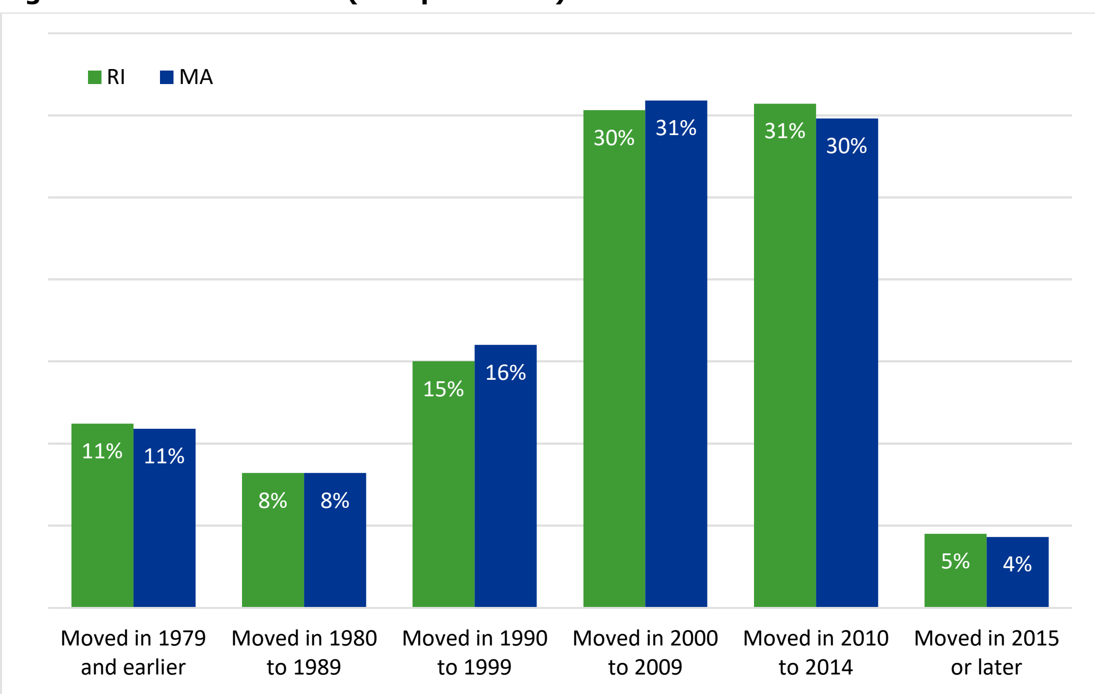
Homes in RI are more likely to be built in the latter 20th century than those in MA. MA homes are more likely to be older (built before 1940) or much younger (built since 2010; Figure 7-3).

Figure 7-3. Year Structure Built (occupied units)



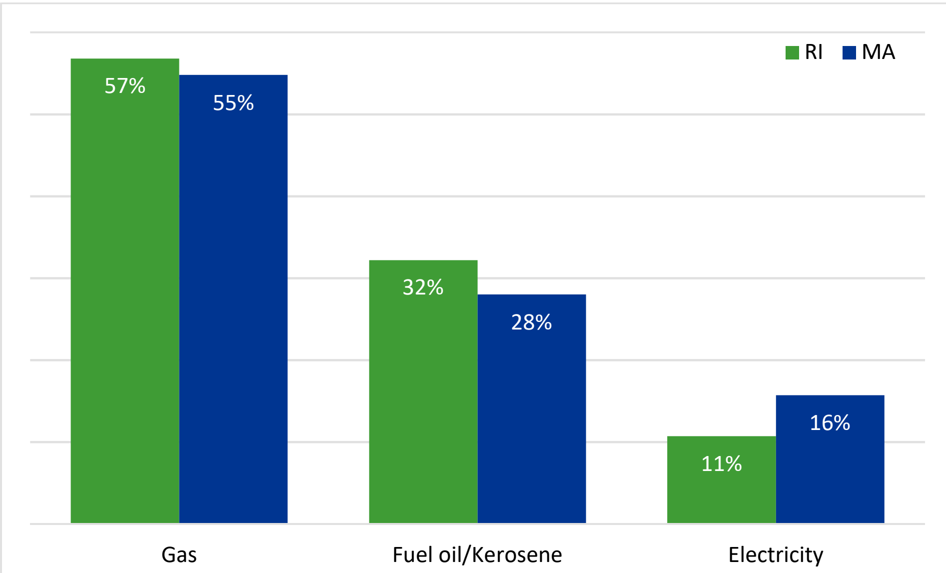
Home tenures are almost exactly the same in both states (Figure 7-4).

Figure 7-4. Home Tenure (occupied units)



RI homes are less likely to be heated via electricity and more likely to be heated with fuel oil or kerosene (Figure 7-5).

Figure 7-5. Home Heating Fuel (occupied units)



7.2 Previous Studies Compared in Meta-analysis

Table 7-4. Studies Reviewed in Meta-analysis

Study Year	Study Name	States Covered
2011	Impact Evaluation of 2011 RI Prescriptive Retrofit Lighting Installations	RI
2011	Impact Evaluation of 2011 RI Custom Lighting Installations	MA+RI
2012	Low-Income Single-Family Program Impact Evaluation	MA
2013	Impact Evaluations of 2011-2012 Prescriptive VSDs	MA
2014	Impact Evaluation of National Grid Rhode Island Commercial & Industrial Upstream Lighting Program	MA+RI
2014	Impact Evaluation of National Grid Rhode Island's Custom Refrigeration, Motor and Other Installations	MA+RI
2014	Impact Evaluation of National Grid Rhode Island C&I Prescriptive Gas Pre-Rinse Spray Valve Measure	MA+RI
2014	Northeast Residential Lighting Hours-of-Use Study FINAL	MA, CT, NY, RI
2014	2013 Commercial and Industrial Programs Free-Ridership and Spillover Study	RI
2014	Northeast Residential Lighting Hours-of-Use Study FINAL	MA, CT, NY, RI
2014	RI Behavioral Program and Pilots Impact Evaluation	RI
2014	Summary of the MA Behavioral Program Impact Evaluations	MA
2014	Impact Evaluation of the Income Eligible Services Single Family Program	RI
2015	RI Small Business Energy Efficiency Program Prescriptive Lighting Study	RI
2015	RI C&I Natural Gas Free Ridership and Spillover Study	RI
2015	2015-2016 MA Single-Family Code Compliance/Baseline Study: Volume 1 – FINAL	MA
2015	2015-2016 MA Single-Family Code Compliance/Baseline Study: Volume 2 – FINAL	MA
2015	2015-2016 MA Single-Family Code Compliance/Baseline Study: Volume 3 – FINAL	MA
2015	2015-2016 MA Single-Family Code Compliance/Baseline Study: Volume 4 – FINAL	MA
2015	2015-2016 MA Single-Family Code Compliance/Baseline Study: Volume 5 – FINAL	MA
2015	Retrofit Lighting Controls Measures Summary of Findings	MA
2015	High Efficiency Heating Equipment Impact Evaluation	MA
2015	Lighting Interactive Effects Study Preliminary Results - Draft	MA
2015	Ductless Mini-Split Heat Pump (DMSHP) Final Heating Season Results	MA+RI
2016	Impact Evaluation of 2014 RI Prescriptive Compressed Air Installations	MA+RI
2016	Impact Evaluation of 2012 National Grid-RI Prescriptive Chiller Program	MA+RI

Study Year	Study Name	States Covered
2016	Impact Evaluation of 2014 Custom Gas Installations in RI	MA+RI
2016	Large Commercial and Industrial On-Bill Repayment Program Evaluation	RI
2016	RI Commercial Energy Code Compliance Study	RI
2016	Multifamily Impact Evaluation	RI
2016	2013 Multifamily Program Gas and Electric Impact Study	MA
2016	ENERGYWISE Impact Evaluation of 2014 EnergyWise Single Family Program	RI
2016	Ductless Mini-Split Heat Pump (DMSHP) Cooling Season Results	MA+RI
2016	Low-Income Single-Family Health- and Safety-Related Non-Energy Impacts (NEIs) Study	MA
2016	Ductless Mini-Split Heat Pump Impact Evaluation	MA+RI
2017	RI 2013-2014 Custom Design Approach	MA+RI
2017	Gas Boiler Market Characterization Study Phase II - Final Report	Multiple
2017	Prescriptive Commercial and Industrial Programable Thermostat Phase 2 Study	MA
2017	Steam Trap Evaluation Phase 2	MA
2017	Final Report on Energy Impacts of Commercial Building Code Compliance in RI	RI only
2017	Impact Evaluation of 2014 Custom HVAC Installations	MA+RI
2017	Impact Evaluation of PY2015 MA Commercial and Industrial Upstream Lighting Initiative	MA
2017	2014 RI Custom Process Impact Evaluation	MA+RI
2017	Multi-Family Program Impact and Net-to-Gross Evaluation (RES 44)	MA
2017	Home Energy Assessment LED Net-to-Gross Consensus	MA
2017	RLPNC 16-7: 2016-17 Lighting Market Assessment Consumer Survey and On-site Saturation Study	MA
2017	2017 Saturation and Characterization Results	MA
2017	2017 MA Single-Family New Construction Mini-Baseline/Compliance Study	MA
2017	RI Statewide Behavioral Evaluation: Savings Persistence Literature Review	RI
2017	MA Cross Cutting Evaluation	MA
2017	Energy Efficiency Program Customer Participation Study	RI
2017	Residential Customer Profile and Participation Study	MA
2017	RI 2017 Code vs. UDRH Study	RI
2017	RI Code Compliance Enhancement Initiative Attribution and Savings Study	RI
2017	MA TXC47 Non-Residential Code Compliance Support Initiative Attribution and Net Savings Assessment	MA

Study Year	Study Name	States Covered
2017	Residential New Construction and CCSI Attribution Assessment	MA
2017	2017 Seasonal Savings Evaluation (Thermostats)	MA+RI
2017	2017 Residential Wi-Fi Thermostat DR Evaluation	MA+RI
2017	Final 2017 UDRH Inputs for the RI Residential New Construction Program	RI
2018	RI 2016 Custom Elec	MA+RI
2018	RI 2016 Custom Gas	MA+RI
2018	Impact Evaluation of PY2016 RI Commercial & Industrial Small Business Initiative	MA+RI
2018	RI Residential lighting market assessment and NTG Estimation	RI
2018	LED Net-to-Gross Consensus Panel Report	MA
2018	Residential Appliance Saturation Survey	RI
2018	RI EnergyWise/HVAC Heat Loan Assessment	RI
2018	HEAT Loan Assessment	MA
2018	RI Baseline Study of Single-Family Residential New Construction	RI
2018	Impact Evaluation of PY2015 RI Commercial and Industrial Upstream Lighting Initiative	MA+RI
2018	Home Energy Services Impact Evaluation (Res 34) August 2018	MA
2019	Rhode Island 2017 Lighting Sales Data Analysis	MA+RI
2019	2018 Rhode Island Shelf Stocking Study	MA+RI
2019	Appliance Recycling Impact Factor Update	RI
2019	MA19R01-E Appliance Recycling Report	MA
2019	Impact Evaluation of PY2016 Custom Gas Installations in RI	MA+RI

8 PARTICIPANT DEFINITIONS FOR COMMERCIAL PROGRAMS

8.1 Prescriptive Lighting

RI definition of participant:

- (2014) Projects from DNV_RI PY2014 DSM_Eval_(015)_Free_Ridership-Spillover_LCI-SBS_6-4-15.xls
  - Where Program not equal to "SBS"
  - and sub_program = "Lighting"
- (2015) Projects from RI PY2015-PROD DSM EVAL_(015)_Free_Ridership-Spillover_LCI-SBS 4-27-16.xls
  - Where Program not equal to "SBS"
  - and sub_program = "Lighting"
- (2016,2017) Projects from LCI_Electric_Projects.xls
  - Where installation_type = "Prescriptive"
  - and end use = "Prescriptive Lighting"
  - and detailed_measure_char contains "LED" or "Lighting"
  - or measure_installed variables contain "LED" or "Lighting"

MA definition of participant:

- (2014, 2015, 2016, 2017) Projects from standardized MA database
  - Where pa_dnv="NGRID"
  - and tracking_type="E"
  - and project_track_dnv="Prescriptive"
  - and project_class_dnv in ("Custom" "New Construction" "Retrofit")
  - and end_use_impacted_dnv in ("LIGHTING")
  - and core_initiative_dnv not in ("C&I Multifamily Retrofit" "C&I Small Business")

8.2 Upstream Lighting

MA all years:

- track_2014, track_2015, track_2016, track_2017
  - if tracking_type= "E"
  - and project_track_dnv= "Upstream"
  - and end_use_impacted_dnv= "UPSTREAM LIGHTING".

RI definition of participant:

- (2015): Projects from PY 2015 RI LCI Upstream lighting.xlsx
- (2016, 2017): Projects from Rebate_Projects.xlsx
  - Where Program_initiative_name = "LCI Upstream Lighting"

8.3 Custom Electric Non-lighting

RI definition of participant:

- (2014) Projects from DNV_RI PY2014 DSM_Eval_(015)_Free_Ridership-Spillover_LCI-SBS_6-4-15.xls
  - Where program="D2" or "EI"
  - and sub_program="CUSTA"
  - Where and Installed_Measure_Report_Group does not contain "LIGHT","LED", "CDA", "Comprehensive Design", or "CHP"
- (2015) Projects from RI PY2015-PROD DSM_Eval_(015)_Free_Ridership-Spillover_LCI-SBS 4-27-16.xls
  - Where program="D2" or "EI"
  - and sub_program="CUSTA"
  - and Installed_Measure_Report_Group does not contain "CDA", "LIGHT","LED", or "CHP"
- (2016,2017) Projects from LCI_Electric_Projects.xls
  - Where installation_type = "Custom"
  - and end_use ≠ "Lighting"
  - and detailed_measure_char does not contain "LED", "Lighting", "CDA", "Comprehensive Design", or "CHP"
  - and measure_installed variables did not contain "LED", "Lighting", "CDA", "Comprehensive Design", or "CHP"

MA definition of participant:

- (2014) Projects from MA PY2014 DSM_Eval_(015)_Free_Ridership-Spillover_LCI-SBS_6-9-15_v2.xls
  - Where program="EI" or "D2" and sub_program="CUSTA"
  - and Installed_Measure_Report_Group did not contain "LIGHT","LED", or "CHP"
- (2015, 2016, 2017) Projects from standardized MA database
  - Where pa_dnv="NGRID"
  - and tracking_type="E"
  - and project_track_dnv="Custom"
  - and project_class_dnv in ("Custom" "New Construction" "Retrofit")
  - and end_use_impacted_dnv in ("BUILDING SHELL" "COMPRESSED AIR" "FOOD SERVICE" "HOT WATER" "HVAC" "MOTORS / DRIVES" "OTHER" "PROCESS" "REFRIGERATION")
  - and core_initiative_dnv not in ("C&I Multifamily Retrofit" "C&I Small Business")

8.4 Custom Electric Lighting

RI definition of participant:

- (2014) Projects DNV_RI PY2014 DSM_Eval_(015)_Free_Ridership-Spillover_LCI-SBS_6-4-15.xls
  - Where sub_program = "CUSTA"
  - and installed_measure_report_group contains "LIGHT" or "LED"

- (2015) Projects from RI PY2015-PROD DSM_Eval_(015)_Freed_Ridership-Spillover_LCI-SBS 4-27-16.xlsx
  - Where sub_program = ”CUSTA”
  - and installed_measure_report_group contains “LIGHT” or “LED”
- (2016,2017) Projects from LCI_Electric_Projects.xls
  - Where installation_type = ”Custom”
  - and end_use = ”Lighting”
  - and detailed_measure_char contains “LED” or “Lighting”
  - or measure_installed variables contain “LED” or “Lighting”

MA definition of participant:

- (2014, 2015, 2016, 2017) Projects from standardized MA database
  - Where pa_dnv="NGRID"
  - and tracking_type="E"
  - and project_track_dnv="Custom"
  - and project_class_dnv in ("Custom" "New Construction" "Retrofit")
  - and end_use_impacted_dnv in ("LIGHTING")
  - and core_initiative_dnv not in ("C&I Multifamily Retrofit" "C&I Small Business")

8.5 Small Business Electric

RI definition of participant:

- (2014) Projects from DNV_RI PY2014 DSM_Eval_(015)_Free_Ridership-Spillover_LCI-SBS_6-4-15.xls
  - Where Program=”SBS”
- (2015) Projects from RI PY2015-PROD DSM_Eval_(015)_Free_Ridership-Spillover_LCI-SBS 4-27-16.xls
  - Where Program=”SBS”
- (2016,2017) Projects from SBS_Projects.xls
  - Where project_fuel_type= ”Electric”

MA definition of participant:

- (2014) Projects from standardized MA database
  - Where pa_dnv="NGRID"
  - and sector="C&I"
  - and tracking_type="E"
  - and project_class_detailed_dnv="Small Retrofit"
- (2015) Projects from standardized MA database
  - Where pa_dnv="NGRID"
  - and sector="C&I"
  - and tracking_type="E"

- and program_verbose_dnv contains ("SBS")

- (2016) Projects from standardized MA database
  - Where pa_dnv="NGRID"
  - and sector="C&I"
  - and tracking_type="E"
  - and program_verbose_dnv = ("Small Business Services")

- (2017) Projects from standardized MA database
  - Where pa_dnv="NGRID"
  - and sector="C&I"
  - and tracking_type="E"
  - and core_initiative_dnv = ("C&I Small Business")

8.6 Prescriptive Non-lighting

RI definition of participant:

- RI 2014: DNV_RI PY2014 DSM_Eval_(015)_Free_Ridership-Spillover_LCI-SBS_6-4-15.xls,
  - where Program ne "SBS" and
  - Sub_Program not equal ("Lighting" "CUSTA")

- RI 2015: RI PY2015-PROD DSM_Eval_(015)_Free_Ridership-Spillover_LCI-SBS 4-27-16.xls
  - where Program ne "SBS" and
  - Sub_Program not equal ("Lighting" "CUSTA")

- RI 2016-2017: LCI_Electric_Projects.xls
  - where installation_type= "Prescriptive"
  - and end_use does not equal "Lighting"

MA definition of participant:

- MA 2014: track_2014
  - if tracking_type= "E"
  - and project_track_dnv= "Prescriptive"
  - and project_class_detailed_dnv ne "Small Retrofit"
  - and end_use_impacted_dnv ne "LIGHTING"

- MA 2015: track_2015
  - if tracking_type= "E"
  - and project_track_dnv= "Prescriptive"
  - and program_verbose_dnv does not contain "SBS"
  - and end_use_impacted_dnv ne "LIGHTING"

- MA 2016: track_2016
  - if tracking_type= "E"

- and project_track_dnv= "Prescriptive"
- and program_verbose_dnv not in ("Small Business Services" "Energy WiseC&I Multifamily Retrofit")
- and end_use_impacted_dnv ne "LIGHTING"

- MA 2017: track_2017
  - if tracking_type= "E" and project_track_dnv= "Prescriptive"
  - and core_initiative_dnv not in ("C&I Small Business" "C&I Multifamily Retrofit")
  - and end_use_impacted_dnv ne "LIGHTING"

8.7 Custom Gas

RI definition of participant:

- RI 2014: DNV_RI PY2014 DSM_EVAL_(025-G)_Gas_Participation_6-4-15.xls
  - where input source="Gas Custom Application"
- RI 2015: RI PY2015-PROD DSM_EVAL_(025-G)_Gas_Participation 5-19-16.xls
  - where input_source= "Gas Custom Application"
- RI 2016-2017: Gas_Custom_Projects.xls
  - all observations

MA definition of participant:

- MA 2014: track_2014
  - if tracking_type= "G"
  - and project_track_dnv= "Custom"
  - and project_class_detailed_dnv not equal "Small Retrofit"
- MA 2015: track_2015
  - if tracking_type= "G"
  - and project_track_dnv= "Custom"
  - and program_verbose_dnv does not contain "SBS"
- MA 2016: track_2016
  - tracking_type= "G"
  - and project_track_dnv= "Custom"
  - and program_verbose_dnv not equal ("Small Business Services","Energy WiseC&I Multifamily Retrofit")
- MA 2017: track_2017
  - if tracking_type= "G"
  - and project_track_dnv= "Custom"
  - and core_initiative_dnv ne ("C&I Small Business","C&I Multifamily Retrofit")

8.8 Prescriptive Gas

RI definition of participant:

- 2016, 2017: Rebate_projects.xls
  - where Installation_type="Prescriptive" and project_fuel_type="Gas"

MA definition of participant:

- 2016: track_2016
  - Tracking_type="G"
  - And pa_dnv="NGRID"
  - And project_track_dnv="Prescriptive"
  - And project_class_dnv="Retrofit"
  - And direct_install_flag_dnv not equal "Direct Install"
- 2017: track_2017
  - Tracking_type="G"
  - And pa_dnv="NGRID"
  - And project_track_dnv="Prescriptive"
  - And project_class_dnv="Retrofit"
  - And direct_install_flag_dnv not equal "Direct Install"
  - And dnv_core_initiative not equal"C&I Small Business"