

**R1706 Residential Appliance Saturation Survey**

**Appendix G, Database User Guide**

UPDATED

October 1, 2020

SUBMITTED TO:  
Connecticut Energy Efficiency Board Lisa Skumatz, Ralph Prahl, and Bob Wirtshafter, EEB Evaluation Administrators

SUBMITTED BY:  
NMR Group, Inc.

[](http://www.nmrgroupinc.com)

Table of Contents

[Appendix G Database User Guide 3](#_Toc21076734)

[Table of Contents 3](#_Toc21076735)

[Data Dictionary Tab 3](#_Toc21076736)

[Raw Data Tab 4](#_Toc21076737)

[Adjustment Factors Tab 4](#_Toc21076738)

[Summary Statistics Tab 5](#_Toc21076739)

[SF\_Efficiency Levels Tab 7](#_Toc21076740)

1. Database User Guide

In June 2019, NMR delivered the RASS database (*CT R1706 RASS Database 13June2019.xls*) to the Connecticut Energy Efficiency Board as part of the *R1706 Residential Appliance Saturation & R1616/R1708 Residential Lighting Impact Saturation Studies*. This appendix to the report – also delivered in June 2019 – serves as a user guide for that database. In mid-2020, NMR enhanced the database, as part of R2023 RASS Database Enhancement, and accordingly updated this user guide.

The database consists of seven worksheets or tabs:

* **Table of Contents.**  List of the tabs and what each contains
* **Data Dictionary.** Variable and value labels for the raw data
* **Raw Data.** Unweighted case-level raw online survey and on-site verification data with billing data
* **Adjustment Factor.** Web-survey statistics compared to verified on-site statistics
* **Summary Statistics.** Penetration, quantity, and average units per household for key end-uses (i.e., measures, devices, equipment) – cross-tabulated by key respondent characteristics
* **SF\_Efficiency Levels.** Average efficiency levels for common heating, cooling, water heating equipment. and insulation for single-family homes
* **MF\_Efficiency Levels.** Similar to *SF-Efficiency Levels* tab, but for multifamily homes

This user guide will help stakeholders adeptly leverage the database for their own purposes by walking users through the database tab-by-tab. While reading this, users should have the database open so that they can follow along.

Table of Contents

The *Table of Contents* tab contains a listing of the tabs and what each contains.

Data Dictionary Tab

The *Data Dictionary* identifies and defines the full set of variables included on the *Raw Data* tab. For example, the tab identifies how the variable TENURE refers to self-reported tenure (owner versus renter). As Table 1 below shows, 1 = Owner and 0 = Renter.

Table 1: Data Dictionary Tab – Example Variables

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Variable Name** | **Variable Description** | **Source** | **Type** | **Detailed Description** |
| TENURE | Tenure | Web survey | Yes/No | 1 = Owner; 0 = Renter |
| A1\_1S | Count - Clothes washer (On-Site) | On-site verification visits | Quantity | Number of clothes washers (collected on-site) |

Raw Data Tab

The *Raw Data* tab houses all survey responses and on-site data at the case level. As described in Appendix C, analysts coded open-ended responses and back-coded[[1]](#footnote-1) close-ended responses as needed. In addition to showing survey and on-site data, the tab includes two elements for future analyses:

* It devotes variables to the weights associated with each respondent but does not apply those weights (other tabs contain weighted values and the column headers for those values identify where it is weighted or not).
* Household-level electric consumption from billing data provided by the Companies – each month, dating back to December 2015,[[2]](#footnote-2) is represented by a variable with the household’s total electric usage (kWh) for that month. Appendix C details the billing data cleaning process; for example, consumption variables aggregate daily usage into calendar monthly kWh.

The tab clusters multiple variables associated with specific pieces of equipment using numbered pre-fixes. For example, a clothes washer observed on site was associated with the prefix *CW1\_* and any characteristics such as type (*CW1\_TYPE*), ENERGY STAR status (*CW1\_ENERGYSTAR*), and efficiency (*CW1\_MEF*) included that prefix. A second clothes washer would have variables that would follow the same pattern, using a prefix of *CW2.* Note, the tab also shows *don’t know*[[3]](#footnote-3) responses while statistical analyses exclude those responses.

Adjustment Factors Tab

The *Adjustment Factor* tab compares penetration and average number of units per home by end use between web-survey responses and on-site verification visit observations. The comparison results in adjustment factors based on the ratio between self-reported and observed values. The tab indicates to what extent the results vary between the two methods and if they statistically vary enough to apply adjustment factors for analyses. It includes several components:

* **End-use information.** Columns A through D provide basic information about the end-use, including an end-use ID to link end-uses to the *Summary Statistics* tab if desired. This tab explores only end uses asked about online *and* investigated on site.
* **Penetration summary.** Columns E includes the number of web-survey respondents (*don’t know* and outlier/erroneous responses excluded), then column F shows the unadjusted weighted penetration from the web-survey results, column G shows the adjusted weighted penetration, and column H shows the adjustment factor used to revise the web-survey results.[[4]](#footnote-4) As points of comparison, columns I and J show the adjusted weighted penetration from the Rhode Island 2018 and Massachusetts 2017 studies.[[5]](#footnote-5)
* **Average units per home summary.** Similarly, the average units per home summary in columns K through O follow the same pattern. They compare original weighted web-survey results, adjusted weighted results, adjustment factors, and Rhode Island and Massachusetts results. Note that average units per home was relevant to only a few end uses: appliances, thermostats, and advanced power strips.
* **On-site sample results.** Columns P through V show results among the on-site homes only, including number of visits, unweighted count of homes with the end use, unadjusted weighted results based on their web-survey responses and the weighted results from on-site observations.
* **Initial adjustment factors.** Columns W and X use the ratio of column T versus S and V versus U, respectively, to estimate initial adjustment factors (that may or may not have been applied).
* **Significance testing.** Columns Z through AM present the significance testing which determines if initial adjustment factors should be used. The analysis applies adjustment factors if on-site and web results statistically significantly differ at the 90% confidence level. If either columns AI *or* AM note statistically significant differences, then both adjustment factors in columns W and X are applied (as reflected in columns H and M).

Summary Statistics Tab

If overall adjusted weighted penetration for a particular end-use was at least 3%, then the *Summary Statistics* tab analyzes it in greater detail, with up to 35 cross-tabulations (rows) per end use. These bullets explain the structure of this tab:

* **Sub-group.** Columns A through F identify the sub-group being examined. Spreadsheet users can filter results as desired based on drop-down choices (i.e., Excel’s *filter* function). For example, a user could select single-family, low-income, and owner-occupied households and look at statistics for any desired end use for that demographic. The discussion below explains the sub-groups.
* **End-use information.** Columns G through I provide basic information about the end-use, including an end-use ID to link end uses to the *Adjustment Factor* tab if desired. This tab explores only end uses asked about online *and* searched for on site.
* **Sample size.** Column J lists the number of customers providing a valid response in the web survey related to that measure. It flags small sample sizes (n = < 20) by shading cells in grey.
* **Penetration rates.** Columns K shows the weighted adjusted penetration. M through P show the unweighted number of homes reporting the end use, unadjusted penetration rates with web-survey weights applied, and absolute precision and standard error associated with the unadjusted weighted penetration. These values relate to the calculation of the final reported value (the weighted adjusted penetration rates shown in column K).
* **Quantity per home.** Columns Q through T present the weighted proportions of households with none, only one, two, or three or more units of the end use.
* **Average units per home.** Column L shows the weighted adjusted average units per home. Columns U through W present the weighted unadjusted average units per home with absolute precision and standard error. These values relate to the calculation of the final reported value (the weighted adjusted average units per home shown in column L). These columns denote if the number of units per home was not measured with “NA.”

As part of the database enhancement, NMR added the ENERGY STAR and energy-efficiency summary statistics of the on-site homes:

* **On-site sample sizes.** Column X shows the number of on-site homes included in each sub-group. Column Y indicates the number of on-site homes in that sub-group that contain the measure identified in Column I; Column Z identifies the quantity of those measures that were found at these homes. (Some homes have more than one unit of a given measure; e.g., two refrigerators in a given home.)
* **ENERGY STAR status.** Column AA shows the number of measures found on-site that have a verifiable ENERGY STAR status. If no ENERGY STAR label was visible, technicians used the measure’s model number or serial number to determine whether the measure was ENERGY STAR qualified at the time of manufacture. The count in this column excludes any measures with an unknown ENERGY STAR status (e.g., if the model number was faded or not found in any known product databases). The next columns illustrate ENERGY STAR saturation overall and by age:
  + **Overall saturation.** Column AB shows the unweighted count of all measures verified as ENERGY STAR qualified at the time of their manufacture. For example, if a unit was manufactured in 2010 and it earned the ENERGY STAR certification under the standards at the time, the unit is counted here as being ENERGY STAR-certified. If that same model were manufactured today, it may no longer meet ENERGY STAR specifications; the column does not account for that caveat. Column AC shows the overall ENERGY STAR saturation – the count in Column AB as a weighted percentage of Column AA.
  + **Saturation by age.** Columns AD through AM distribute the overall ENERGY STAR saturation across age brackets or *bins* associated with year of manufacture. Table 2 outlines the bins with an example of refrigerators. Overall, 37% of refrigerators were ENERGY STAR-certified: 1% of all refrigerators were ENERGY STAR-certified and manufactured in 2003 or earlier, 4% of all refrigerators were ENERGY STAR-certified and manufactured between 2004 and 2008, 9% of all refrigerators were ENERGY STAR-certified and manufactured between 2009 and 2013, and so on. If a unit had a visible ENERGY STAR label but the technician was unable to ascertain the year of manufacture, it appears in Columns AL and AM (*age unknown*).

Table 2: ENERGY STAR Saturation by Year of Manufacture

|  |  |  |
| --- | --- | --- |
| **Year of Manufacture** | **Bin** | **Saturation**  **(refrigerator example, n=227)1** |
| 2003 and older | 1 | 1% |
| 2004 to 2008 | 2 | 4% |
| 2009 to 2013 | 3 | 10% |
| 2014 to 2018 | 4 | 21% |
| Age unknown | n/a | 2% |
| **Overall** | **n/a** | **37%** |
| 1 Technicians found 273 refrigerators on site, but only 228 had a discernable ENERGY STAR status, so the denominator was 228. | | |

* **Efficiency levels.** Columns AN through AV include measure efficiency levels overall (Columns AN through AP) and split by residential (Columns AQ through AS) and commercial (Columns AT through AV). The *Efficiency Metric (n)* columns indicate the quantity of observed measures in the sub-group that have discernable efficiency data. Missing efficiency data indicates that the unit was unreachable by technicians or the model number was either faded or yielded no results in a database search. *Units of Measurement* indicate what is being analyzed (e.g., HPSF, SEER). *Average Efficiency* shows the weighted average efficiency rating for measures in each sub-group.

As mentioned above, the cross-tabulations devote up to 35 rows to each end-use:

* First, it cross-tabulates dwelling type (single-family, multifamily, all) by income (low-income, non-low-income, all) by tenure (own, rent, all), yielding 27 rows per end-use.
* Next, it cross-tabulates by program participation (participant versus non-participant), primary heating type (fuel oil, natural gas, electric, other), and company (Eversource versus United Illuminating), yielding another 9 rows per end-use.

Table 2 presents the groups and flags small sample sizes.

Table 2: Summary Statistics Tab – Cross-Tabulations

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Demographic** | | **Dwelling type** | |  |
| **Single-Family, 1-4 units** | **Multifamily, 5+ units** | **All** |
| Low-income | Owners |  | ⬩ |  |
| Renters |  |  |  |
| All tenures |  |  |  |
| Non-low-income | Owners |  |  |  |
| Renters |  |  |  |
| All tenures |  |  |  |
| All incomes | Owners |  |  |  |
| Renters |  |  |  |
| All tenures |  |  |  |
| **Limited Cross-Tabulations** | | | | |
| **Primary heating fuel** | |  |  |  |
| Natural gas | |  |  |  |
| Fuel oil | |  |  |  |
| Electricity | |  |  |  |
| Other | |  |  |  |
| **Participation** | |  |  |  |
| Participant | |  |  |  |
| Nonparticipant | |  |  |  |
| **Electric company** | |  |  |  |
| Eversource | |  |  |  |
| United Illuminating | |  |  |  |
|  Indicates that the database analyzes this group.  ⬩ Flags that the database analyzes this group, but sample sizes are small. | | | | |

Efficiency Levels Tabs

The *SF\_Efficiency Levels* tab shows the weighted average efficiency ratings found in single-family homes: AFUE ratings for furnaces and boilers (by fuel type); the average HSPF, SEER, EER, and COP ratings (as applicable) for common electric equipment, including central ducted air source heat pumps, mini-split heat pumps, ground-source heat pumps, central air conditioning, and room air conditioning; and the EF for water heaters. Columns L through P provide frequencies of predominant R-values for wall, ceiling, floor, and duct insulation.

The *MF\_Efficiency Levels* tab similarly shows the weighted average AFUE, HSPF, SEER, EER, COP, and EF ratings in multifamily buildings. Average efficiencies are broken out by residential and commercial equipment, where applicable.[[6]](#footnote-6)

Survey questions did not ask about efficiency levels, so results are on-site derived only.

1. Back coding refers to the process of reviewing a respondent’s answers and correcting errors that other responses indicate are erroneous or completing incomplete answers that other questions answer. For example, if a respondent lived in a building with five units and reported they had five dishwashers, the team assumed they had one dishwasher per housing unit and revised the quantity from five to one. [↑](#footnote-ref-1)
2. For some customers who were also panelists in the R154 study, billing data dates as far back as December 2014. [↑](#footnote-ref-2)
3. As shown in the data dictionary, *don’t know* is often denoted with the number 98 for quantity type questions. [↑](#footnote-ref-3)
4. If equal to 100%, no adjustment factor was applied. Adjustment factors were inapplicable if none of the on-site respondents reported the end-use *or* none were observed on site (noted as “NA”). [↑](#footnote-ref-4)
5. Source: *MA Statewide Final Saturation Results 2017-07-16.xls* and *NGrid RI2311 RASS database 20181020.xls*. [↑](#footnote-ref-5)
6. The Team used available data from multifamily site visits to organize the equipment into commercial or residential breakouts. This varied by equipment type and available data – available variables in the data led NMR to use the “Serves” variable as the main determinant of residential (serves one unit) or commercial (serves multiple units). [↑](#footnote-ref-6)