Measurement & Verification of Residential Refrigerator Energy Use

Final Report

2003 – 2004 Metering Study

July 29, 2004

Submitted by:

Michael Blasnik
on behalf of project team Proctor Engineering Group Limited, Michael Blasnik & Associates, and Conservation Services Group

Contents

Appendix C provided separately summarizes the data and findings for each refrigerator in the project.

Acknowledgments

This project would not have been possible without the cooperation and hard work of the auditors and program managers at the CSG, HDMC, and RISE Engineering who oversaw the deployment of the metering equipment and field data collection while juggling their other responsibilities to deliver the programs to the customers. Bryon Martinez at CSG went above and beyond the call of duty in ensuring the retrieval of the metering equipment for the entire project. The author would like to thank the Evaluation Working Group members for providing encouragement throughout the project, valuable suggestions during the design and planning, and insightful feedback on the reports. Of course, all errors and omissions are the sole responsibility of the author.

1. Executive Summary

The Refrigerator Measurement & Verification (M&V) project is a research project designed to better understand the field performance of refrigerators, assess the accuracy of refrigerator audit/diagnostic methods, and measure the energy and load impacts of refrigerator replacements more reliably. The project was sponsored by NSTAR Electric & Gas, National Grid local utilities Massachusetts Electric Company (MECO) and Narragansett Electric Company (NECO), and Northeast Utilities' local utility Western Massachusetts Electric Company (WMECO). This final report describes the project activities and the results of the data analysis.

Methodology

The research design included monitoring electricity usage and room temperatures for 160 existing and 30 new Energy Star rated replacement refrigerators for 2-4 week periods in customer homes from June 2003 through May 2004. The sample of refrigerators was drawn from program participants in five utility programs: MECO RCS, NSTAR RCS, NSTAR RHU, NECO EnergyWise, and WMECO RCS, operated by three different Program Installation Vendors (PIVs): CSG, Honeywell DMC, and RISE Engineering. The PIVs screened units and deployed metering equipment as part of their regular program work. Equipment was retrieved by CSG staff as part of the evaluation team.

Currently, DMC and CSG estimate the usage of customer refrigerators by looking up the make and model in a database of rated usage values derived from DOE-specified testing that is used to produce the Energy Guide labels that have appeared on refrigerators for the past 25 years (referred to as the label-rated usage). DMC and CSG each have their own algorithms to adjust the label-rated usage value to account for the age and other features of each refrigerator (referred to as the PIV rated usage). RISE estimates the usage of each refrigerator by metering the electric usage for 60 to 90 minutes and extrapolating to an annual value based on the length of metering and the room temperature. DMC and CSG also use short-term metering, ranging from 45 minutes to more than two hours, for the 10%-15% of the units they can't locate in the rated usage database.

We analyzed the metered data to develop estimates of the annual energy usage of each unit. We used these results to:

- assess the accuracy of existing refrigerator audit approaches employed by the PIVs;
- develop a simple approach for estimating the usage of refrigerators in the field based on adjusting the rated usage;
- develop program savings adjustment factors that estimate the proportion of PIV-estimated savings being achieved by the programs (i.e., savings realization rates); and,
- develop load shape estimates for existing and new refrigerators.

Findings

Key characteristics of the refrigerators metered are summarized in Table 1. Additional data can be found in Section 5.

Table 1. Refrigerator Characteristics and Energy Usage – by utility program

Characteristic	MECO RCS	NECO EW	NSTAR RCS	NSTAR RHU	WMECO RCS	All Existing	New Units
# Units Metered*	45	15	41	27	30	158	30
Year Built (median)	1985	1982	1985	1984	1983	1984	2003
Volume	17.9	17.9	18.5	18.5	18.6	18.3	20.0
Label-Rated Usage*	1207	1255	1249	1226	1302	1244	484
PIV-Estimated Usage	1830	1969	1655	1703	1656	1743	N/A
Actual Detailed Metered Usage	1288	1640	1338	1391	1452	1383	425
% of Label-Rated	105%	135%	106%	113%	111%	**111%**	**88%**
% of PIV Estimated	70%	83%	81%	82%	88%	79%	N/A
Door Style:							
top freezer	89%	80%	71%	67%	80%	78%	60%
side-by-side	11%	20%	20%	26%	13%	17%	17%
bottom freezer	0%	0%	5%	4%	7%	3%	23%
single door	0%	0%	0%	4%	0%	1%	0%
Icemaker							
internal	4%	7%	15%	7%	10%	9%	27%
thru-the-door	7%	13%	10%	15%	7%	9%	17%
Located in Kitchen / Living Space	73%	100%	83%	59%	93%	80%	83%

* The number of units listed includes all units where metering was deployed. Two units did not have useable data, leaving 156 existing units for analysis. Label rated usage data were only found for 151 of these 156 units

The table shows that the typical existing refrigerator in the project was a 20 year old 18 cubic foot top freezer unit located in the kitchen. The label-rated usage averaged 1,244 kWh/yr and the PIV audits estimated that the usage was 1,743 kWh/yr. The detailed metering found that the true usage averaged 1,383 kWh/yr – equal to 111% of the label rated usage, but just 79% of the PIV audit estimate. The new Energy Star replacement units used an average of 88% of the label-rated usage.

Further analysis found that:

- existing refrigerators in basements used 9% less electricity than the label-rated usage on average, primarily due to cooler temperatures and smaller occupant loads in basements, while units in the main living space used about 16% more than the label rating on average;
- about 8% of the existing refrigerators had essentially constant usage indicating a potential malfunction causing the unit to run all the time. These units used an average of 158% of rated usage and tended to be units that were acquired used, had low power factor, and/or had large gaps around the door seals. These units provide a large savings potential and benefits to the programs.

We analyzed the relationship between the actual usage and the label-rated usage to develop a new audit approach. The analysis focused on estimating refrigerator usage as a percentage of label-rated usage based on auditable refrigerator and household characteristics. Modeling results varied depending on the specific model specification and estimation approach and whether the analysis included units with "flat" usage and units outside the living space. The analysis identified several factors with impacts that were generally statistically and practically significant across a range of models. Many factors had no apparent impacts. We used the results from the modeling to develop the new audit approach. The results of the analysis and the proposed new audit approach are summarized in Table 2. The estimated impact column shows the range of model coefficient point estimates across models and samples, not a statistical confidence interval. The temperature impact is from the detailed usage data modeling and was not included in the regression analysis.

Table 2. Factors Affecting Refrigerator Usage & Proposed New Audit Approach

Factor	Estimated Impact (% rated usage)	Proposed Audit Approach	Notes
# Occupants (primary unit only)	4.3%-5.5%	+5% /occupant	only for refrigerators used regularly by occupants. probably declines per occupant as # occupants grows
Anti-sweat Switch "On"	12% - 22%	+20%	Actual energy use from switch is about 15% of rated usage, but other factors correlate with switch existence
Icemaker: through-the-door	8%-17%	+15%	statistical significance marginal, but better predictor than side-by-side door style
Door Seal has noticeable gaps	13%-27%	+15%	related to occurrence of continuous running flat usage
Unit was bought used	18%-27%	+20%	also related to occurrence of continuous running flat usage
Average Room Temperature	2.65% /°F	+5% if avg. winter T in low 70s -5% if avg. winter T in low/mid 60s	Annual temperature effects can be roughly accounted for by adjusting usage up or down if avg. winter thermostat settings outside typical range 65°-70°F
Base Level Usage	81%-84%	85%	usage estimate if no occupants, no anti-sweat on, no TTD ice, OK door seals, unit bought new or came with house, typical winter thermostat settings

The proposed new audit approach involves simply adding up the relevant factors in the "Proposed Audit Approach" column and multiplying that sum by the rated usage.

We did not find significant effects from many factors, including:

- the refrigerator's age (all units were relatively old so any early degradation would have occurred);
- refrigerator door style (better accounted for through presence of through-the-door icemaker);
- presence of internal icemaker;
- occupant-reported schedules or frequency of cooking;
- auditor-estimated overall unit condition;
- unit location details such as clearances, recessed into cabinets, solar gain or outside wall; and,
- many other unit and site characteristics collected as part of the study.

The study included few units less than 10 years old, limiting our ability to find any age effect. In addition, almost all units had automatic defrost, preventing an analysis of defrost impacts. We found that our modeling, and the new audit approach, only works well for refrigerators located in the main living space. Refrigerators in basements, garages, and other semi-conditioned or unconditioned spaces will require short-term metering to assess.

Efficacy of Refrigerator Audit Methods

A primary task of this study was to compare methods of estimating the annual energy use of refrigerators. The efficacy of a method can be judged by these criteria:

- How accurately does the method estimate annual energy use?
- How well does the method qualify existing refrigerators for replacement where 0% mistaken replacements (i.e., false-positives) and 0% lost opportunities (i.e., false negatives) define perfect?
- What is the value of the estimation method – what proportion of the potential net benefits from refrigerator replacements are captured by using each method to make replacement decisions?

Table 3 summarizes the findings for each audit method. Note, this table applies to primary kitchen refrigerators only. The net benefits analysis in the table is based on assumptions from a 2001 NSTAR planning document and may differ substantially for other utilities or timeframes, but the relative performance of different audits should be comparable for programs with comparable replacement rates.

Table 3. Summary of Audit Method Performance – kitchen refrigerators only

	Statistical Accuracy		Decision Making Errors (using RCS 1,175 kWh/yr threshold)		Value	
Audit Method	**Average Error**	**Average Bias**	**Lost Opportunities**	**Mistaken Replacements**	**Net Benefits $/audit**	**Notes**
Best Estimate from detailed meter data	0%	0%	0%	0%	$55	"gold standard" for assessing other methods
2 Hour *Ideal* Metering	16%	0%	9%	7%	$51	Based on analysis of all possible short-term metered values from data logging
1 Hour *Ideal* Metering	20%	0%	12%	8%	$48	
New Audit Adjusted Rated Usage	19%	-2%	6%	11%	$50	developed in this study
111% Rated Usage	21%	-3%	11%	11%	$42	simpler ratio of label rated value
PIV Adjusted Rated Usage (as is)	31%	+20%	0%	36%	$17	PIV audit as found
PIV Adjusted Rated Usage (corrected)	26%	+11%	2%	28%	$23	PIV audits with CSG error fixed for 2003 units
PIV Metered *	37%	+27%	0%	25%	n/a	PIV short-term field meter audits (n=24)

Notes:
-"*Ideal* Metering" refers to using the data logger data to calculate short-term metered results of each unit and reflects an ideal version of short-term metering without auditor mistakes or disturbances.
*-"PIV Metered" refers to the 24 kitchen units that were audited with short-term metering by the PIVs. These are different units than the remainder of the table which was restricted to only rated usage audit units.
-"Average Error" is the mean absolute error (discrepancy) expressed as a percentage of true usage.
-"Average Bias" is the mean error and indicates whether the method systematically over or under estimates usage.
-"Lost Opportunities" are the percentage of units audited that should have qualified for replacement, but didn't.
-"Mistaken Replacements" are the percentage of all units that should not have qualified for replacement, but did.
-"Net Benefits" is the net present value of the lifetime energy savings minus applicable purchase costs from replacing units according to each audit method, divided by the number of units audited. **Financial assumptions for these calculations were based on a 2001 planning documents from NSTAR and will vary for each utility**. The cost of the audit itself is not included in the analysis, but the results allows one to weigh audit costs against benefits.

The table shows that:

- 2 hour "ideal" metering is the best audit approach – it has the highest accuracy, the most reliable overall decision making and the highest net benefits per audit.
- 1 hour "ideal" metering is a little less accurate than 2 hour metering but captures nearly the same net benefits per audit because most of the decisions affected by metering length involve units that are marginal from a cost effectiveness standpoint.
- Actual short-term metering (avg. 75 minutes) as performed by the PIVs in the small sample of meter audits performed considerably worse than the ideal short-term metering results (which had an average error of 25% for one hour metering in the same sample). This difference may be due to problems with field practices and protocols (incorrect, inconsistent, or non-existent temperature adjustments, incorrect time of day adjustments, rounded elapsed times), and perhaps due to real world difficulties in capturing undisturbed average metered data. An improved and consistently applied metering protocol is needed to realize the full potential of short-metering.
- The new proposed adjusted rated usage audit performs comparably to short-term metering, but requires collecting information on several refrigerator characteristics.
- Simply estimating usage as equal to 111% of the label-rated value also performs well but did suffer from a few more lost opportunities and provides $8 lower net benefits per audit than the proposed new adjusted rated usage audit (the difference was $11 in a larger sample). Still, this approach has the advantages of simplicity and objectivity (no judgment calls about door seals).
- The existing PIV rated usage approaches were not very accurate and overstated usage by 20% on average, leading to a high rate of mistaken replacements. A variation on the PIV methods that corrected for an error discovered in CSG's rated usage approach (which they corrected themselves halfway through the project) still found the PIV methods lacking. The simple 111% approach performs much better than the PIV approaches and the new proposed approach produces more than double the net benefits per audit.

One of the findings from this project is the inherent uncertainty in field data collection especially, but not solely, when that data comes from occupant reports. This finding argues for parsimony in field data collection for any recommended audit, weighing the value of information against its cost and uncertainty.

For refrigerators located in basements and garages, the findings were quite different. While short-term metering worked well for these units, all approaches based on rated usage performed poorly. The cost-effectiveness analysis revealed that all rated usage approaches produce substantial **negative** net benefits when used to audit refrigerators located outside the main living space. The problem with rated usage approaches is that many basement units use less than the label-rated value due to low occupancy loads and cooler temperatures, but a significant proportion use much more than the rated value because they are malfunctioning. We were unable to find a reliable method for distinguishing between the two types except short-term metering.

Other Findings and Observations

Program Savings Adjustment Factors: We developed initial program savings adjustment factors (a.k.a. realization rates) for each audit approach. This analysis found that the PIV audit approaches predicted savings of about 1,295 kWh/yr per qualified refrigerator but the actual savings would average 988 kWh/yr, yielding a savings realization rate of 76%. If one excludes the units audited by CSG prior to their correction of the temperature default, the realization rate was 80%. The estimated realization rates varied from 62% to 85% by program.

Load Shapes: We analyzed the load shapes of the refrigerators and found that the new replacement units have a much "peakier" load shape **--** with peak hour usage nearly 24% above average, compared to existing units where peak hour usage is less than 9% above average.

PIV Metering Protocols: Our analysis of the results from the small sample of units where the PIVs used short-term metering suggest that the metering protocols and calculations may need revision and that actual in-field short-term metering audits may suffer from some inaccuracies due to rounded recording of elapsed time, changes in door openings and occupancy loads, and perhaps unplug/re-plug cycles that occur during the audit. We developed guidelines for auditing in general and metering protocols in particular (see Section 11 on page 49)

Cost Effectiveness and RCS Program Rules: Our analysis of cost-effectiveness found that 20% of the units with usage above the 1,175 kWh RCS threshold were not cost-effective to replace based on program planning assumptions (generally larger units with marginal usage) -- the simplified RCS program rule of offering incentives to all units using more than 1,175 kWh/yr instead of basing the decision on a cost-effectiveness calculation is reducing potential program net benefits by about $12 per audit – equal to about 16% of the potential benefits from refrigerator replacements in these homes.

Recommendations

- Existing PIV refrigerator audit approaches should be revised or replaced.
- Refrigerators in the living space should be audited using either the new proposed adjusted rated usage approach developed in this study, the 111% of label-rated usage approach, or through short-term metering of at least one hour and preferably two hours.
- The moderate improvement in accuracy from extending metering from one hour to two hours produces only a small incremental benefit to the program because the decisions most likely to be affected by longer metering are for units with usage near the cost-effectiveness threshold. The units with large net benefits and high usage are generally properly identified by shorter term metering. Regardless of length, short-term metering requires consistently following a field protocol and applying a temperature correction. A well designed and executed field protocol is likely to be more valuable than metering for longer periods.
- Refrigerators in basements, garages, and other unconditioned or semi-conditioned spaces should be audited using short-term metering. These refrigerators are often logistically easier to meter than kitchen refrigerators so a dual audit approach should not pose much problem.
- A key factor in maximizing cost effectiveness for any audit approach is to properly identify all very high use units without diluting net savings too much by replacing many low use units.
- The refrigerator replacement protocol for RCS should be revisited to ensure that qualified replacements are at least projected to be cost-effective.

2. Background

The Refrigerator M&V project is a research project designed to better understand the field performance of refrigerators, assess the accuracy of refrigerator audit/diagnostic methods, and measure the energy and load impacts of refrigerator replacements more reliably. The project is sponsored by NSTAR Electric & Gas, National Grid local utilities Massachusetts Electric Company (MECO) and Narragansett Electric Company (NECO), and Northeast Utilities' local utility Western Massachusetts Electric Company (WMECO).

Utility Programs

The specific utility efficiency programs included in the project were:

- NSTAR Residential High Use (RHU) program;
- NSTAR Residential Conservation Service (RCS) – RCS is also known as the Massachusetts Home Energy Service;
- MECO RCS;
- WMECO RCS; and,
- NECO EnergyWise single family.

These five programs all serve residential customers and offer rebates for encouraging customers to replace inefficient refrigerators with new Energy Star rated units as part of their more extensive services. The RCS programs in Massachusetts share a common design – they are open to any residential customer and all offer rebates of $300 for a new refrigerator if the existing unit is believe to use more than 1,175 kWh per year. The RHU program, which was cancelled in early 2004, had been targeted to higher use customers and provided more extensive incentives than RCS. Refrigerator rebate levels varied with the style, size and usage of the existing unit. The Energy Wise program refrigerator incentives also vary with the size and usage of the existing unit. Table 4 summarizes the refrigerator qualifying criteria and incentives for the five programs.

Table 4. Utility Program Refrigerator Incentives and Qualifying Criteria

RCS Programs			NSTAR RHU			NECo EnergyWise		
Unit Type	Usage	Rebate	Unit Type / Size	Usage	Rebate	Size	Usage	Rebate
All	1,175	$300	TF <=16.5	1,112	$300	13+	1,000	$100
			TF 16.6 to 18.9	1,200	$325	14+	1,200	$150
			TF 19.0 to 20.9	1,376	$400	15+	1,400	$200
			TF >=21.0	1,563	$450	16+	1,600	$250
			SbS <=21.9	1,885	$550	18+	1,800	$300
			SbS 22.0 to 23.55	1,904	$550	19+	2,000	$350
			SbS 23.6 to 25.4	2,017	$600	22+	2,200	$400
			SbS >=25.5	2,095	$625	24+	2,400	$450
			BF <=21.0	1,448	$400			

Three quarters of the units monitored in this project were from the RCS programs, so the simplified criteria from that program were the focus for assessing the decision making accuracy and cost-effectiveness of different refrigerator audit approaches.

Prior to 2003, the RCS program had varying, but generally larger, incentive levels and employed a site-specific 7 year payback criteria for qualifying units rather than a simple usage threshold. These program changes reportedly led to a considerable decline in the proportion of units qualifying and a decline in the proportion of qualifying units that actually were replaced.

Program Contractors and Refrigerator Audit Approaches

The five utility programs are implemented by a total of three contractors, referred to as Program Installation Vendors (PIVs). The MECO RCS and NSTAR RCS programs are operated by Conservation Services Group (CSG). The NSTAR RHU and WMECO RCS programs are operated by Honeywell DMC. The NECO EnergyWise program is operated by RISE Engineering. The project plan called for having the PIVs screen sites, collect data, and deploy metering equipment as part of their regular program work. The metering equipment would then be retrieved by CSG, as a contractor to the evaluation team.

Each PIV has a different approach for estimating the annual usage of a refrigerator but there are two main methods – adjusted rated usage and short-term metering.

Adjusted rated usage methods involve looking up the original label-rated energy usage for each refrigerator, typically using a large database with information on tens of thousands of different models from DOE-based testing used for the energy guide labels. The DOE test procedure puts the refrigerator in a 90°F room but has no food loading or door openings. If the unit has an icemaker, the icemaker is not operated. The high room temperature is meant to offset the elimination of occupancy loads. The label rated usage value from the database is then adjusted for factors believed to affect real world usage such as the age of the unit, defrost type, and door style. Rated usage approaches have the advantage of quick decision making (no waiting for a metering period) and avoid problems associated with moving refrigerators and unplugging them and then re-plugging them in. Some of the challenges for the adjusted rated usage approach include:

- not all refrigerators have readable model information;
- some models can't be found in the database or model number matches are imperfect leading to potentially incorrect rated usage values;
- the relationship between original rated usage and actual usage is not clear and the adjustment factors used by contractors often have little empirical basis; and,
- some models malfunction and run continuously, using much more energy than the rating (although some contractors will opt to meter a refrigerator if they suspect that it's a high user)

Short-term metering methods involve plugging the refrigerator into a power meter that accumulates the total kWh of consumption over a relatively short period of time – typically one or two hours. The measured usage is then scaled up to an annual estimate based on the elapsed time and often an adjustment for room temperature is made. The disadvantages of short-term metering include:

- difficulties in moving some refrigerators to access the power cord and potential damage to customer flooring or the refrigerator itself;
- the added time needed for metering, particularly if the home visit would otherwise be shorter than the metering period;
- the cost of metering equipment;
- the potentially large errors if a defrost cycle occurs during metering (usage often triples during a defrost as heaters are used in the freezer compartment to melt frost build up) – some audit protocols require either monitoring for a defrost cycle by tracking peak watts or checking freezer temperatures, or avoiding defrost by using a timer screw accessible on some units;
- random variations in short term consumption that may make a short period unrepresentative; and,
- potential errors if large warm food loadings occurred immediately preceding the metering.

In terms of the length of time to meter, many program protocols require at least two hours of metering and some require three hours based on various research reports and trade press articles[1]. The length of metering has become an issue especially for programs where the length of the home visit is sometimes shorter than the required metering time. The target programs for this project have allowed the PIVs wide latitude in developing refrigerator audit approaches. One of the objectives of the research project is to assess the strengths and weaknesses of the differing approaches and develop guidelines and recommendations for an optimal approach.

Prior to conducting either type of audit, each PIV employs some pre-screening so that audits are only conducted on relatively older refrigerators that customers may be willing to replace. This screening eliminates about half of the refrigerators from consideration. This project is focused on assessing the accuracy of the audits in estimating annual energy usage for units that pass this initial screening.

Audit Methods: RISE Engineering (NECO EnergyWise)

RISE Engineering used short-term metering as their audit method. They metered each unit for 60 to 90 minutes and used a software application (AMP Calc from NGRID) that adjusts for how the current room temperature varies from the annual average. The software provides choices such as 3°F-5°F warmer (cooler) than normal, 6°F-10°F warmer, and then takes the midrange of the temperature bin and adjusts the usage by 2.5% per °F. According to RISE staff, the temperature adjustment is often not made.

Audit Methods: CSG (MECO RCS and NSTAR RCS)

CSG primarily employed an adjusted rated usage approach. They used a model developed by Pacific Northwest National Labs[2] that adjusts the rated usage based on age and defrost type and includes an adjustment to reflect the probability that a unit is malfunctioning and running all the time. CSG did not include a temperature adjustment mentioned separately in the study as an adjustable default. In the first waves of project metering, we found that the CSG's usage estimates were quite high -- averaging 150% of rated usage. A closer reading of the PNNL study revealed that their default indoor temperature was 79°F(!) based on the study locale of public housing in New York. We developed adjusted usage estimates by multiplying CSG's estimates by 0.818 to shift the default temperature to 70°F. These modified estimates averaged 123% of rated usage. CSG revised their audit software on January 2, 2004.

CSG was unable to find the model in the rated usage database for 15% of the units and used short-term metering instead. The metering typically lasted for 45-60 minutes. CSG scales the short-term metered results to a year and adjusts the result by 2.15% per degree difference from 68°F and also uses a small time of day adjustment based on load profile information from a study in California[3].

Audit Methods: HDMC (NSTAR RHU and WMECO RCS)

HDMC used a proprietary rated usage approach that adjusts for size and age. The model estimated usage at 131% of rated on average but ranged from 106% to 159% (these extremes are likely mistakes) with the vast majority of estimates between 124% and 140% of rated usage. In 9% of the HDMC cases, the rated usage couldn't be found and they used short-term metering. HDMC metered for exactly 2 hours in each case and simply scaled the metered usage to a full year without any adjustment for room temperature.

---

[1] see, for example, Cavallo, J. and J. Mapp, 2000. "Monitoring Refrigerator Energy Usage," Home Energy May/June 2000; Moore, A., 2001. "Incorporating Refrigerator Replacement Into The Weatherization Assistance Program: Information Tool Kit", D&R International; Kinney, L. "Refrigerator Monitoring, A Sequel," Home Energy Sep/Oct 2000.

[2] "The New York Power Authority's Energy-Efficient Refrigerator Program for the New York City Housing Authority—1997 Savings Evaluation", R.G. Pratt and J.D. Miller, 1998. see www.eere.energy.gov/buildings/emergingtech/pdfs/sear3.pdf (also see sear2.pdf and sear1.pdf). The model was based on just 46 auto defrost units and most of those units were small (14 ft^3 avg.) and relatively new (avg. 5 years old).

[3] "PG&E Refrigerator Metering Costing Period – Part Two," Proctor, J., M. Blasnik, Z. Katsnelson, G. Dutt and A. Goett, Proctor Engineering Group / PG&E 1994.

3. Research Plan

The project research plan included the following main tasks:

- monitor the hourly electricity usage and room temperatures for 160 existing refrigerators and 30 new Energy Star replacement units in a sample of homes drawn from four target programs;
- use the program PIVs to screen potential sites, perform site data collection about the refrigerators and households, and deploy metering equipment as part of their regular work;
- deploy the metering in five waves spread over the course of year to reflect varying weather and other seasonal effects;
- retrieve the meters (using CSG in their role as part of the evaluation team) within two to three weeks of deployment;
- develop a model of indoor temperatures by analyzing the temperature data with weather data and information about refrigerator location and occupant-reported thermostat settings;
- analyze each site's usage and temperature data to develop an estimate of annual usage, correcting for differences in temperature between the metering period and an estimate of the site's annual temperature;
- assess the accuracy of the different PIV refrigerator auditing techniques including adjusted rated usage and short-term metering;
- attempt to develop an improved refrigerator auditing technique based on refrigerator and site characteristics directly observable during a typical field audit, such as rated usage, refrigerator age and condition, household size, and estimated indoor temperatures., and assess the value of this approach compared to short-term (<=2 hour) metering;
- assess the energy usage of new replacement refrigerators compared to their rated usage values
- develop program savings adjustment factors, to the extent feasible, based on the PIV-estimated energy savings and actual usage results from the detailed data
- develop load shape estimates for the existing and new refrigerators to assess load impacts.

The only significant revisions to the plan involved timeline delays due to difficulties encountered by the PIVs in achieving metering deployment goals.

Sampling Plan

Given the research orientation of the project objectives, the target population for sampling did not need to be based strictly on representing current program participants. The sampling task was further complicated by revisions in the RCS program design that changed incentive levels and qualifying criteria as well as changes in PIV auditing approaches. These changes led to an initial decline in customer refrigerator replacements and some discussion that additional program design revisions may be needed.

Given the potentially shifting nature of the main program design, the Evaluation Working Group decided to define the target population as units that were likely to be replaced under the current, prior, or potential future program designs (depending on incentive levels). Therefore, a unit was considered part of the population of interest if the unit's usage is estimated by the PIV to meet the current program replacement incentive usage threshold (or >80% of the threshold for NSTAR RHU and MECO EW) and the homeowner expresses an interest in potentially replacing it. The 80% threshold for two of the programs was selected so that the sample would include at least some units that current procedures classified as unqualified, allowing some assessment of potential lost opportunities.

Another sampling design issue concerned secondary refrigerators. The Working Group decided to include secondary refrigerators in the project, opting for a more representative sample at the cost of greater expected variability.

The overall sampling approach was to select a random sample stratified by program (to meet the program target metering levels) and clustered by auditor (i.e., select a sample of auditors within each program). By working with fewer auditors, we expected to get higher quality and more consistent data collection. During each metering wave, each auditor was assigned a sample size typically between 3 and 6 units and was provided with all equipment needed for metering their sample. On the start date of each wave, the auditors began deploying the metering on a "first-come first-served" basis for all houses that met the following criteria:

- the auditor has not already deployed a meter that day if the day's work was scheduled to minimize travel time (and has not already deployed a meter in the same house);
- the contractors' normal screening/auditing of the refrigerator indicates that it is a candidate for replacement and has usage that meets or exceeds the program's replacement threshold (or 80% of the threshold for NSTAR RHU and NECO EW);
- the customer states that they are interested in replacing the unit if they qualify for an incentive;
- metering is feasible (unit can be moved, outlet considered safe); and
- the customer agrees to be metered and be home for meter retrieval in two or three weeks.

At the start of the project, we anticipated that about 50% of all units would meet the usage criterion, about one third to one half of those units would meet the customer interest criterion, and two thirds or more of those units would meet the last two criteria. Based on these assumptions, we expected meter deployment to require about two to three weeks for each wave. The actual experience in the project found lower qualifying rates and other challenges to meter deployment, leading to delays.

The project plan included metering 30 new replacement refrigerators. These sites were recruited from the customers whose existing refrigerators were metered previously in the project. CSG, as a member of the evaluation, performed all customer recruitment, meter deployment, meter pick ups in these units.

Analysis Plan

The analysis plan included four primary components:

- An analysis of refrigerator energy usage *within* sites due to variations in temperature was used to estimate annual usage for each unit. Two versions of annual usage were developed – one at a "typical" indoor temperature (70°F) to reflect a standard condition, and one at an estimate of the site-specific annual temperature (based on temperature modeling described below). These estimates were then compared to estimates from the various contractor's auditing approaches, variations on those approaches, and potential new approaches developed as part of this project.
- An analysis of refrigerator usage variations *between* sites due to differences in rated consumption, refrigerator characteristics (style, icemaker, condition, size, age), operating environments (indoor temperature, refrigerator location / confinement), settings (anti-sweat switch, compartment temperatures), and occupancy. This analysis assessed each of the auditing approaches, particularly comparing the value of short-term metering compared to rated usage methods.
- An analysis of indoor and outdoor temperatures to develop an indoor temperature model based on measurable site characteristics such as reported air conditioner usage and summer and winter thermostat settings. This analysis provides a basis for predicting annual usage for each sites.
- An analysis of hourly load shapes based on energy usage and other measurable factors. Load shapes models were developed to estimate hourly loads during periods of interest.

4. Data Collection and Meter Deployment

The project required an extensive data collection effort that involved energy usage and temperature data logging at nearly 200 houses and collecting site characteristic data during each meter deployment and pickup.

Metering Equipment

The project used three types of equipment: kWh dataloggers, temperature dataloggers, and refrigerator thermometers.

The kWh datalogger was a WattsUp Pro (see www.doubleed.com) . This plug-in logger records cumulative true kWh usage. The meter logs 1000 records of data with a time adaptive resolution, starting at 1 second and deleting every other record when memory fills and re-starting at record 501. For data logging of two to three weeks (up to 23 days), the resulting data resolution is 34 minutes. For periods of 23-46 days, the resolution is 68 minutes. Because the logger stores cumulative kWh, deleting every other record will not affect the kWh readings. Data are stored in non-volatile memory and can be downloaded using a serial cable and supplied software. Several meters failed and required re-metering of some sites.

Temperature logging was done using a HOBO model H08-002-02 temperature logger that can store nearly a year of hourly average temperature data. The loggers were positioned in the room with the refrigerator in a place thought to be representative of the temperature, but not intrusive to the customer.

Two standard commercial refrigerator thermometers were provided to each meter deployment auditor and pick-up technician to measure freezer and refrigerator compartment temperatures during each visit.

Site Characteristic Data Collection

Information about the refrigerators and the households was required for developing models of refrigerator usage and creating a new and improved adjusted rated usage audit approach. A list of potentially useful data elements was developed and refined with Working Group input during the development of the project research plan. The items included on the field data collection form included:

Refrigerator Information: make, model, year of manufacture, volume, rated usage, door style, defrost type, icemaker information, color (was it a popular 1970’s color like harvest gold or avocado?), overall condition and door seal condition (as rated by the PIV), location in the home, whether it is recessed into cabinets or not, clearances on each side, special location information including high sun exposure, located on an exterior wall, wood stove nearby

PIV Audit Information: the PIV’s estimate of annual usage and, if short-term metering was used, the meter reading and length of metering

Refrigerator Settings: the presence and position of the anti-sweat heater switch, the settings for the fresh food and freezer temperature controls, spot measurements of the fresh food and freezer compartment temperatures

Occupant Information: how the refrigerator was obtained (bought new, came with house, got used), any prior repairs, occupant reported use of anti-sweat switch and temperature controls, number of occupants (break out into adult, school age children, and pre school age children), frequency of people home during weekdays, reported winter thermostats setting and setbacks, cooling system information (type, settings, and frequency of use), and whether any special conditions may affect usage (including frequent cooking, rare cooking, refrigerator doesn’t keep food cold, refrigerator runs all the time, refrigerator room not heated)

All of this information was collected using a single page one-sided form that was provided to the contractors. The same information was collected again when the metering equipment was retrieved by

CSG. This duplicate data collection provided a means to assess how consistently and accurately the field information could be obtained. A copy of a field data collection form is provided in appendix A.

In addition to the data collection forms, customers who agreed to metering signed a customer agreement form that delineated their responsibilities (i.e., don't remove the metering equipment or replace the refrigerator until after the metering equipment is retrieved) and stated that they would be paid $25 to thank them for their participation.

Michael Blasnik, the evaluation team project manager, provided training to each PIV at their offices during June 2003. The training covered an overview of the project and the contractors' role in screening sites, collecting data and deploying meters. At each training session, the PIV's were provided with sufficient metering equipment for all sites for the first wave of metering. They were also provided with all paperwork, including step-by-step instructions, data collection forms, and customer agreement forms. Each contractor identified a sufficient number of auditors for meter deployment so that no auditor would need to deploy more than 6 meters in a wave. In future waves, the necessary metering equipment was shipped to each PIV at the start of the wave.

Meter Deployment

The project work plan called for the five waves of refrigerator metering to occur in June, July, August, and November 2003, and January 2004. Metering was planned for four programs – NSTAR RHU, MECO RCS, WMECO RCS, and NECO EnergyWise. These plans changed due to problems with PIV meter deployments. A fifth program was added – NSTAR RCS – to shift some of the burden from the RHU PIV to the NSTAR RCS PIV. Later in the project the RHU program was canceled and the remaining metering was shifted to RCS. The deployment problems led to the delay of subsequent metering waves throughout the project so that the final metering wasn't completed until late May (originally scheduled to end in February and later postponed until March).

In the first wave of metering, begun in mid-June 2003, one of the three PIVs required 50 days to deploy their meters (the other two PIV completed their deployments within 16 days). The reasons for this delay were never clear. The delay of wave 1 into August eliminated any chance of completing the planned July and August waves before the end of the summer. Instead, we set the wave 2 deployment to begin on August 11th. To avoid another major delay, NSTAR metering was shifted from the RHU program to the RCS program (operated by a different PIV). The wave 2 deployment went more smoothly, but late August vacations caused some metering to slip into early September. Seven units shifted back to RHU for wave 3 which didn't begin until early November. The length of metering was extended for most units in wave 3 because of the Thanksgiving and Christmas holidays. Extended metering would allow the analysis to exclude the major holidays if needed. The metering delays forced the "January" metering wave into mid-February through March. The final wave began as the wave 4 meters were picked up. By this time, the RHU program had been cancelled and all NSTAR metering shifted back to RCS. More delays led to meter deployments being suspended on May 14th, two units short of the goal, to allow time for analysis. The meters were all retrieved by May 21, shortening the data collection for some units, but shortening the analysis and report writing time even more.

Figure 1 graphically summarizes the metered data collection for the project. The shaded area shows the number of sites metered on each date with the five humps corresponding to the peak of each metering wave. The vertical lines show the minimum and maximum temperatures in Boston for each day (see the right hand scale), allowing an assessment of how different weather patterns were covered by the data.

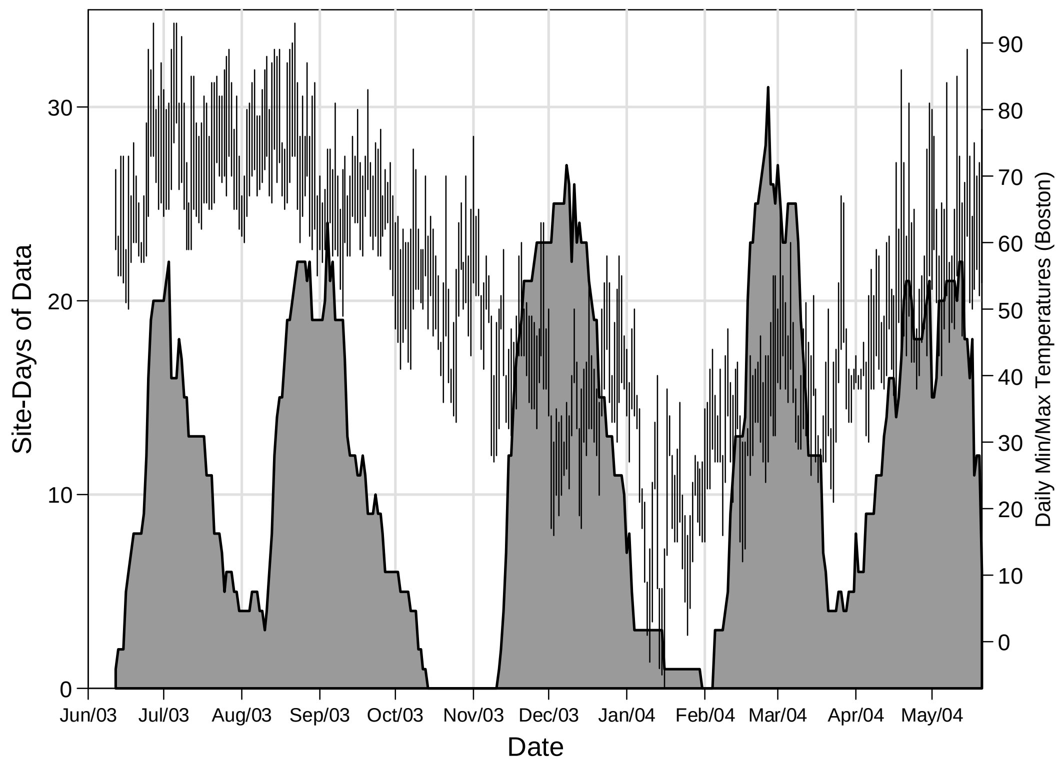


Figure 1. Site-Days of Metering over the project, with daily temperature ranges.

The figure shows that hot weather was fairly well represented by the first wave of metering due to heat waves in late June and early July. Moderate weather was mostly missed in the Fall, but was captured in the Spring of 2004. Cold weather is well represented in the third wave, although the extreme cold in mid January was mostly missed.

Data Quality: Metered Data

Several data quality issues were encountered during the project:

- The electric usage loggers failed at several houses. In all but one case, the CSG retrieval technician was able to deploy a replacement meter and came back in two to three weeks. In one of these cases, the second meter also failed and no further metering was pursued.
- Ten customers apparently didn't abide by the terms of the customer agreement form because they decided to unplug the electric usage loggers themselves. In some cases, they didn't want to make the retrieval appointment so they left the logger outside or with a neighbor. In other cases, they replaced their refrigerators during the metering. In a few cases, no reason was given. Although the meters maintain their data when unplugged, problems arose from the fact that they have no real time clocks[4]. One customer never returned the temperature logger.

[4] The lack of a true clock in the meter was considered a small inconvenience when planning the project since CSG technicians would be there to time stamp the data file (time stamping could not be accomplished during deployment because the PIVs lacked the skills and computers). When the customer removes the meter, data collection stops and the pick-up technician does not know the proper ending time. The PIVs sometimes recorded the deployment time using the appointment time or a rounded off time leading to some uncertainty in the precise time for the data.

- Temperature loggers recorded unrealistic data in a few houses. In one case, a kitchen temperature was recorded at over 90°F and in another case the logger was placed next to the furnace and recorded average temperatures of about 80°F in a basement.

Overall, two sites were completely lost from the analysis due to a lack of electric usage data and four sites had a limited analysis due to a lack of temperature data (we developed estimates of annual usage by imputing the average temperature based on weather conditions and location in the house using data from other sites). Overall, we had sufficient data to develop annual usage estimates for 186 refrigerators out of the goal of 190 – two units were never metered and two had no usage data due to meter problems. Table 5 shows the distribution of metered sites by program in comparison to the original plan.

Table 5. Meter Deployments by Program

		Existing Refrigerators				New Units	
Program	PIV	Original Plan	Revised Plan	Actual	Actual (has data)	Original Plan	Actual (has data)
NSTAR RHU	HDMC	70	34	27	27	6	6
NSTAR RCS	CSG	0	36	41	41	6	6
MECO RCS	CSG	45	45	45	44	10	8
NECO EW	RISE	15	15	15	15	3	5
WMECO RCS	HDMC	30	30	30	29	5	5
Total		160	160	158	156	30	30

The adaptive time resolution of the watt hour loggers led to the goal of picking up each meter within 23 days of deployment so that data would be recorded at about a half hour resolution (34 minute actually). Overall, 138 of the meters were picked up within that timeframe (five had 17 minute resolution and one unit with shortened data collection had 9 minute resolution). Delays in meter pickups, some due to the holidays, led to 46 meters having 68 minute resolution and two meters having 137 minute resolution.

We developed automated routines for reading the data files from the two types of loggers and combining this information with data from the data collection forms. Special attention was paid to ensuring proper time synchronization of the data files and a cross check was made with contractor information on meter deployment and pick-up times. Anomalies were investigated and typically resolved to within one hour.

Data Quality: Field Data Collection

In terms of field data collection by the PIVs, the data quality was about as expected – some items were left missing on some forms, some entries were illegible, some entries appear suspect, and some entries were in conflict between the deployment contractor's form and the pick up contractor's form. Data discrepancies were resolved following a set of rules:

- if an entry was missing or not noted on one form (e.g., deployment or pick up), then the entry from the other form was assumed to be correct;
- numerical items (such as thermostat settings and refrigerator temperatures) and scaled ratings (such as door seal condition) that could be coded numerically were averaged between the two forms;

There were particular problems with recording the position of the anti-sweat heater switch – about 25% of the time the entries differed between the deployment and pickup form. This discrepancy may have been at least partly due to understanding how to interpret the varying labels provided by manufacturers (e.g., "saves energy" means off and "reduces condensation" means on). There was even a discrepancy about whether some units had a switch.

We also found some notable variations in items such as reported thermostat settings and frequency of air conditioner use that may be due in part to different occupants answering the questions on different visits but also to people answering the same question differently on each occasion.

A key item in the field data collection was the label-rated usage of the refrigerator. We used the value from the PIV's deployment form for assessing that PIV's audit approach, but we performed an independent, and perhaps more careful, lookup of each unit using a California Energy Commission database and other sources. If a model could not be found by the principal investigator, CSG's value from the meter pickup form was used (CSG's field personnel use the AHAM database). Although the vast majority of rated usage values from the PIVs agreed closely with the values found in the CEC database, there were discrepancies greater than 10% of usage in 11% of the CSG deployments and 27% of the HDMC deployments. The mean differences were near zero, implying no systematic bias, but we did find that mistakes could be made and some units could not be found or had only potential matches in the databases.

Perhaps one of the findings from this project is the inherent uncertainty in field data collection especially, but not solely, when that data comes from occupant reports. This finding argues for parsimony in field data collection for any recommended audit approach, weighing the value of information against its cost and uncertainty.

Other Data Collection

In addition to the field data collection forms and metered data, the project required data from two other sources: program tracking system data from the utilities and weather data.

The NSTAR, MECO and WMECO programs all provided tracking system data from 2002 and parts of 2003. This data included information about refrigerators replaced by the programs including model numbers, refrigerator costs, incentive amounts, new unit rated usage, and the estimated usage of the existing unit. We used this information to estimate the costs and rated usage of new units for use in the cost-effectiveness analysis and assessment of program savings adjustment factors.

We acquired daily temperature data for Boston, Pittsfield, and Worcester for the period of metering for use in the analysis and modeling of indoor temperatures. We also used TMY2 (Typical Meteorological Year) data for Boston and Worcester to characterize the typical outdoor temperatures for these cities. The lack of TMY2 data for Pittsfield and the similarity of the Pittsfield and Worcester temperature data led us to employ only two weather stations for the analysis.

5. *Refrigerator & Site Characteristics*

Overall, 158 existing units and 30 new replacement units were metered during the project. We used the project work plan as a blueprint for guiding the analysis.

Table 6 summarizes some basic information about the metered sites based on the data from the field data collection forms with break outs for each utility program.

Table 6. Refrigerator and Site Information – by utility program

Characteristic	MECO RCS	NECO EW	NSTAR RCS	NSTAR RHU	WMECO RCS	All Existing	New Units
# Units	45	15	41	27	30	158	30
Year Built (median)	1985	1982	1985	1984	1983	1984	2003
Volume	17.9	17.9	18.5	18.5	18.6	18.3	20.0
Temperature: fresh food	40	41	40	41	43	41	44
Temperature: freezer	5	5	5	9	7	6	10
Label Rated Usage (PIV, n=120)	1185	N/A	1195	1238	1275	1217	N/A
Label Rated Usage (evaluator, n=151)	1207	1255	1249	1226	1302	1244	484
Predicted Usage (PIV, n=158)	1823	1969	1655	1703	1649	1740	N/A
Door Style:							
top freezer	89%	80%	71%	67%	80%	78%	60%
side-by-side	11%	20%	20%	26%	13%	17%	17%
bottom freezer	0%	0%	5%	4%	7%	3%	23%
single door	0%	0%	0%	4%	0%	1%	0%
Icemaker							
internal	4%	7%	15%	7%	10%	9%	27%
thru-the-door	7%	13%	10%	15%	7%	9%	17%
Anti-Sweat Switch							
On	36%	53%	49%	44%	43%	44%	0%
Off	40%	7%	24%	15%	33%	27%	0%
None	24%	40%	27%	41%	23%	29%	100%
Door Seal has gaps	13%	13%	7%	4%	7%	9%	0%
Location:							
Kitchen / Living Space	73%	100%	83%	59%	93%	80%	83%
Heated Basement	7%	0%	0%	11%	3%	4%	0%
Unheated Basement	18%	0%	17%	30%	3%	15%	17%
Unheated Garage/Porch	2%	0%	0%	0%	0%	1%	0%
Not Primary Refrigerator	27%	0%	17%	44%	7%	21%	17%
Bought/Got Used	24%	13%	12%	30%	23%	21%	0%
Harvest Gold / Avocado Color	20%	0%	12%	15%	17%	15%	0%
# Occupants	2.7	2.5	3.0	3.8	2.5	2.9	3.3
Home all day	38%	27%	54%	41%	50%	44%	60%
Avg. Winter Thermostat Setting	65.6	66.5	66.4	66.2	67.4	66.4	65.1
Central Air Conditioning	16%	13%	20%	19%	3%	15%	17%
Room Air Conditioning	9%	7%	2%	7%	7%	6%	3%
Adjust Anti-Sweat	17%	13%	17%	26%	23%	19%	0%

The data on existing units indicates that :

- the typical unit was about 20 years old and 18 cubic feet;
- more than three quarters of the units had a top freezer and only about one in six had side-by-side doors;

- rated usage averaged about 1,200 kWh/yr. and the PIVs' predicted annual usage averaging 1740 kWh/yr. For the 118 existing units audited with a projected rated usage approach, the average projected usage was 1680 kWh/yr., 37% greater than the 1220 kWh/yr. rated usage of those units;
- fewer than 20% of the refrigerators had icemakers and half of those were through-the-door;
- most of the units with anti-sweat heater switches had them turned on and more than 80% of the occupants report that they never adjust the anti-sweat heater switch;
- about 80% of the units were located in the living space and about 80% were the primary refrigerator;
- about 20% of the units were bought or acquired as a used appliance (excluding those that came with the house);
- nearly half of the occupants reported that people are home all day – while this value may seem high, it may be representative of the participant populations for these programs;
- occupants reported an average winter thermostat setting of just about 66°F;
- only about 20% of the refrigerators were located in air conditioned spaces;
- only 5 units had manual defrost (not shown in the table due to low frequency)

One surprising finding is that more than 40% of the NSTAR RHU units were not in the main living space of the house, perhaps reflecting a larger proportion of two refrigerator households among that high use population.

Figure 2 shows the number of existing refrigerators in the project by year of manufacture. The PIV pre-screening eliminated nearly all of the units built after 1990.

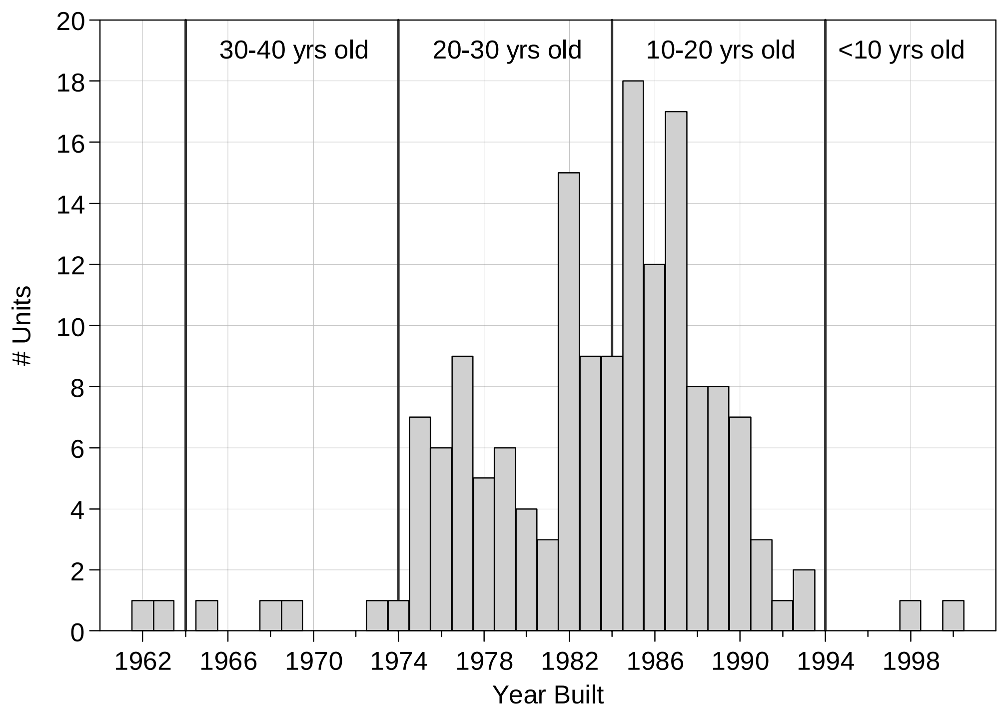


Figure 2. Distribution of Refrigerators by year of manufacture

Among new refrigerators, we found:

- there were many more bottom freezer units (most likely reflecting their increasing popularity in recent years);
- the new units tended to have slightly warmer temperatures in the fresh food and freezer compartments than the existing units -- the average measured temperatures were about 5°F warmer than the temperatures used for DOE rated usage testing (DOE uses 5°F in the freezer and 38°F in the fresh food compartment);
- new units tended to be about 2 cubic feet larger than existing units on average -- this difference is evident even among just the replaced existing units; and,
- new units are more than twice as likely to have icemakers than existing units – in fact only 10% of the units replaced by the new units had icemakers.

Upsizing and extra features may affect program savings and cost-effectiveness. Among the 30 new units, four side-by-side units replaced top freezers and four bottom freezer units replaced top freezers. The six existing side-by-side units were replaced by 1 side-by-side, 3 bottom freezers, and 2 top freezers.

6. Meter Data Analysis

The first step in the analysis of the metered data was to "look" at the data using time series plots and scatter plots of usage versus temperature. The time series plots revealed that several units had usage patterns consistent with a malfunction, such as loss of refrigerant charge or badly damaged door seals, that caused continuous operation. These sites tended to use much more than the rated values and some had high refrigerator temperatures or reports of malfunctions.

Figure 3 shows time series plots of the hourly usage data for a "typical" unit and a potentially malfunctioning unit. The typical unit's usage varies from day to day reflecting temperature and occupancy effects as well as defrost cycles. The "flat" usage site shows a constant usage baseline with frequent defrost cycles, consistent with a unit running continuously. The occupant reported that the refrigerator did not work properly and the refrigerator temperature was 50°F and 32°F in the freezer. This unit was in a basement where poor performance may be more acceptable.

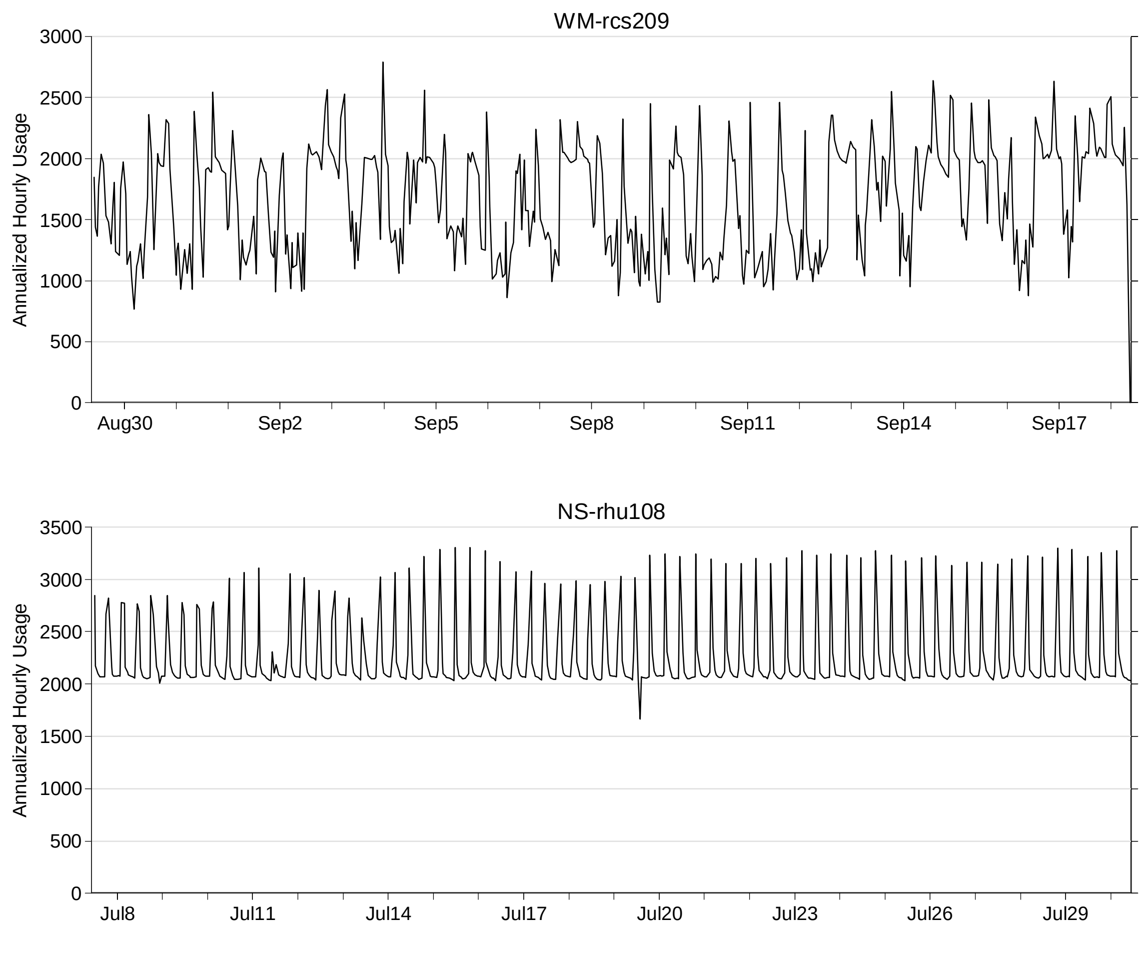


Figure 3. Usage Time Series Plots: "Typical" unit on top, "Flat" usage site

The presence of potentially malfunctioning units requires care in the analysis. These units represent a portion of the population and should be kept in any analysis of audit accuracy or program savings. However, these units may be outliers in modeling usage variations between sites. We developed statistical criteria to identify flat usage units[5] and found 14 of the 186 units (7.5%) exhibiting such usage.

[5] We classified a unit as having flat usage if three quarters of all hours' usage are within 10% of the median and either half of all hours' usage are within 3% or half are within 4% and the temperature range is larger than 5°F.

Six of these units were located in basements. Some units with flat usage may be secondary refrigerators in spaces with fairly constant temperatures and small occupant loads. In addition to the “flat” usage problem, one site apparently malfunctioned during the metering, jumping to very high consumption halfway through the data collection.

Assessing Annual Refrigerator Usage

The electric usage of a refrigerator varies with the room temperature as the rate of heat gain into the cabinet is proportional to the temperature difference. Given a room temperature of 70°F and an average refrigerator/freezer temperature of about 30°F, one should expect usage to vary by about 2.5% per °F (1/(70-30)=.025). The impact may be greater than 2.5% since the efficiency of the refrigeration cycle declines as the temperature rises. *Figure 4* plots the usage/temperature relationship for 15 refrigerators along with a regression fit line and shaded 90% confidence bands for each.

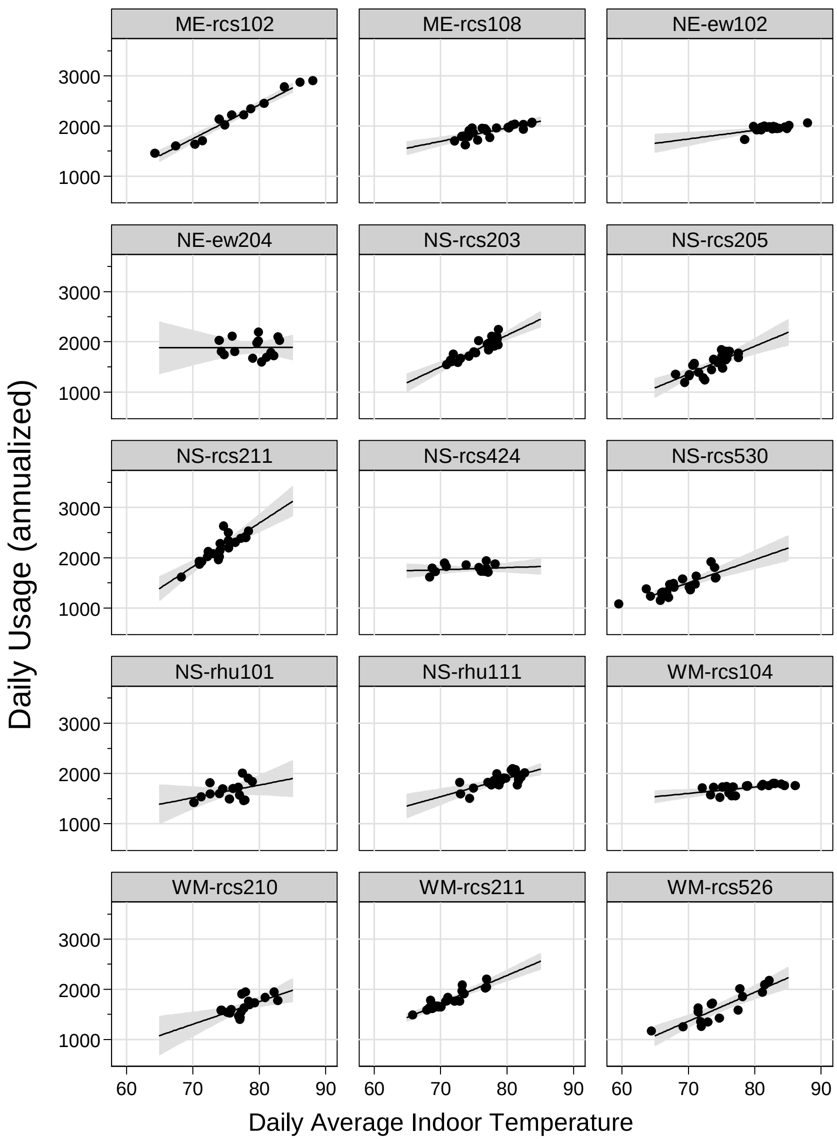


Figure 4. Usage vs. Temperature graphs for 15 refrigerators

The figure shows fairly wide variations in the relationship from one unit to the next. For some units, daily usage varies by almost a 2 to 1 over the course of the metering while other units have little variation in usage. Some units have a clear slope while others show a wider scatter.

Temperature effects may vary between refrigerators due to differences in internal refrigerator temperatures, differences in how room temperatures affect the average temperature difference across the unit and at the condenser, and perhaps variations in refrigerator design. Because of the uncertainty in the slope of the temperature/usage relationship and the likelihood that this relationship varies between units, the work plan called for estimating the temperature impact for each unit

We explored the temperature/usage relationship using graphical and statistical techniques and discovered that occupancy patterns over the course of the day tend to make temperature impacts appear larger than they really are because people are usually sleeping during the coldest indoor temperatures (before 5 AM) and actively using the refrigerator during the warmest indoor temperatures (dinner time). To avoid this bias, we analyzed the usage/temperature relationship using daily aggregate data. We also found that many sites experienced a relatively narrow range of temperatures during the metering, making it difficult to reliably estimate a site-specific temperature slope. We addressed this issue by employing a random coefficients regression analysis that estimates each site's temperature effect as essentially a weighted average of the site specific result and the average result across all sites. Sites with well determined temperature impacts are relatively unaffected by the results from other sites, but sites with poorly determined temperature effects have their estimates pulled toward the overall average, avoiding unreasonable results caused by poor statistical fits. Appendix B provides a more detailed discussion of the usage/temperature modeling.

The regression analysis found an overall estimated temperature slope across all existing units of 2.65%/°F, consistent with engineering principles and prior research. Site specific temperature slopes varied from -0.1% to 6.6% with a median value of 2.67% and half the values falling between 1.86% and 3.59%. An effect of 2.65% may seem small, but the monitored temperature data found that hourly indoor temperatures varied over a range of about 15°F on average within each house's metering period and even daily average temperatures varied by 8°F on average.

The results of the analysis provided an estimate of each unit's annual usage normalized to a room temperature of 70°F. We applied a correction factor to this figure for anti-sweat heater switch usage for 17 units where they appeared to actually use it (see Appendix B for details). The resulting adjusted usage estimates provide a measure of refrigerator efficiency – a usage that is temperature adjusted to 70°F. In reality, the refrigerators experience varying operating conditions and temperatures. The annual averages temperature may differ from 70°F, particularly for units located outside the conditioned living space.

Indoor Temperature Modeling

We used the monitored indoor temperature data to develop a model of indoor temperatures as a function of outdoor temperatures, occupant reported thermostat settings, and the presence of air conditioning. Unheated basements were modeled separately.

We assigned each unit to a weather station based on the distance to one of two selected stations -- Boston and Worcester. Available data from the Pittsfield station in Berkshire County indicated that those temperatures are more similar to Worcester than Albany (and we didn't have access to long-term weather norms for Pittsfield). The outdoor daily average temperatures ranged from -4°F to 86°F and averaged 51°F across the 3,598 site-days of data.

Figure 5 shows the relationship between inside and outside daily average temperatures for all days at all sites (except for one site located in an unheated garage in the winter) using box plots. The "box" part of the plot shows the range of the middle half of the data (25th percentile through 75th percentile) with a line at the median. The vertical lines extend through the furthest point no more than 1.5 times this inter-quartile range away from the median. The graph on the left shows the relationship for the living space (including kitchens, heated basements, and other conditioned spaces). The indoor temperatures are relatively constant in the winter and typically ranged from the mid 60s to about 70°F. The average kitchen temperature in the winter was measured at 67.4°F. During warmer weather, indoor temperatures increase noticeably, reflecting the relatively low penetration of air conditioning in these houses. The graph on the right shows the same relationship for the 28 unheated basements in the sample. Basements are generally cooler than the living space but exhibit a much wider variability and a more gradual increase in warm weather, reflecting their ground coupling.

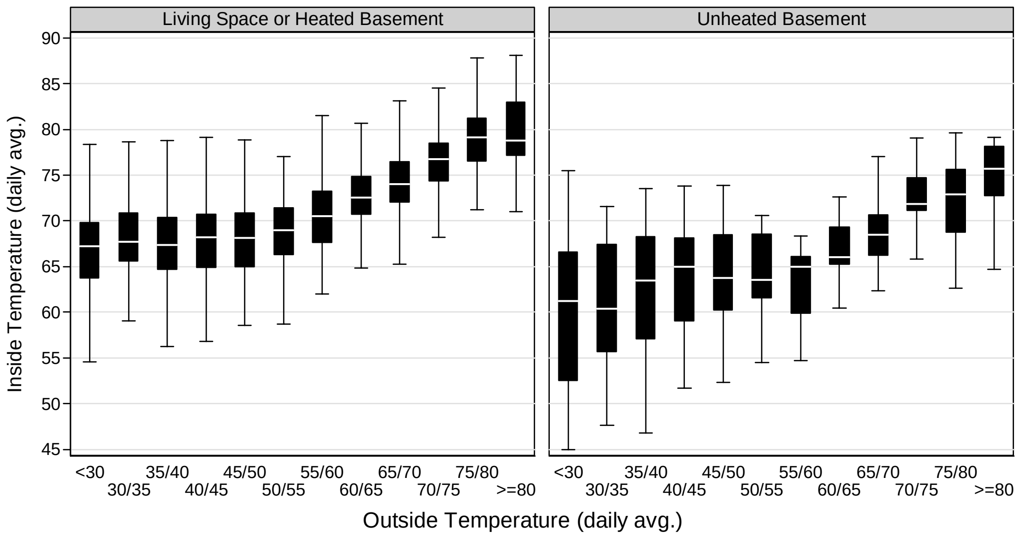


Figure 5. Indoor-Outdoor Temperature relationships by type of space
box plots show the median and 1st and 3rd quartiles as the box, and outer values beyond

Figure 6 focuses on the relationship in warmer weather and breaks out the living space units by the presence of air conditioning.

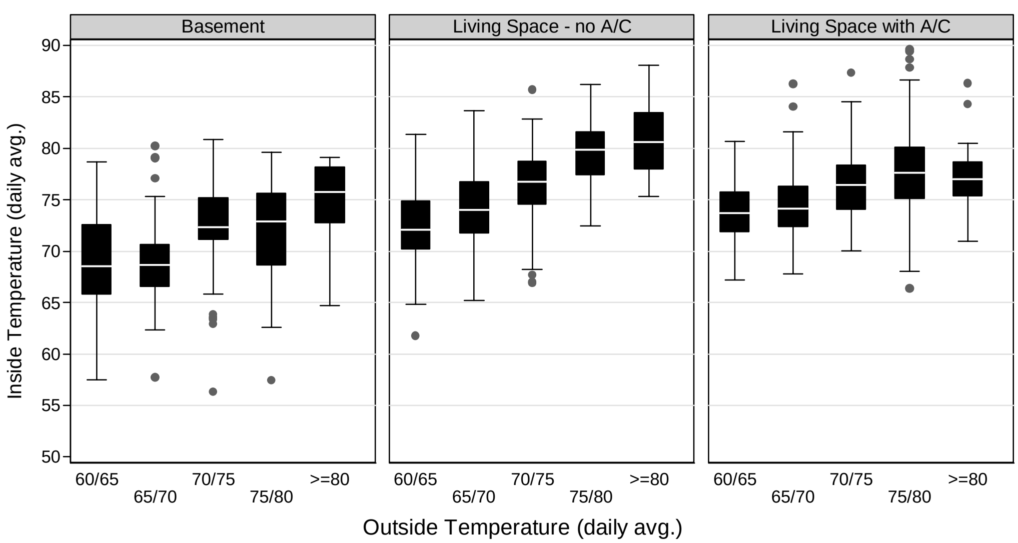


Figure 6. Summer Temperature relationships by type of space

As expected, the graph shows that air conditioned rooms are cooler during hot weather. However the difference in temperature is not very large, only about 5°F during the hottest days, perhaps due to the limited or intermittent use of air conditioning in some houses (e.g., in homes unoccupied during the day).

Based on the examination of temperature patterns, we developed an approach to estimating annual temperatures by splitting the year into three regimes: winter, summer, and mild weather.

For winter weather (defined as days with an average temperature below 60°F), we modeled living space temperatures as a function of occupant reported thermostat settings and heating degree days (base 60°F). For unheated basements, we modeled winter temperatures as just a function of heating degree-days. For other unheated spaces (garages and unheated porches), we had data for just one site and so decided to assume that the temperature would be halfway between inside and outside (an educated guess, at best). We modeled mild weather temperatures as simply the mean of the indoor temperatures when the outdoor daily average temperatures were between 60°F and 70°F. This value was calculated separately for basements and living space and assumed at 65°F for other unheated spaces. For warmer weather, we modeled temperatures as a function of cooling degree days (base 70°F) with separate slopes for air conditioned spaces and basements and a separate constant for unheated basements.

We estimated the annual temperature as the average temperatures during these three seasonal regimes, weighting each regime by its frequency in a typical weather year based on TMY2 weather data for Boston and Worcester. The model coefficients are shown in Table 7.

Table 7. Temperature Model Coefficients

	Living Space	Unheated Basement	Other Unheated Space
Winter (<60°F)			
Constant	22.7	65.9	30
Average Thermostat Setting	.71	n/a	.5
Heating DD60 per day (B=19.2. W=21.5)	-.05	-.16	-.5
Summer (>70°F)			
Constant	76.5	71.4	76.5
Cooling DD70 per day (B=6.0, W=4.5)	.40	.21	.40
A/C or heated basement CDD70/day	-.32		
Mild Weather (60-70°F)	73.7	68.3	65

The proportion of the year in each regime is .627 winter, .171 summer, and .202 mild for Boston and .676 winter, .109 summer, and .215 mild for Worcester. As an example calculation, a kitchen in Boston with a reported average winter thermostat setting of 68 degrees and no air conditioning would have an average annual indoor temperature calculated as:

$T_{winter} = 22.7 + (.71 * 68) – (.05*19.2) = 70.0°F$

$T_{summer} = 76.5 + (.4 * 6) = 78.9°F$

$T_{mild} = 73.7°F$

$T_{annual} = .627*70.0 + .171*78.9 + .202*73.7 = 72.3°F$

The model results can be easily pre-calculated -- the only house-specific input (beyond the type of space) is the reported winter thermostat setting, which proved to be a surprisingly good predictor of indoor temperatures (the regression model t-statistic was 26.6).

Results from the model for the annual temperature in the living space are shown in Table 8.

Table 8. Some Temperature Model Results:
annual living space temperature estimates

Winter T-stat Setting	Boston no AC	Boston w/AC	Worcester no AC	Worcester w/AC
62	70	69	69	69
63	70	70	69	69
64	71	70	70	70
65	71	71	70	70
66	71	71	71	71
67	72	72	71	71
68	72	72	72	71
69	73	72	72	72
70	73	73	73	72
71	74	73	73	73
72	74	74	74	73
73	75	74	74	74
74	75	75	75	74
75	75	75	75	75

The values in the table show a narrow range of estimated temperatures and no real impact from air conditioning. For refrigerators in the living space, the indoor temperature may be reasonably approximated as about 70°F for households that keep very low winter thermostat settings, 72°F for most households with normal temperatures, and 74°F for households with fairly warm winter temperatures. Higher temperatures could occur in houses with very high thermostat settings, perhaps in some senior housing and upper floors of high rise buildings. In these cases, the estimated average winter temperature can be used as the annual average.

One slightly unexpected finding was the relative unimportance of air conditioning because of how few hours per year are affected by it in the moderate New England climate and the relatively small reduction in average summer temperatures for air conditioned houses. This combination of factors makes the impact of air conditioning on average annual temperatures negligible -- less than one half degree.

The temperature model projected average indoor temperatures of 71°F in the living space (ranging from 68°F-75°F), about 65°F in unheated basements, and 58°F in the one garage unit. These average values could be used as defaults for auditors to adjust short-term metered data without losing much accuracy. The advantage of just using average values is to eliminate data collection (occupant reports) and avoid the use of a calculation-based detailed model.

The temperature model may also be used to develop estimates of refrigerator loads for specific weather conditions, such as in forecasting summer or winter peaks. For example, the estimated indoor temperature when it is 90°F outside for a house without air conditioning would be:

$T = 76.5 + (.4 * (90-70)) = 84.5°F$

Therefore, a refrigerator’s usage on that hot day would be estimated at .0265*(84.5-71) = 36% greater than it’s annual average daily usage, if the refrigerator was located in a room with a 71°F average temperature. Hourly demand estimates would be calculated by using this daily usage figure and the load shape estimates developed in section 10.

7. *Annual Usage Analysis Results*

We used the results of the temperature modeling to adjust the annual usage estimate of each refrigerator to reflect the difference between the average temperature and the 70°F usage level using the site-specific temperature slopes. These results provide the best estimate for the true annual usage of each refrigerator at each site. For the remained of this report, we refer to these usage estimates as the “true” or “actual” usage since they represent the most detailed and accurate estimate available, but one needs to recognize that these values are still estimates with uncertainty from the usage/temperature model fit and the estimate of the site-specific annual temperature.

The results of the usage analysis and comparisons to rated usage are summarized in Table 9.

Table 9. Annual Refrigerator Usage (kWh/yr) – averages and comparisons to rated

		Label	Usage @ 70°F		Usage @ T-in	
	#	Rated	Actual	% rated	Actual	% rated
Existing Refrigerators						
All Units	156		1369		1383	
Units with label-rated usage	149	1247	1368	110%	1384	111%
- Side-by-Side	27	1456	1697	117%	1722	118%
- Top Freezer	115	1185	1289	109%	1304	110%
- in Basement	22	1300	1304	100%	1181	91%
- in Living Space	120	1233	1387	112%	1428	116%
New Refrigerators						
All Units	30	484	425	88%	425	88%
- Side-by-Side	5	623	595	95%	594	95%
- Top-Freezer	18	436	367	84%	364	83%
- Bottom Freezer	7	509	456	90%	461	91%

The average annual usage for the 156 existing units was 1369 kWh/yr at 70°F and 1383 kWh/yr at the site-specific temperature estimate. For the 149 units with known rated usage values, the average annual usage was 11% greater than the original rated value. Based on simple comparisons, it appears that units with side-by-side doors tended to have greater usage relative to their rating than units with top freezers. The table also shows that refrigerators located in unheated basements tended to use less than their rated usage primarily due to the lower temperatures found in basements. The usage of these units at 70°F averages almost exactly the original rated usage while units in the living space average 12% greater than rated at 70°F and 16% greater than rated at the estimated indoor temperature. The usage difference at 70°F can likely by attributed to the lower occupancy loads of secondary units (fewer food loadings and door openings). The average usage of basement units masks a wide distribution with nearly one third of the units using less than 80% of rated and nearly 20% using more than 150% of rated.

New Energy Star refrigerators used 425 kWh/yr. on average, equal to 88% of the average label-rated usage of 484 kWh. Side-by-side units tended to use a little more relative to the rated values than top freezer units, but the sample size is too small to draw any reliable and statistically significant conclusions. This finding of usage less than rated for new units is consistent with other research in moderate or heating climates[6]. The higher temperatures in the fresh food and freezer compartments for the new units in this study (compared to DOE test values) could explain all of the observed difference in usage.

The relationship between actual usage and original rated usage for the existing units is shown in *Figure 7*. Refrigerators in basements are shown as hollow squares and potentially malfunctioning units with flat or shifting usage are shown as gray filled circles. A line of agreement with ±20% error bounds is drawn on the diagonal.

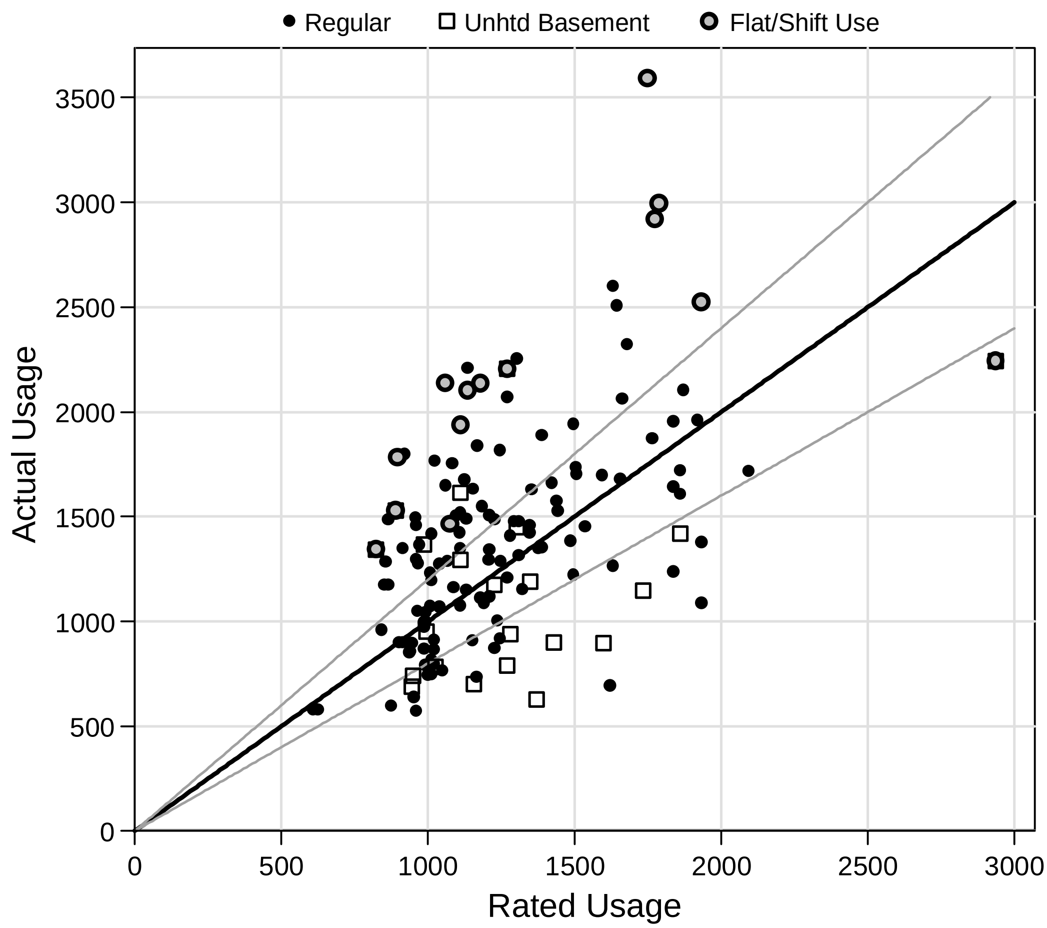


Figure 7. Actual vs. Rated Usage – existing refrigerators

The figure shows a wide scatter around the line of agreement. A regression fit finds a highly significant coefficient on the slope (.79 with a standard error of .11), but the model only explains 27% of the variation in actual usage. As shown in the prior table, most refrigerators in basements use less than their rated usage indicated by the large proportion of squares below the line of agreement, but some use much more. Refrigerators with flat or shifting usage often use much more than the rated value, forming a line almost parallel to the line of agreement but shifted upward by more than 50%.

---

[6] see for example, "Large Scale Residential Refrigerator Field Metering" J. Proctor, G. Dutt, M. Blasnik, A. Goett, E. Galawish, and D. Quigley, Proceedings of the ACEEE 1994 Summer Study on Energy Efficiency in Buildings, Asilomar, CA 1994.

The 13 flat usage units (8% of the sample) used about 58% more than rated on average. This finding is remarkably similar to that by Pratt and Miller (footnote 2, Section 2) – they found that 13.5% of the refrigerators in New York public housing had a "high duty cycle" and that these units used an average of 57% more than rated. These two values imply that units operating according to their rated usage should have a duty cycle of about 63% (1/1.58). The flat usage units posed a particular challenge to the modeling of usage.

Figure 8 shows the distribution of refrigerator usage as a percentage of rated usage for the existing units and *Figure 9* shows new units. Flat usage units are shown in lighter gray. Although the mean ratio for existing units is 111%, many units use less than the rated value. The figure also shows that the majority of units using more than 160% of rated usage have flat usage. The variability of usage among new units is smaller and clearly shows most units using less than rated.

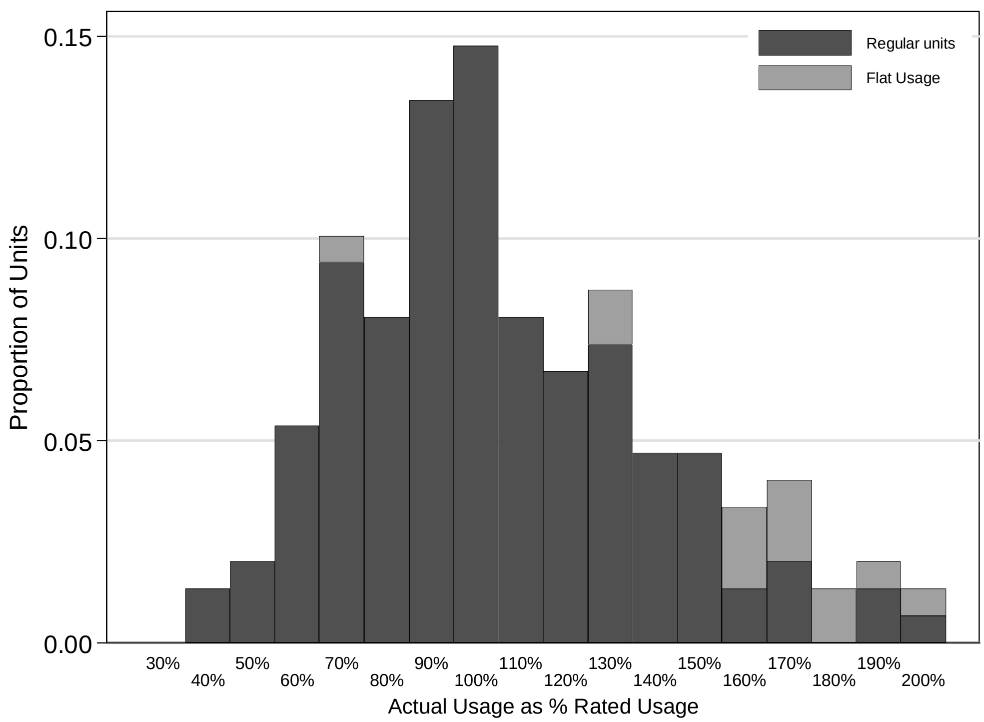


Figure 8. Distribution of Existing Refrigerator Usage: % of rated

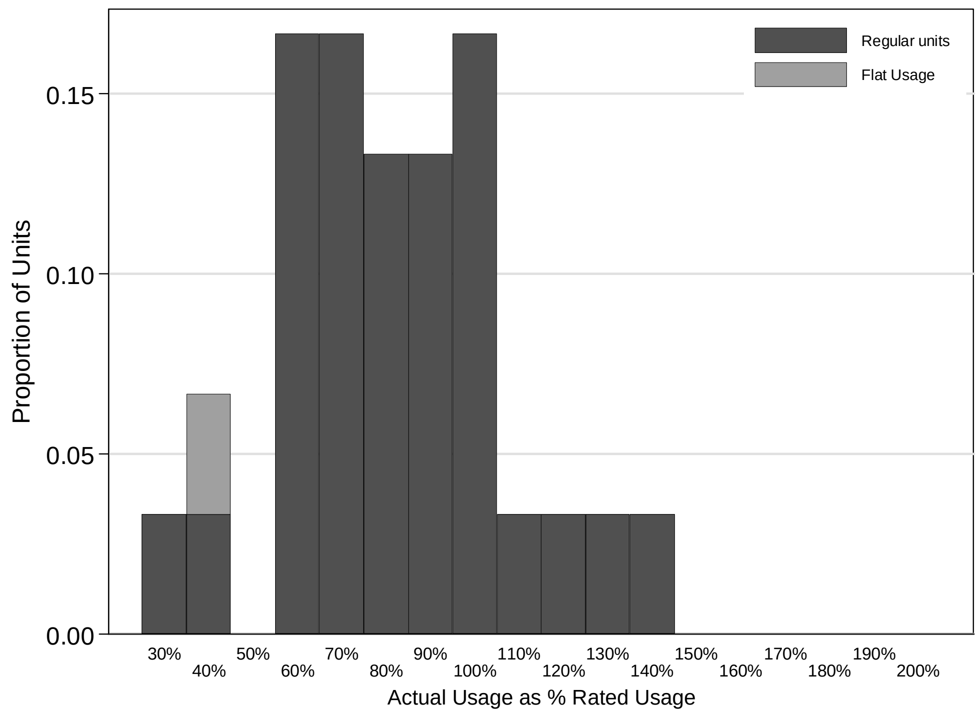


Figure 9. Distribution of New Refrigerator Usage: % of rated

Figure 10 shows the distribution of usage compared to rated values for the 29 existing units located in basements ("heated" or not). The usage of basement units is much more variable than living space units. Many units use much less than the rated amount while others, especially those with flat usage, use much more.

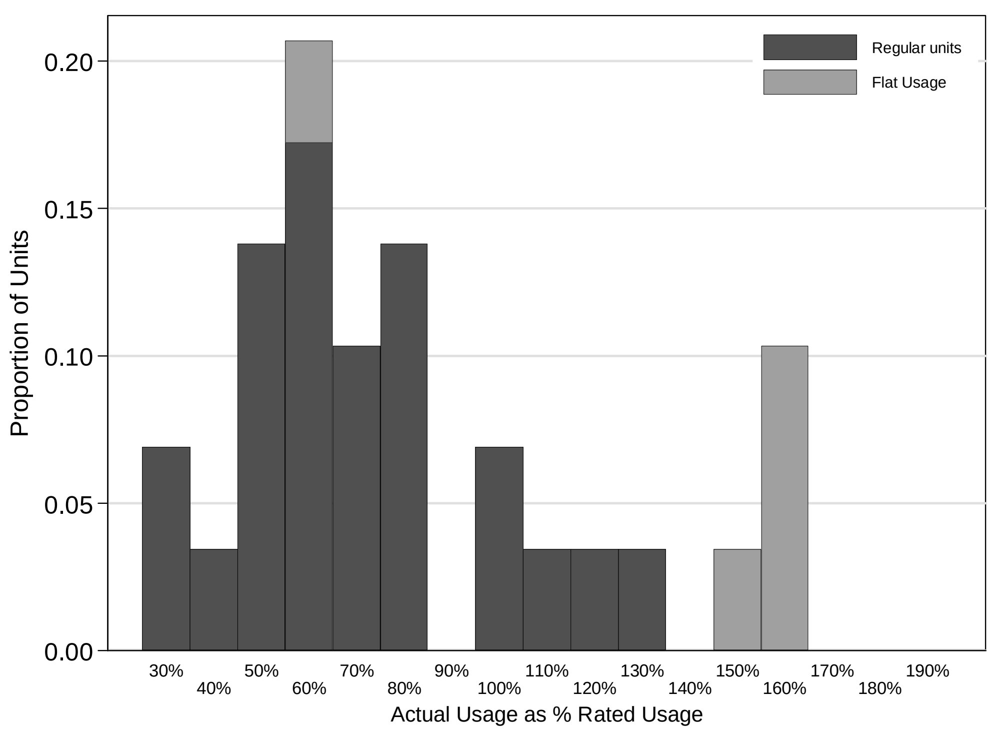


Figure 10. Distribution of Existing Basement Refrigerator Usage: % of rated

8. Modeling Refrigerator Usage and New Audit Development

A major project objective is to develop a "best" refrigerator audit approach, preferably one based on adjusted label-rated usage to keep implementation costs low. We planned to develop a regression model that explains variations in actual usage based on rated usage and other "auditable" characteristics such as refrigerator age, door style, features, and settings. The large number of potential explanatory variables and the potential problem of over-fitting to idiosyncrasies in the data required a careful approach.

We began the analysis by examining the likely nature of the relationship. We decided to express the true annual usage of each unit as a percentage of its rated usage and then model this ratio. This specification allows the candidate variables to affect usage on a percentage basis – for example, usage increases by X% per year of age or usage increases by Y% if an icemaker is present -- rather than a kWh basis. We used the usage at 70°F rather than at the estimated actual indoor temperature because it makes sense to perform a separate temperature adjustment as part of an audit strategy. One wouldn't expect the presence of an icemaker or the condition of the door seal to have very different percentage impacts at 65° versus 70°F.

We focused the analysis on the 149 existing units where we had usage results and known label-rated usage. We eliminated the one unit with shifting usage as an outlier and then proceeded to develop models with and without including the 13 units that had flat usage.

Flat Usage Effects

Given the patterns shown in the prior section, one could argue that flat usage units are outliers that will skew the modeling. However, flat usage units are real and excluding them can create bias. A statistically optimal approach would employ some type of two stage modeling process where the likelihood of flat usage is modeled directly and then the usage is estimated separately for flat usage vs. normal usage sites (perhaps through Heckman selection type models or endogenous switching models). The two part approach was attempted by Pratt and Miller in the New York study, but their technique ended up just increasing every unit's estimated usage by assuming each unit has a 13.5% chance of having flat usage equal to 157% of rated usage. We attempted to model the occurrence of flat usage but realized that having only 13 sites provided too little data to develop reliable results. We did find several factors that appear to be related to flat usage:

- flat usage occurred in 18% of the units that were purchased or acquired as used refrigerators (not including units that came with a house) compared to 6% in units that were not gotten used;
- flat usage occurred in 23% of the units where the door seal had noticeable gaps compared to 7% of the units with good door seals (and very poor door seals were probably directly responsible for several flat usage sites);
- flat usage occurred in 21% of the units with a power factor below .76 but only 1% of the units with higher power factors.

The strongest predictor of flat usage is power factor – only one unit with flat usage had a power factor above .76 while the remaining 12 units had low power factors. Secondary refrigerators and units outside the living space were also more likely to have flat usage as were units with through-the-door icemakers. Although we found many factors associated with flat usage, there were perhaps too many factors, given the small number of actual flat usage units, to be able to develop a reliable model – it would be highly questionable to model the occurrence of flat usage in 13 units using a model with 8 predictors.

Modeling Results

The problem of flat usage led us to examine a range of models fit with and without flat usage sites and examine how the predictors were affected by the model specification. We also employed some instrumental variables regression models where we treated flat usage as endogenous. We also examined models excluding units located outside the main living space. We performed this analysis exploring approximately two dozen potential predictors of refrigerator usage. Throughout the model building process we sought to develop a model that was sensible in both the practical and statistical sense. We also recognized that a simpler model would be easier to implement in the field while a complex model may be based more on idiosyncrasies of the data than true phenomena. Table 10 summarizes the regression modeling results for 9 of the main models fit.

Table 10. Refrigerator Usage Regression Modeling Results

Modeling Approach: dependent variable = actual usage as % of label rated									
Sample:	No Flat Use Units		All Units				Living Space Only		
Estimation:	OLS	Robust	OLS	Robust	IV	IV	OLS	OLS	IV
Model Results:									
# Occupants	0.055***	0.053***	0.047**	0.053**	0.051***	0.055***	0.043*	0.048**	0.043*
Ice TTD	0.104	0.121	0.150^{+}	0.165^{+}	0.113	0.079	0.151+		0.086
Anti-Sweat On	0.149**	0.121*	0.174**	0.161**	0.163***	0.148**	0.218***	0.215***	0.187***
Door Seal Poor	0.129	0.157^{+}	0.242**	0.252**	0.173^{+}		0.273**	0.268**	0.167
Bought Used	0.186**	0.178**	0.270***	0.262***	0.217**	0.188**	0.239**	0.242**	0.178*
Flat Use					0.369^{+}	0.546**			0.469
Side-by-Side								0.130^{+}	
Constant	0.811***	0.799***	0.835***	0.801***	0.819***	0.825***	0.828***	0.808***	0.839***
Model R^2	0.182	0.162	0.206	0.205	0.373	0.379	0.270	0.273	0.375
N	135	135	148	148	148	148	119	119	119

Notes:
asterisks indicate estimated statistical significance: $^{+}$ p<.10, *p<.05, ** p<.01, *** p<.001
OLS = ordinary least squares regression model
Robust= robust regression model down-weighting outliers
IV = instrumental variables regression treating Flat Use as endogenous with instruments including power factor (<.76) and, for second model, door seal condition

The models show a fairly consistent pattern of results. When flat usage units are excluded from the models, the impacts of several variables such as door seal condition, got used, and ice through-the-door decline, reflecting that some of their impact is related to flat usage. These impacts are also reduced when flat usage is included as an endogenous predictor through instrumental variable modeling. Impacts appear somewhat larger when only living space units are included. Our analysis led to the following conclusions:

- usage increases by about 5% per **occupant** *using the refrigerator as their primary refrigerator*, secondary units didn't show an occupancy effect;
- usage is about 15% to 20% greater in units where the **anti-sweat heater switch** is on (mostly reflecting the incremental usage associated with the heaters, but also reflecting a characteristic of units that have such switches) ;
- usage is about 15% to 25% greater in units where the **door seal has noticeable gaps**, the high end of this range reflects its relationship to flat usage units, while the low end is when flat usage units are excluded;
- usage is about 20% to 25% greater in units that were **bought or acquired used**, with the higher end of the range reflecting the flat usage effect;
- usage is about 10% to 15% greater in units with a **through-the-door icemaker** (note that the rated usage test procedure disconnects the icemaker for the test);
- neither the **refrigerator's age or door style had any discernible impact** on how usage varied from the rating, but the program pre-screening that eliminated most newer units prior to the audit makes the lack of an age effect less surprising and the inclusion of ice through-the-door in the model reflected but proved more powerful than the side-by-side effect.

The estimated impacts of occupancy, ant-sweat heater switch, and ice makers are consistent with results from the California study (see Proctor et al, 1994 in footnote 6) . We used the modeling results to develop several variations of a new adjusted rated usage audit approach. We found that if flat usage sites are included in the analysis then decision making accuracy declines as many units without flat usage are overestimated – an inevitable result of least squares regression which avoids large errors at the cost of many smaller errors. This finding led us to select impact estimates for the audit method from the lower end of the range for factors associated with flat usage and then adjust the base level usage upward.

Unexpectedly, **we were unable to develop any rated usage based approach that performed well for units located outside the main living space** – including units in basements ("heated" or unheated) and garages. In fact, none of the model predictors were statistically significant when the model was fit using data only from the 33 units not in a kitchen and several of the estimated effects changed sign! Based on this finding, we concluded that we could not develop an adjusted rated usage approach that works well on refrigerators outside the main living space. From a programmatic perspective, this exclusion is somewhat fortuitous since units in basements and garages are typically the easiest to meter because of easier access to the plug and less worry about damaging floors.

The variability of the estimated effects within and between various statistical models and the need to balance the impact of flat usage sites on the new audit method led us to not select one particular regression model as the new audit procedure (with all of it's coefficients listed to 5 decimal places…), but instead to select values that strike a balance between the statistical models and practical performance. The final model that was developed from this effort can be implemented as a simple scoring procedure. Because the model is only applicable to units in the living space, the simplified version of the indoor temperature model (described previously) can be incorporated into it with minimal performance penalty:

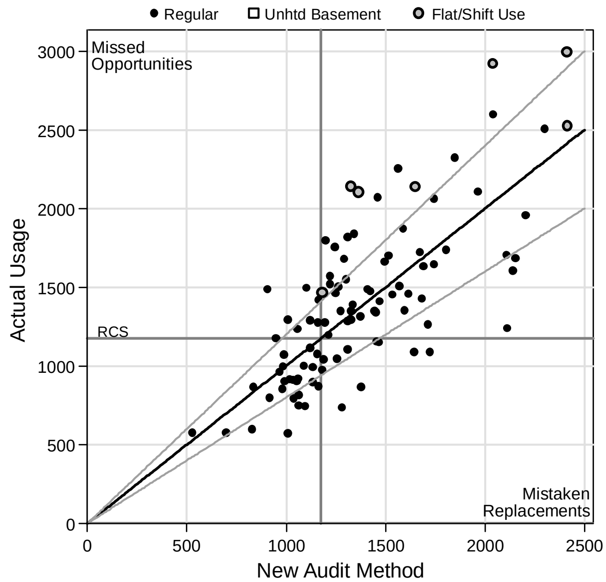
New Adjusted Rated Usage Refrigerator Auditing Procedure: for refrigerators in the living space

Refrigerator Score = 85 + 5*(# occupants if primary fridge) + 20*(anti-sweat on) + 20*(bought used) + 15*(door seal noticeable gaps) + 15*(ice through-door) + 5*(high winter avg. T-stat set point) – 5*(low winter T-stat set point)

Annual Usage = Rated Usage * Score/ 100

The scoring system is quite simple and appears to work well for refrigerators in the living space. High winter thermostat settings are defined as in the low 70's and low settings are defined as in the low to mid 60's (both should be an average that includes any setbacks). For spaces with extremely high or low thermostat settings, another few points can be added or subtracted.

The figures on the following page show the relationship between actual and predicted usage for the new audit method. The left graph shows the results for units in the living space and the right graph shows units in basements (heated or unheated) or other spaces. Data points that represent unheated basements or flat or shifting usage are shown with different symbols (note that one site can be both flat/shifting and basement). The graphs are each divided into four quadrants at the 1,175 kWh RCS usage threshold. Points in lower right are "Mistaken Replacements" (units that shouldn't have qualified but did) and points in the upper left are "Missed Opportunities" (units that should have qualified but didn't) .



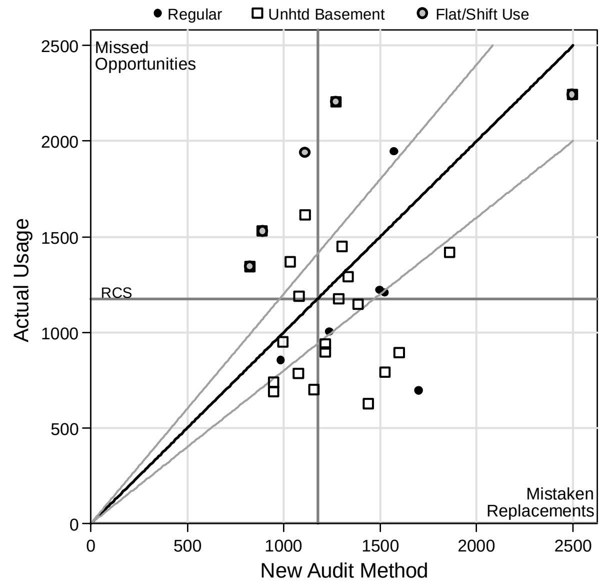


Figure 11. New Audit Method – Living Space

Figure 12. New Audit Method - Basements

Notes: Diagonal solid line=perfect agreement. Dashed lines=±20% error bounds. Vertical and horizontal lines at 1175 kWh indicate RCS threshold in Massachusetts.

The graphs show that the new audit approach does a fairly good job of predicting usage in living space units. The flat usage sites are generally under-predicted, but still tend to be properly selected for replacement. There are relatively few missed opportunities or mistaken replacements. The graph on the right illustrates that there is virtually no relationship between the predicted usage and the actual usage for units outside the living space. The reasons for this scatter may include:

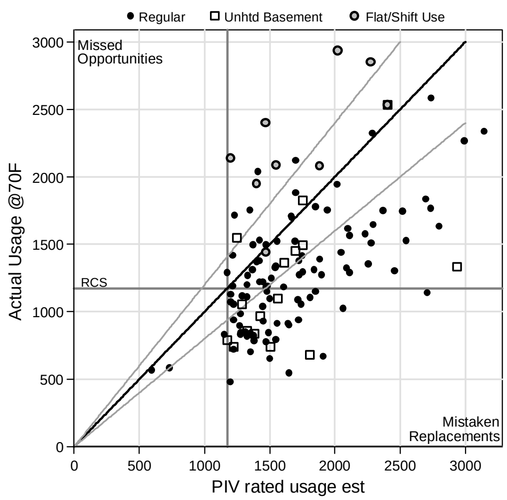
- a wider variation in occupant loads (some units are almost never opened);
- a wider variability in the quality of door seals and condition of the unit and refrigeration system may be found in basement units, particularly units that are malfunctioning; and,
- average annual basement temperatures are more difficult to predict (and ignored in the new audit method as shown).

Note on Occupancy Effects: The estimated impact of occupancy is quite modest – just 5% per occupant. In reality, occupancy likely has a larger effect, perhaps averaging about 30% of total loads, but a base level of occupancy is undoubtedly contained within the base score of 85 and the impact of occupancy is almost certainly not linear – the first occupant adds the most load and each additional occupant adds progressively less and less.

9. Performance Comparison of Auditing Approaches

The report has gotten ahead of itself -- the first step in meeting the objectives of the project was to assess the accuracy of the PIV's auditing methods before developing a new and improved approach. However, the performance comparison of different approaches is easier to follow if the new audit approach is developed prior to the comparisons.

We began our assessment of the accuracy of the current PIV audit usage estimates by graphing estimated usage against "actual" usage. The graphs below show the PIV audit-predicted usage plotted against the "true" usage from the data analysis. The left graph shows the sites that had an adjusted rated usage audit and the right graph shows the sites that were audited with short-term metering.



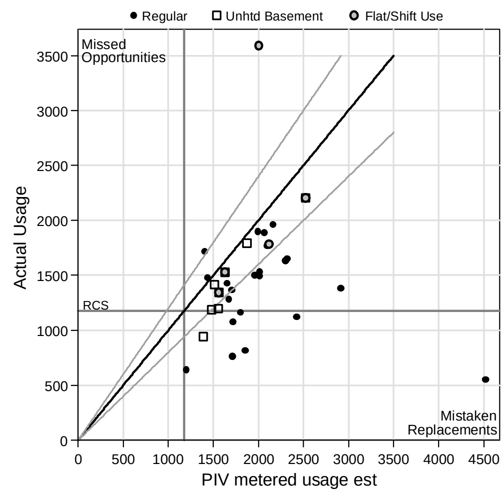


Figure 13. Audit Accuracy: PIV rated usage approach **Figure 14. Audit Accuracy: PIV metering approach**

Notes: Diagonal solid line=perfect agreement. Dashed lines=±20% error bounds. Vertical and horizontal lines at 1175 kWh indicate RCS threshold in Massachusetts. .

Figure 13 shows that, for the PIVs adjusted rated usage methods:

- the audits over-estimated usage for the vast majority of units – 100 of the 123 units (81%) had their usage overestimated and the overestimate averaged 24% (1681 vs. 1354);
- a considerable proportion of units that did use more than predicted had flat or shifting usage;
- few units had their usage predicted within 20% of the true value;
- the PIV's qualified many units for rebates that should not have qualified (mistaken replacements);
- only three units did not qualify for replacement and all three were truly not qualified, leaving no lost opportunities;
- basement units show little if any relationship between the audit predictions and actual usage.

Figure 14 shows that the short-term metered units:

- had a somewhat better relationship between actual and predicted usage -- 14 of the 33 units (42%) had their usage predicted within 20% -- but overprediction is evident here as well and the average bias was a 32% overestimate (influenced strongly by one large outlier in the lower right);
- all 33 units were qualified for replacement by the audit, but there were 8 mistaken replacements;

- unlike the rated usage approach, units with flat usage were generally predicted at least as well as properly functioning units (with the exception of one large outlier).

The two large outliers on the metered usage graph (in the lower right and in the middle of the top) represent a unit in a garage where the temperature was in the mid 30s during the audit, leading to a large but mostly erroneous adjustment to a default 68°F, and the one unit that apparently malfunctioned and had a large upward shift in usage during the metering.

In the description of the PIV audit methods, we noted that CSG's adjusted rated usage algorithm inadvertently incorporated a 79°F default temperature. This problem was rectified by CSG midway through the project. We created adjusted versions of CSG's usage estimates for the units audited prior to the correction. In addition, neither of the adjusted rated usage methods included an adjustment for the estimated site-specific annual temperature. In order to test the other aspects of the PIV's adjusted rated usage methods, we created a temperature adjusted version of each audit using a default 2.5%/°F temperature slope to adjust each site to our estimate of its annual indoor temperature. These new estimates reflect how the current approaches might perform if basic temperature corrections were made. We performed a similar adjustment for the short-term metered sites. These adjusted values are explored in greater detail in the next section.

Only 33 PIV audits were performed using a short-term metering approach, but the project produced metered data on 156 existing units. We used this data to assess the accuracy of potential short-term metering-based predictions by calculating annual usage estimates for each metered data point for each site, adjusting the usage using the default 2.5% temperature slope. We developed audit predictions based on metering period lengths of 0.5, 1, 1.5, and 2 hours (actually 34, 68, 102, and 136 minute metering) examining each period throughout the full two to four weeks of monitoring. In other words, each short period within the full monitoring period were treated as if it were a single audit. We restricted the analysis of these metered usage estimates to potential work hours (7 AM through 7PM weekdays) to reflect likely results from a program audit. These values represent an ideal version of what short-term metering would find since they are undisturbed by the audit process, have the elapsed time properly recorded, and include a consistent temperature adjustment.

In practice, we found that the PIV short-term metered audits did not perform as well as these ideal values. In an actual short-term metered audit the refrigerator door is typically opened to read the name plate, the unit is unplugged and plugged back in (causing the compressor to stop for a few minutes), and the occupants may be told not to open the door or may not act "normally". These differences may reduce the accuracy of a real world short-term metered audit somewhat compared to these measurements. In addition, it appears that auditors may round off the elapsed metering time and the PIVs may employ inaccurate or inconsistent temperature corrections and make other inappropriate calculation adjustments. These problems with the PIV metering will need to be addressed to achieve the performance of the ideal metered results.

In addition to assessing the PIV audits and the all possible ideal short-term metering results, we assessed some alternative audit methods including simply using 111% of the label-rated usage (without adjustments for age or any other factor), where the 111% is the average actual usage as a percent of the average rated usage for the units in this study. We also assessed the performance of the new proposed audit approach.

Accuracy of Audit Approaches

We summarized the accuracy of each of the auditing approaches using a variety of metrics: the mean error (i.e., systematic bias between the audit estimate and the true usage), the mean absolute error (i.e., the average size of the discrepancy between the usage and the audit estimate), and the root mean square error (a statistical measure which gives greater weight to large discrepancies). These values provide insight into the direction and size of the discrepancies but don't show how often the audit estimate is close or far from the true value. To summarize the distribution of errors, we classified the accuracy of audit usage estimates as very good (within 10% of true value), OK (within 10%-20%), not good (20%-40% off) and poor (more than 40% off). Table 11 shows the results of the analysis.

Table 11. Accuracy of Audit-Predicted Usage (compared to best estimate of usage from data logging)

		Mean Usage		Mean Errors (% of true use)			% Cases with Accuracy*		
Audit Method	# units	Audit	Actual	Bias	Absolute	RMSE	V. Good	OK	Poor
PIV Audits									
Rated: CSG	72	1702	1319	29%	35%	44%	22%	11%	42%
Rated: HDMC	51	1653	1403	18%	34%	41%	22%	8%	43%
Rated: All	123	1681	1354	24%	34%	43%	22%	10%	42%
PIV Metered	33	1973	1493	32%	40%	62%	18%	24%	36%
T-Adjusted Audits									
Rated: CSG adj.	72	1509	1319	14%	26%	31%	18%	25%	24%
Rated: HDMC adj.	51	1684	1403	20%	35%	42%	20%	14%	41%
Rated: All adj.	123	1581	1354	17%	30%	36%	19%	20%	31%
Metered: adj.	33	2023	1493	35%	47%	82%	27%	12%	33%
Potential New Audits: for comparative purposes, results only shown for units with PIV rated usage estimates									
111% Label-Rated Usage	123	1382	1354	2%	24%	32%	26%	27%	21%
New Audit Rated Usage	123	1368	1354	1%	21%	27%	29%	27%	11%
Ideal Metering	# "audits"								
Meter: ½ hr (105 units)	29,585	1373	1360	1%	30%	51%	24%	21%	27%
Meter: 1 hr (143 units)	22,614	1385	1382	0%	22%	40%	33%	27%	14%
Meter: 1½ hr (105 units)	9,906	1383	1360	2%	20%	46%	35%	28%	12%
Meter: 2 hr (145 units)	11,709	1427	1418	1%	17%	28%	39%	29%	9%
***** Accuracy categories are defined as: Very Good <10% error; OK 10%-20% ; Poor >40% error									

The table provides numbers to reinforce what the prior graphs showed about the current PIV audits:

- the PIV approaches are not very accurate with a 34% average error (mean absolute % error) for the rated usage approaches and 40% for the metered approach (influenced by an outlier).
- as expected, temperature adjustments had a substantial effect on CSG's audit accuracy, but still just 39% of the corrected audits were within 20% of the true value for both the metering and adjusted rated usage approaches (the results for the retroactively modified CSG usage estimates were essentially the same as the fully temperature adjusted version shown in the table) .

The performance of the simple 111% of label-rated usage approach is quite good. This simple approach is more accurate than any of the current PIV methods. **It appears that the adjustments to the label-rated value that the PIVs make in their audit approaches are worse than just using a simple adjustment of 111%.** Further analysis found that the PIV's methods are worse than no adjustment at all.

The proposed new adjusted rated usage audit performs comparably to the 111% method in most ways but makes fewer large errors of more than 40%. This advantage is due to factors in the method that help properly predict that some units use substantially more than 111% of the label-rated value.

The all possible ideal short-term metering shows that ½ hour metering is not as accurate as the 111% method especially in terms of large errors ("poor" accuracy column). Ideal metering is slightly more accurate than any of the rated usage methods at one hour, with 60% of all units predicted within 20% (a comparison made on just the same set of 123 units shows slightly better metering performance than the full sample results shown here). The improvement in metering accuracy is substantial going from ½ to one hour, but more modest when increasing the metering time to two hours.

Even ideal metering does produce some large errors, especially with shorter intervals, due to defrost cycles and large (and/or hot) food loadings. The defrost problem could be addressed by screening for defrost cycles during the metering by either monitoring peak power draw (defrost cycles typically draw 400 or more watts) or freezer temperatures (which go above 32°F during the defrost). The food loading problem is also easily avoided during short-term metering, but if occupants are not allowed to use the refrigerator during metering a small downward bias would be expected. The relatively large errors shown for PIV metering compared to ideal metering are influence by an outlier but are also based on a sample where the ideal metering did not do as well. When restricted to the same 33 units, one hour ideal metering had an average absolute error of 27% vs. 22% for the entire sample.

Decision Accuracy

Although accuracy is certainly a worthwhile goal for an auditing approach, a more practical issue is how each approach affects the decisions made. We assessed the accuracy of each audit's decision making by tabulating the four possible decision outcomes for each using the 1,175 kWh RCS threshold:

- correct replacement: unit used more than 1,175 kWh/yr and audit correctly identified it;
- lost opportunity: unit used more than 1,175 kWh/yr but audit did not qualify it;
- mistaken replacement: unit used less than 1,175 kWh/yr but audit qualified it; and
- correctly kept: unit used less than 1,175 kWh and audit agreed.

Figure 15 depicts the decisions made by the PIV's rated usage audits. Each bar represents the actual annual usage of a refrigerator compared to the threshold of 1175 kWh. Bars above zero truly qualify for replacement and those below zero do not qualify. Units selected for replacement by the PIVs are shown in black and units not selected are light gray. If decisions were made perfectly, all bars above zero would be black and all bars below zero would be gray. Black bars below zero are mistaken replacements while gray bars above zero are lost opportunities.

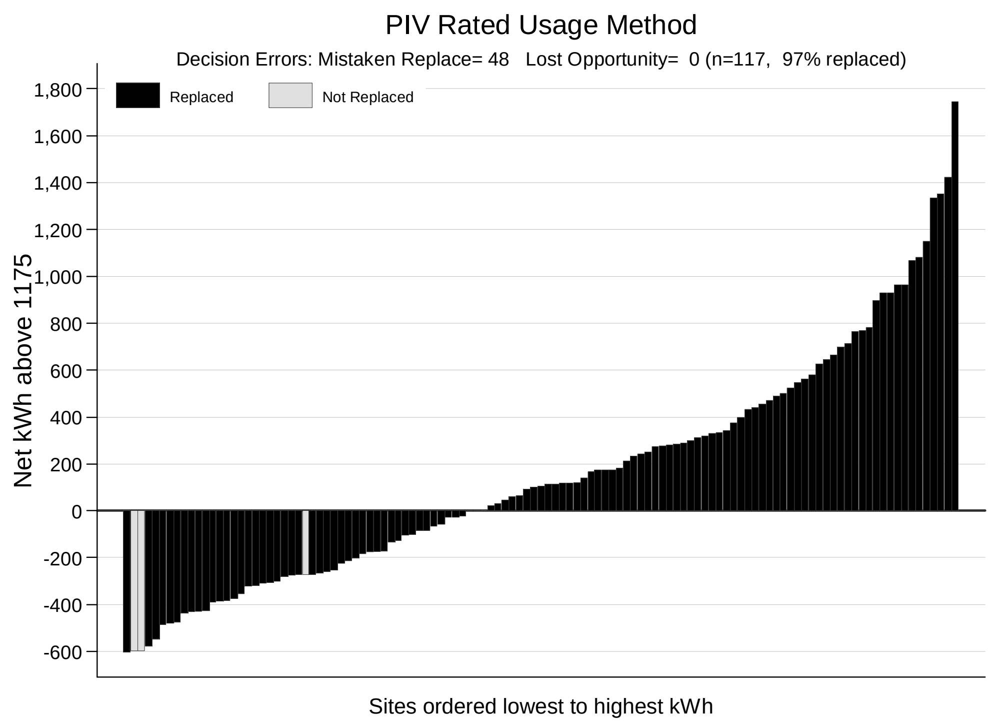


Figure 15. Decision Making: PIV rated usage audits

The figure shows that the PIV audits qualified all but 3 of the units (largely a result of the sample selection strategy as only two of the five programs allowed the PIVs to select units that they did not consider qualified). The problem with the audits is that just 56% of the units actually used more than 1,175 kWh/yr, producing a large number of mistaken replacements. The PIV metered audits (not shown) performed somewhat better -- all 33 audited with metering were deemed qualified for replacement but only 76% truly qualified, yielding 8 mistaken replacement in 33 units.

The decision making of two alternative rated usage audit approaches is shown in *Figure 16*.

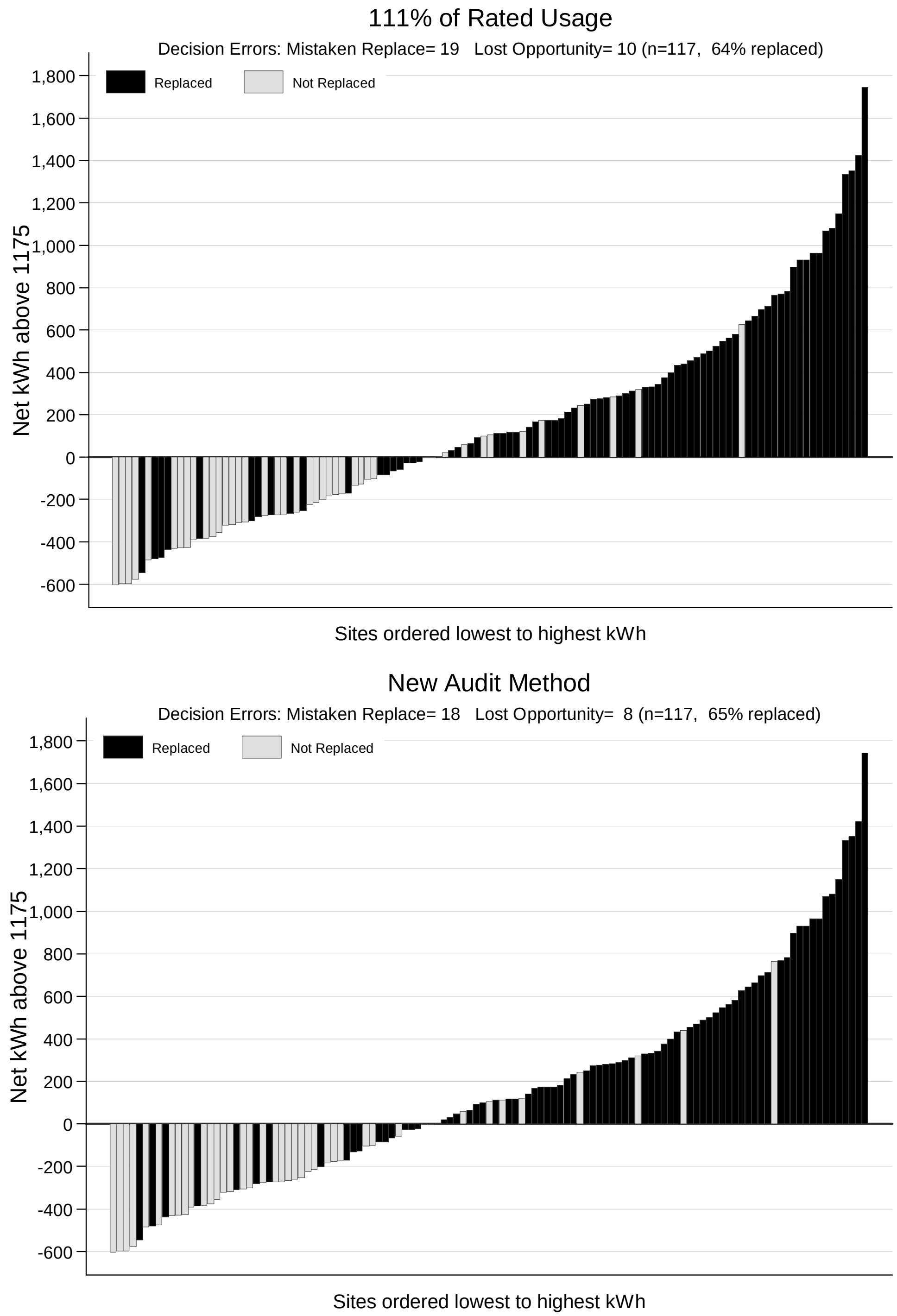
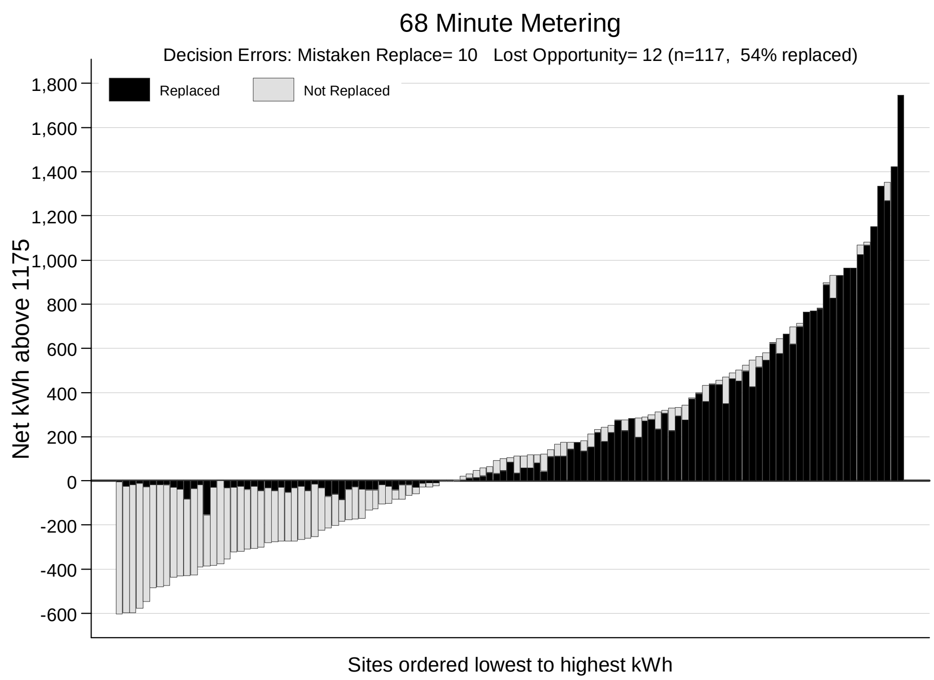


Figure 16. Decision Making of Alternative Rated Usage Audits

The simple 111% of label-rated usage and the new proposed rated usage approach both produce far fewer mistaken replacements (fewer black bars below zero), but also create some lost opportunities (gray bars above zero). The new audit approach makes 3 fewer mistakes than the simple 111% method. Each of these methods has one unit that was a relatively large missed opportunity with usage more than 500 kWh/yr above the 1,175 kWh threshold (the tallest light gray bar).

Figure 17 shows the decision making for one hour (68 minute) and two hour (134 minute) metering. For these methods, each bar can have both black and gray fills representing the proportion of all potential metering periods for each unit that indicate usage above and below 1175 kWh/yr, respectively.



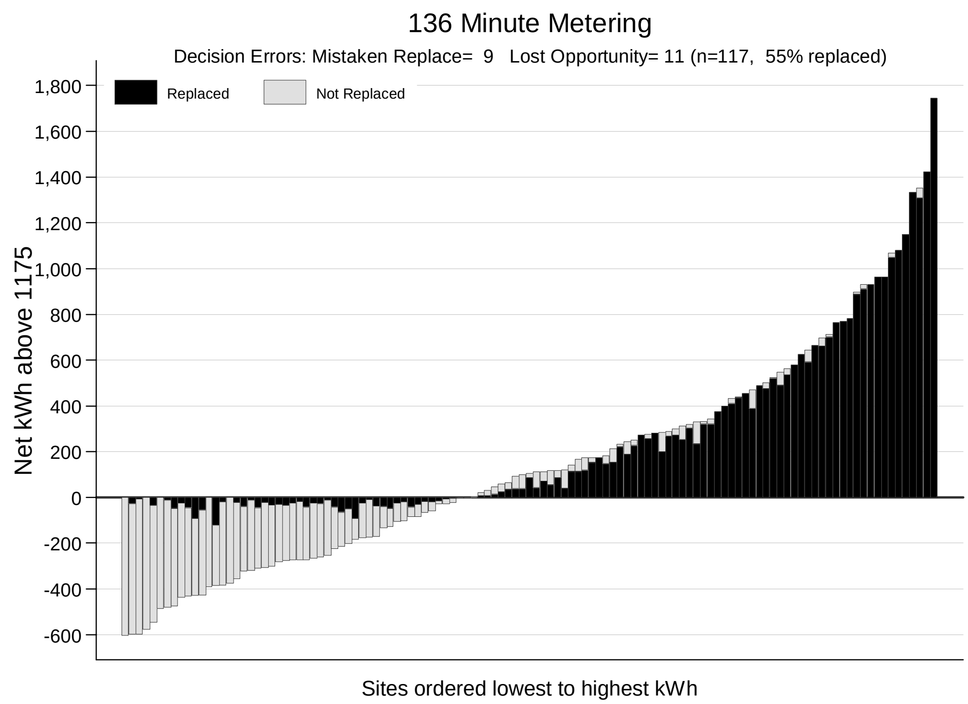


Figure 17. Decision Making: Short-term Metering: Ideal Metered Audits

It is clear that both one hour and two hour metering classify nearly all units correctly the vast majority of the time. The greatest variability in the replacement decision occurs in units close to the 1175 threshold, indicating that the mistakes that are made are usually quite small in terms of kWh impact on program savings. Therefore, although the metering methods show more "lost opportunities" than the rated usage methods, these opportunities are concentrated among the lower usage units (the total number of mistaken replacements and lost opportunities for these units is the sum of the proportion of the times that each unit is misclassified, rounded to the nearest integer). Among the higher use units, the bars are virtually pure black, indicating the metered audits would always make the right decision. A close examination of the two metering graphs shows that two hour metering does somewhat better than one hour metering, but the difference appears modest.

The small fraction of black in nearly every bar below zero reflects the impact of defrost cycles and probably large food loadings that will lead to some mistaken replacements from short-term metering. An audit protocol to avoid defrost cycles, such as monitoring peak wattage or freezer temperatures that indicate a defrost cycle mentioned previously, would eliminate nearly all of these mistaken replacements. Even with the defrost cycles, the rate of mistaken replacements is only half the rate found for the rated usage methods. One reason why defrost cycles don't have a larger negative impact is that a little more than half of the units audited qualify for replacement, so defrost errors on those units don't affect the decision. Therefore, decision errors from defrost cycles are likely to be no more than 4% overall since they affect less than half the decisions for less than 8% of the time.

The graphs depicting audit decision making also expose a major limitation in just tabulating the number of mistakes and lost opportunities. A missed opportunity is not much of a loss if the unit only qualified marginally and a mistaken replacement does not impose much net loss to the program if a unit just barely missed qualifying. In terms of the graphs, the small bars are not as important as the larger bars to overall program savings, but each bar counts the same in tabulating decision accuracy. This realization led us to examine how each auditing approach affects the overall program savings and costs.

Value of Audit Approaches

We approached the issue of audit value by calculating the net benefit that would be derived from replacing each refrigerator and then calculating the average net benefit per audit for each auditing approach based on the replacements that each would select. This analysis does not include the cost of performing each audit approach, but instead allows one to compare the net benefits that may be derived from each audit approach to assess what incremental cost may be justified for each. The net benefits were calculated as the value of the energy savings minus the assignable program cost using these formulas:

Net Benefit = \$0.31 * (TrueUsage - .877 * 510) – (.353 * \$1043) if volume >= 19 ft^3

Net Benefit = \$0.31 * (TrueUsage - .877 * 437) – (.353 * \$629) if volume < 19 ft^3

Sources:

1. NSTAR provided program planning assumption document developed by VEIC in 2001 that indicated the present value of lifetime energy savings equals \$0.31 per annual kWh (based on a five year savings lifetime) and that program costs should include .353 times the total refrigerator cost (the remaining .647 is considered enhanced customer value and O&M savings)
2. Program tracking system data from NSTAR, MECO and WMECO showed a sharp divide in new unit costs at 19 ft^3 – average costs were \$1043 for units >= 19 ft^3 and \$629 for smaller units. Rated usage averaged 510 kWh for larger units and 437 kWh for smaller units
3. Analysis of the 30 new replacement units found that usage averaged 0.877 of rated usage for new units
4. TrueUsage is the estimate of the existing unit's annual usage at the site specific temperature from the detailed meter data analysis.

The assumptions used to develop these formulas may not apply to other utilities or time frames. The calculation required a two tiered approach because of the large jump in average new refrigerator costs for units 19 ft^3 and larger. We assumed that the size of the new unit will equal the size of the existing unit.

This assumption ignores the fact that new units averaged almost 2 cubic feet larger than existing units and customers often purchased a new unit with a different door style and added features. One could argue that this upgrading reflects added customer value and shouldn't affect the costs assigned to the program.

We summarized the net benefits of each potential audit strategy by calculating the average net benefits achieved per audit if that audit approach had been used to select which units are replaced. For the many results of the all possible ideal metering, we calculated the proportion of the audit results that would qualify the unit and multiplied that proportion by the net benefit of the replacement.

For comparative purposes, we also calculated the maximum possible net benefit per audit if all decisions had been made with "perfect" knowledge (i.e., based on the best estimate of annual from our site-specific modeling) and the RCS incentive threshold of 1,175 kWh were the usage threshold for replacement.

An example of this approach may provide some clarification. If a program audited three refrigerators and the calculations indicated that the net benefits of replacement were $300, $75, and -$100 for the units respectively, then the maximum possible net benefit per audit would be $125 -- the two units with positive benefits would be replaced, yielding $375 total benefits, divided by the three audits equals $125 per audit. If an audit strategy selected all three for replacement, the net benefits per audit would drop to ($300+$75-$100)/3 = $93. If another audit strategy only identified the one unit with the greatest benefits, the net benefits per audit would be $300/3 = $100. This latter audit strategy would be $7 more cost effective per audit than the one that replaced all three units, but it would miss a cost-effective replacement opportunity.

One interesting and complicating finding from the net benefit analysis is that 20 of the 97 units that should qualify for replacement based on the 1,175 kWh threshold yield negative net benefits – they are not cost-effective replacements. These units are typically larger units with higher assumed replacement costs that have usage levels only moderately above 1,175. In addition, 5 of the units that should not qualify actually would produce small positive net benefits. **The simplification of the RCS program guideline to a single usage threshold has led to a program design that does not necessarily select cost-effective replacements.** In comparing the cost-effectiveness of different audits designed to select units based on the RCS threshold, audits that make programmatically "correct" decisions may be penalized if the selected unit is not cost-effective. This may affect comparisons between approaches but can only be properly rectified by changes in program rules or cost-effectiveness assumptions.

The net benefit calculations also revealed the skewed distribution of benefits. A few refrigerators with very high usage provide a substantial fraction of the potential net benefits while many units are marginal contributors where wrong decisions about replacement would have a minor impact on cost-effectiveness (the small bars on the decision making graphs). This distribution implies that improving the usage prediction accuracy of an audit approach by a few percent is unlikely to have much impact on cost-effectiveness – if a few percent change in estimated usage affects the program decision for a unit, then the net benefit or net loss from the replacement decision will be fairly small either way. **The key to maximizing cost-effectiveness is to properly identify all very high use units without diluting the net savings too much by replacing many low use units**.

We performed the net benefit calculations for each audit approach and calculated the maximum potential net benefit if site specific cost effectiveness screening were used instead of the 1,175 kWh threshold.

In addition to the audit methods described thus far, we analyzed two replacement rules that have been used or proposed elsewhere in the nation – a rule to replace all units built prior to 1990 and a rule to replace all units that have 1970's colors such as harvest gold. We provide results for the current audit approaches for all units where they apply, results for varying length metering periods for matched sets of cases (to assess the value of extending metering), results for all 117 cases that have PIV rated usage audits, results for all 95 of those units in the living space, and results for the 115 living space units regardless of PIV audit method. Table 12 summarizes the results of this analysis and tabulates the decision making accuracy for each audit approach.

Table 12. Decision Making Accuracy and Cost-Effectiveness of Audit Approaches

		Decision Making Accuracy						
		Qualified Units (true use >1175 kWh)		Unqualified Units (true use <1175 kWh)			Net Benefit $/audit	
Audit Method	N	Replaced: Correct	Kept: Lost Opportunity	Replaced: Mistake	Kept: Correct	% Wrong Decisions	**Actual**	Program Possible
PIV Audits: assess the performance of the current PIV audits								
PIV Rated: CSG	72	37	0	32	3	44%	**-5**	49
PIV Rated: HDMC	51	35	0	16	0	31%	**39**	68
PIV Rated: All	123	72	0	48	3	39%	**13**	57
PIV Metered	33	25	0	8	0	24%	**72**	93
T-Adjusted PIV Audit: assess the value from temperature-adjusted version of the current PIV approaches								
PIV Rated: CSG adj.	72	35	2	25	10	38%	**7**	49
PIV Rated: HDMC adj.	51	34	1	13	3	27%	**40**	68
PIV Rated: All adj.	123	69	3	38	13	33%	**21**	57
PIV Metered: adj.	33	25	0	7	1	21%	**76**	93
All Possible Ideal Meter: compare the performance and value of different metering period lengths								
Sites w/½ hr data:								
Meter: ½ hr	112	52	16	11	33	24%	**45**	58
Meter: 1 hr	112	56	12	10	34	20%	**50**	58
Meter: 1½ hr	112	57	11	9	35	18%	**52**	58
Meter: 2 hr	112	58	10	9	35	17%	**52**	58
Sites w/ 1 hr. data:								
Meter: 1 hr	150	75	16	12	47	19%	**53**	61
Meter: 2 hr	150	77	14	12	47	17%	**55**	61
w/ Rated & Metered: compare all auditing approaches on a set of units with full data for all methods								
PIV Rated: All adj.	117	65	1	38	13	33%	**15**	52
111% Label-Rated Usage	117	56	10	19	32	25%	**31**	52
New Audit Rated Usage	117	58	8	18	33	22%	**35**	52
Meter: 1 hr	117	54	12	10	41	19%	**44**	52
Meter: 2 hr	117	55	11	9	42	17%	**46**	52
All Pre-1990	117	61	5	44	7	42%	**12**	52
Living Space Units Only: compare all auditing approaches on refrigerators in the living space								
PIV Rated: All adj.	95	57	1	28	9	31%	**23**	55
111% Label-Rated Usage	95	48	10	10	27	21%	**42**	55
New Rated Usage Audit	95	52	6	10	27	17%	**50**	55
Meter: 1 hr	95	47	11	8	29	20%	**48**	55
Meter: 2 hr	95	49	9	7	30	17%	**51**	55
All Pre-1990	95	54	4	31	6	37%	**21**	55
Living Space Units Only: not restricted to units with PIV rated usage audits, maximum sample for living space units								
111% Label-Rated Usage	115	59	14	13	29	23%	**49**	63
New Rated Usage Audit	115	65	8	12	30	17%	**60**	63
Meter: 1 hr	115	60	13	9	33	19%	**55**	63
Meter: 2 hr	115	62	11	9	33	17%	**58**	63

Notes:
- The first four columns under "Decision Making Accuracy" tabulate the four possible decision outcomes: properly replaced units, lost opportunities (should have been replaced but wasn't), mistaken replacements (shouldn't have been replaced but was), and properly kept units. The sum of the middle two columns is the number of errors, shown in percentage terms in the "% Wrong Decisions" column.
- The "Net Benefit $/audit" columns show the calculated net benefit of each auditing approach excluding the cost of the audit itself (based on NSTAR planning assumptions). The "Actual" column shows the results of the specific auditing strategy. The "Program Possible" column is calculated if all decisions were in accordance with the program rules using the true usage.

The table shows that:

- PIV rated usage audits qualified nearly every unit when only 59% (72 of 123) actually used more than the threshold. The audits achieved only $13 of a potential $57 net benefit per audit. Simple temperature adjustments only help a little.
- PIV metered audit sites captured $76 of the $93 in net benefits – this performance is largely due to the high usage and high qualifying proportion since 100% were qualified by the audits.
- Net benefits per audit increase by $5 by extending metering from ½ hour to 1 hour (from $45 to $50) and by just $2 more from extending the metering to 2 hours. The 150 units with 1 hour data confirm this finding. **The incremental value of extending metering from one hour to two hours is quite small, just about $2 more per site.**
- The section labeled “w/ Rated & Meter” shows results for the 117 units where we have both metered data and adjusted rated usage estimates allowing comparisons of many methods across the same units. Net benefits per audit increase from $15 to $31 by using 111% of label-rated usage instead of the PIV rated usage approaches. The new proposed rated usage audit would increase the net benefits by another $4 per audit to $35. Metering is superior to the new rated usage approach, but this comparison includes units outside the living space where the new method is not recommended.
- The sections labeled “Living Space Units Only” show results for the 95 units located in the living space that have data on all audit approaches including PIV rated usage, and the 115 units that have audit estimates regardless of the PIV audit approach. All audit methods perform better on units in the living space, reflecting the greater predictability of usage for these units. The new proposed rated usage audit approach does particularly well, producing net benefits comparable to the short-term metering results and even slightly better for the larger sample of 115. **The new rated usage audit method may be the most cost-effective audit approach for units in the living space**. It is more cost-effective than using 111% of the label-rated usage primarily because of fewer lost opportunities.
- Replacing all pre-1990 units performed poorly by replacing far too many units. Replacing all units with 1970’s colors (not shown) performed even worse as it qualified only about 10% of te units and missed most savings opportunities.

Other findings from the analysis include:

- The performance of methods based on rated usage is very poor when applied to the 22 units located outside the living space (results not shown in the table). In fact, **all of the rated usage based audit approaches produce negative net benefits when applied to the units in basements and garages – averaging a $20-$30 net loss per audit**. In other words, the program would be better off just ignoring all refrigerators outside the living space than trying to use one of the rated usage audit methods to assess them. In contrast, short-term metering does very well with these units, capturing $28-$29 in net benefits per audit out of a program possible $35.
- the simplified RCS program rule of offering incentives to all units using more than 1,175 kWh/yr instead of basing the decision on a cost-effectiveness calculation is **reducing potential program net benefits by about $12 per audit** – equal to about 16% of the potential benefits to be gained from refrigerator replacements in these homes (not shown in table).

The primary conclusion of the net benefit analysis is that the new proposed rated usage audit performs comparably to ideal metering for refrigerators in the living space and may be the most cost-effective audit approach for those units. Simply estimating usage at 111% of the label-rated

value also performs well. Refrigerators in basements, garages, and other unconditioned or semi-conditioned spaces must be audited using short-term metering to provide cost-effective decisions.

The results concerning length of metering are generally at odds with other published recommendations. The primary reason for this discrepancy is that most of the research and recommendations have focused on the accuracy of metering for different lengths of time without regard to how it affects the decision making in the field and the net benefits to the program. Refrigerator replacement programs tend to get a large proportion of their net benefits from replacing a relatively small fraction of units that have very high usage. The decision to replace these units is generally not affected by moderate errors because their usage is far from the replacement threshold. Therefore even very short-term metering is likely to identify them. The refrigerators where the replacement decision is affected by moderate gains in audit accuracy tend to have usage rates near the replacement threshold, making the net impact of a wrong decision fairly small.

Note: The results of the cost-effectiveness calculations and the conclusions drawn from them may not apply well to other programs, particularly programs with very different replacement thresholds. For example, a program with a usage threshold that leads to a low replacement rate may find few units with very large net benefits, making small changes in usage estimation accuracy more important. Such a program would need to do a good job at identifying very high use flat usage units and may find that metering has a larger advantage than it does here. Such a program may also find that the incremental benefits from two hour metering are larger than those here.

Savings Realization Rates

We developed estimates of how program savings may have varied from audit-predicted savings by examining how the actual usage of existing units varied from the usage estimated by the PIVs and how the actual usage of new units varied from their rated values. We performed this analysis under the assumption that every refrigerator that the PIVs qualified for incentives would actually be replaced (or that the units actually replaced would be a random sample in terms of how usage varied from PIV predictions).

The analysis found that the average unit that qualified for replacement would be predicted to save 1295 kWh/yr but the actual savings would average 988 kWh/yr, yielding a savings realization rate of 76%. The rate varied a little between programs, with realization rates averaging 74% for CSG's rated usage audits (67% prior to the change in default temperature and 86% after the change in default temperature), and 84% for DMC's rated usage audits.

We calculated realization rates if alternative audit approaches had been used to estimate the usage and savings for each unit. For living space units, realization rates were 113% for the 111% method and 114% for the new rated usage audit. Both audits would understate savings due to lower usage among new units than rated and also due to a modest underestimation of the existing units' usage.

For PIV metering, the realization rate was 72% for the 33 metered audits or 78% if an outlier is removed. After removing the outlier, the ratio of PIV metered to actual usage was 80%±8% indicating systematic over-estimation of usage. Temperature adjustments improve the ratio (to 84% ±9%), but bias remains. Some bias is due to a time of day correction employed by CSG that shouldn't be made because the effects are captured in the temperature. It may be worth exploring the metering protocols, including the potential impact of door openings, metering length, recording of elapsed time (one PIV always recorded 120 minutes), and temperature adjustments. Half of the metered audits involved an hour or less of metering. These shorter periods could suffer from greater bias due to the door opening and unplug/re-plug events at the start of metering. Although the sample is small, it appears that short-term metering as practiced in the field by the PIVs may be less accurate than the ideal "all possible meter results" used in our comparative analyses. A more consistent and rigorous approach to short-term metering may be needed.

We compared the project sample with the program population using the average projected savings. The combined program-wide tracking database of 1,572 units showed 1296 kWh/yr predicted savings – quite similar to the 1295 kWh for our sample.

10. Load Shape Analysis

We calculated load shapes by first calculating the average usage separately for each hour of the day for each site and the overall average load for each site across all hours. These values were then used to fit a ratio estimate for each hour of the day across all sites. This approach gives greater weight to high use sites than low use sites in proportion to their usage compared to approaches that average the individual site load ratios. The resulting load shape is a ratio of the hour's usage to the average hourly usage for the day. This approach allows one to apply the load shape to varying usage levels to represent different units and different operating conditions.

We performed the load shape calculations for existing units and new units separately and also calculated break-outs by day type and location for existing units.

Figure 18 shows the load shape for the pool of all existing units, with a 90% confidence interval shaded in gray. The figure shows that average loads range from about 6% below average at 4AM to about 8% above average at 6PM. *Figure 20* shows the load profile for the 30 new refrigerators. The new refrigerators' load is much more variable – the vertical scale covers a much wider range with usage dropping about 20% below the average from 2AM to 4AM and peaking at nearly 25% above the daily average at about 6PM. Meal times, especially lunch, are more apparent in the new units.

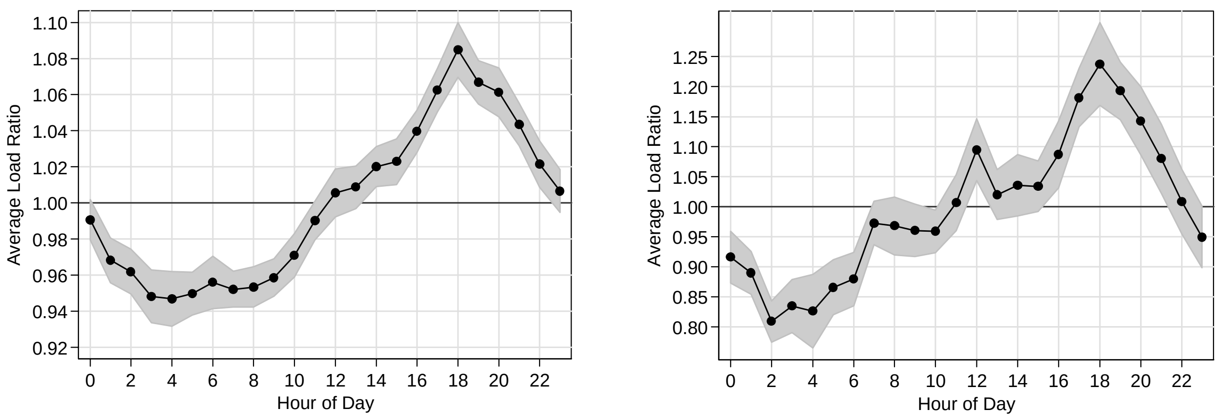


Figure 18. Hourly Load Profiles– existing refrigerators (left) and new refrigerators (right)
gray shading shows and 90% confidence interval

The sites with new refrigerators tended to have a greater number of occupants and the occupants were more likely to report that they were home all the time than the average existing unit household (see Table 6). This slightly greater occupancy could not fully explain the dramatic difference in load shapes. Instead, the data confirm prior research (see Proctor et al in footnote 3 of Section 2) indicating that newer refrigerators have a more "peaky" load shape than older units. A key reason for this difference is that as the refrigerator box becomes more efficient, the proportion of refrigerator loads due to occupancy effects increases. It still requires the same number of Btus to remove the heat from a pot of hot leftovers, whether you put it in a well insulated box or a poorly insulated box. Therefore, energy usage and load shape will vary more with occupancy effects as the box is better insulated. This greater "peakiness" of new units means that the relative energy impacts during peak loads will be smaller than average savings.

We explored load shape differences based on several factors. For each of these comparisons, we have superimposed the two load shapes on one graph and included 90% confidence intervals. These graphs are on the next page.

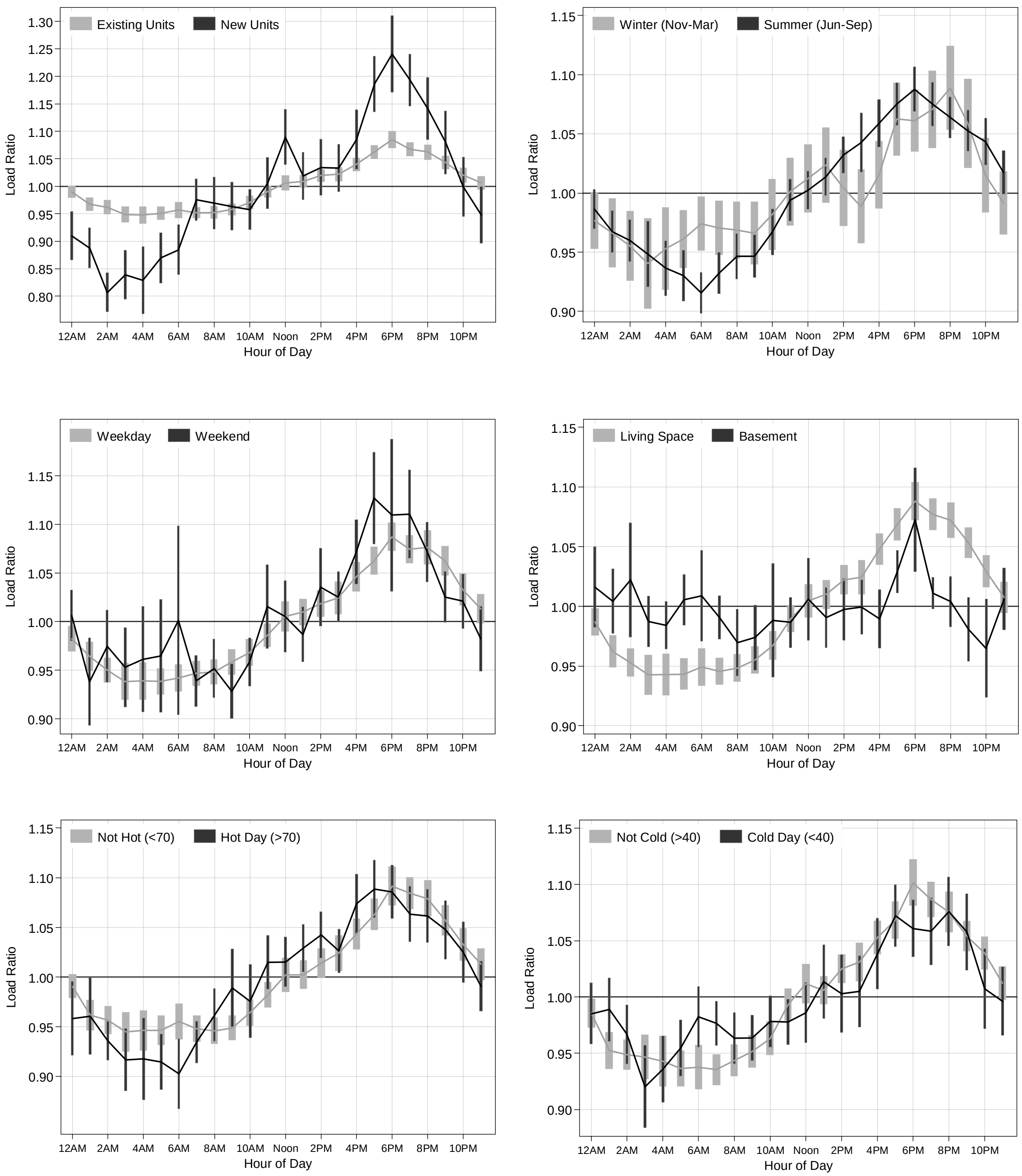


Figure 19. Hourly Load Profile comparisons: Existing v. New Units; Winter v. Summer; Weekday v. Weekend; Living Space v. Basement; Hot v. Not Hot Days; Cold v. Not Cold Days

note: bars show 90% confidence intervals – when bars don't overlap, differences are statistically significant.

The graphs show that:

- new and existing refrigerators have very different load shapes (as discussed previously);

- summer and winter load shapes are generally similar, with a deeper early morning drop off , a higher mid afternoon load and a slightly earlier peak in the summer;
- weekends tend to have a slightly earlier and larger evening peak than weekdays;
- basement refrigerators continue to show their odd character – exhibiting a nearly random daily load shape with a modest peak around 6PM. This finding is consistent with low occupant loads and unpredictable temperatures;
- hot days (>70°F average temperature) tend to have a larger dip in early morning usage (perhaps a house cooling off finally from night ventilation?) but are otherwise similar to other days;
- cold days tend to have an earlier and sharper morning increase (set back thermostat recovery?) and a lower evening peak.

Example Load Impact Calculation

A worked example may help clarify how to use the load shape data to develop estimates of peak demand impacts. The first step in estimating the wattage for a unit is to take the annual usage estimate and adjust it for the temperature conditions during the specific time period of interest using the temperature model developed in Section 6. For example, if the period of interest is a peak summer weekday with an average outdoor temperature of 90°F, then the following temperature adjustments would be used, based on site conditions:

1+.0265* (76.5 + (.4 * (90-70)) – 71) = 1.358 if living space with no air conditioning

1+.0265* (76.5 + (.08 * (90-70)) – 71) = 1.188 if living space with air conditioning

1+.0265* (71.4 + (.21 * (90-70)) – 65) = 1.281 if basement

where:

.0265 = average temperature effect on usage per degree F
76.5, .4,70,.08,71.4,.21 are all values from the temperature model
90 is the outdoor daily average temperature during the peak
71 is the estimated annual temperature for kitchen units
65 is the annual estimated temperature for basements

For program average peak impacts, a weighted average of these three values can be used based on the population treated. For the current study units, 15% were in unheated basements, 20% were in air conditioned living space and 65% were in non air conditioned living space (ignoring 2 garage/porch units), yielding an average temperature correction of 1.31. The average wattage at 4PM can then be calculated as:

Existing Unit Watts (4 PM)= 1.04 * 1.31 * 1383 * 1000/8760 = 215 W

where:

1.04 = load shape factor for existing units at 4 PM (see table on next page)
1.31 = average temperature correction calculated above
1383 = average annual kWh of existing units
1000/8760 = converts kilowatt hours to average watts

The new replacement unit wattage can be estimated using the same temperature correction as:

Replacement Unit Watts (4 PM)= 1.085 * 1.31 * 0.88 * 484 * 1000/8760 = 69 W

where:

1.085 = load shape factor for new units at 4 PM (see table on next page)
1.31 = average temperature correction calculated above
0.88 = average actual usage as percentage of label-rated for new units
484 = average label-rated usage of new unit

Therefore, demand savings at 4PM during a hot summer day is calculated as 215-69=146 watts per unit replaced. This impact represents a 68% demand reduction (vs. a 70% reduction in annual energy usage). Other load shape factors can be used to estimate demand impacts at other times of day.

Table 13 provides the details on the load shapes for all of the breakouts in *Figure 19*.

Table 13. Load shape data break outs (Watts/Watt ratio of hour's load to daily average load)

			Existing Units		Existing Units in Living Space (Kitchen)							
Hour	Exist	New	Kitchen	Bsmt	Winter	Summer	Weekday	Wkend	<70°F	>70°F	>40°F	<40°F
0	0.990	0.910	0.987	1.016	0.976	0.986	0.982	1.006	0.991	0.958	0.986	0.985
1	0.967	0.888	0.962	1.004	0.966	0.967	0.964	0.938	0.961	0.961	0.952	0.989
2	0.962	0.807	0.953	1.022	0.955	0.960	0.950	0.975	0.957	0.936	0.949	0.967
3	0.949	0.839	0.943	0.987	0.940	0.948	0.938	0.953	0.945	0.917	0.946	0.920
4	0.948	0.829	0.943	0.984	0.953	0.936	0.939	0.961	0.946	0.918	0.943	0.936
5	0.951	0.870	0.943	1.005	0.961	0.930	0.938	0.964	0.946	0.914	0.936	0.955
6	0.957	0.885	0.949	1.009	0.974	0.915	0.942	1.001	0.955	0.903	0.938	0.982
7	0.952	0.975	0.946	0.991	0.970	0.932	0.947	0.939	0.948	0.935	0.935	0.977
8	0.953	0.970	0.948	0.970	0.969	0.946	0.948	0.952	0.946	0.962	0.944	0.963
9	0.958	0.963	0.955	0.974	0.966	0.947	0.958	0.928	0.949	0.989	0.951	0.964
10	0.970	0.958	0.967	0.988	0.982	0.967	0.968	0.958	0.964	0.976	0.963	0.978
11	0.990	1.006	0.989	0.987	1.001	0.994	0.986	1.015	0.982	1.015	0.993	0.978
12	1.006	1.089	1.005	1.006	1.012	1.002	1.005	1.005	1.002	1.015	1.012	0.986
13	1.009	1.019	1.010	0.991	1.024	1.013	1.010	0.987	1.002	1.029	1.006	1.014
14	1.020	1.034	1.022	0.997	1.004	1.032	1.018	1.035	1.014	1.043	1.025	1.003
15	1.022	1.033	1.024	0.999	0.989	1.042	1.024	1.025	1.024	1.026	1.031	1.005
16	1.040	1.085	1.048	0.990	1.015	1.059	1.046	1.072	1.043	1.074	1.053	1.038
17	1.062	1.186	1.069	1.029	1.062	1.075	1.063	1.127	1.063	1.089	1.068	1.072
18	1.085	1.241	1.088	1.073	1.061	1.088	1.087	1.110	1.091	1.086	1.102	1.061
19	1.067	1.193	1.077	1.011	1.071	1.075	1.074	1.111	1.084	1.063	1.086	1.058
20	1.062	1.141	1.072	1.004	1.089	1.064	1.076	1.072	1.079	1.061	1.075	1.076
21	1.043	1.079	1.053	0.981	1.059	1.053	1.062	1.025	1.057	1.048	1.054	1.058
22	1.020	1.000	1.029	0.965	1.015	1.043	1.033	1.021	1.033	1.025	1.039	1.007
23	1.006	0.948	1.008	1.006	0.991	1.017	1.013	0.982	1.013	0.991	1.012	0.996

Note: the load shape data shown here should **not** be applied as a time of day correction factor for short-term metering (as has been recommended in articles in Home Energy Magazine). Much of the daily load shape is due to diurnal temperature variations which will already be used to adjust the data. A temperature adjusted version of the average existing unit daily load shape is flatter – all values are within 3% of the daily mean except from 5PM through 8PM when temperature-corrected loads are 4%-6% higher than average.

11. Refrigerator Auditing Guidelines

This report has provided a detailed and often dense examination of refrigerator energy usage and alternative auditing approaches. In this section, we provide some recommendations for practical refrigerator auditing approaches based on the findings of this project.

The "best" approach for auditing refrigerators can only be selected when "best" is defined – is it the most accurate approach or the most cost-effective approach? There is typically a trade-off between accuracy and cost – at one extreme, one could require every refrigerator to be metered for a full year before making a replacement decision, at the other extreme one could select units based on color, age, or flipping a coin. The best approach involves making an informed trade-off between cost and accuracy and therefore depends on the costs of implementing different methods. In addition, program managers may prefer an approach that is simpler, or less subject to occupant reported data, contractor interpretation, or potential tampering, even if it may be slightly less accurate than another approach. For these reasons, we cannot make one clear cut recommendation for the best auditing approach, but can provide guidelines for selecting an approach.

In addition, this project analyzed the energy usage and auditing approaches for refrigerators that were initially pre-screened by the PIVs to eliminate units that were relatively new (apparently almost all units built after 1990) or the customer was unwilling to consider replacing. Therefore, any recommendations are applicable only to older units.

Audit Approach Selection Guidelines

Based on the results from this project, we make the following recommendations concerning audit approaches:

1. The current PIV adjusted rated usage methods should not be used – they systematically over-estimate usage and qualify far too many units for replacements.
2. Short-term metering for two hours, done properly (see next section), is the most accurate audit approach examined and captures 90% of the potential net benefits available. This approach should be used on all units if there is no incremental audit cost compared to other methods. One hour metering is acceptable if the home visit is otherwise completed in less than two hours.
3. The new proposed rated usage approach (adjusted based on number of occupants, presence of icemaker, anti-sweat heater switch position, door seal condition, units bought used, and estimated winter thermostat settings) is worth using for refrigerators in the living space if the incremental cost of two hour metering is more than $1 above the cost of this rated usage approach.
4. A simple usage estimate of 111% of the label-rated usage can be used instead of the more complicated approach developed in this study if program managers are concerned about subjective judgments (e.g., door seal condition) or otherwise prefer a simpler approach. This 111% approach performs fairly well -- providing $8 lower net benefits per audit than the more complicated method ($42 vs. $50).
5. Units where the label-rated usage value can't be found in look-up tables should be audited with short-term metering. The PIVs were able to find rated usage values for about 90% of all units encountered in the field. In some cases a bad match was made and the rated value was entered incorrectly by the PIV. The procedures for looking up each model should be reviewed and made as "smart" as possible by incorporating string matching algorithms that find likely mistakes (e.g., S vs. 5) and provide a good indication of the quality of a match.
6. All refrigerators located in basements, garages, porches, and other spaces outside the main living space should be metered for at least one, and preferably two hours (see metering

recommendations below). These units should not be audited using an approach based on rated usage since those approaches perform poorly in these units.

7. Audits based solely on unit age, color, or power factor perform poorly compared to methods based on rated usage or metering and should not be used.

Short-Term Metering Protocol Recommendations

Short-term metering as practiced by the PIVs in this study did not perform very well compared to its potential. The field protocols need some revisions and must be implemented more carefully. Some specific recommendations include:

Metering Period Length

Metering should be done for at least one hour and preferably two hours. The incremental benefit from metering for two hours compared to one hour is relatively small, but if the audit visit is expected to last long enough, then metering should continue as long as feasible during the home visit without substantially increasing costs. One hour metering would only be preferred to two hour metering if the cost savings were at least $2 per audit and preferably $5 per audit (to reflect some potential problems with real world very short-term metering). A combined approach could be used which recommends two hour metering (to be used whenever the home visit otherwise lasts long enough) but allows as little as one hour metering if the home visit is completed quickly. There could be a small reduction in the audit fee (e.g., $5) as a disincentive to shorter metering when allowed. The length of time for the metering should be entered accurately by the PIVs – they should not use values rounded to the nearest 15 minutes or half hour.

Temperature Adjustment

In addition to adjusting for the length of metering compared to a year, the metered usage should be adjusted for the difference between the current room temperature and the estimated average annual temperature in the space. This annual average can be estimated as 71°F in the living space, 65°F in basements, and 58°F in unheated garages and porches for houses in New England. More precise annual indoor temperature estimates can be made using methods and formulas shown in this report. The usage adjustment should be made using this equation:

$$Usage_{avg} = Usage_{test} * (T_{avg}-33)/(T_{test}-33)$$

where $_{avg}$ refers to annual average conditions and $_{test}$ refers to test conditions.

This approach provides a 2.6%/°F adjustment for most living space units, 3.1%/°F for most basement units, and 4%/°F for most garage and porch units.

Defrost Cycles

In automatic defrost units, defrost cycles during metering can often be avoided by either:

- adjusting a set screw in the unit to cycle past a defrost before metering;
- monitoring wattage draw during metering for short periods greater than about 400 watts;
- monitoring freezer temperatures during metering (using a min/max freezer thermometer) and flagging units where the maximum temperature is greater than 32°F and yet frozen foods (e.g., ice cream, ice cubes) aren't all thawed.

These precautions can be taken to improve metering accuracy but they are only likely to affect about 4% of the replacement decisions (for programs replacing about half the units audited) and therefore can be considered optional. If defrost cycles are avoided, then metered usage can be adjusted upward by 8% to reflect the typical incremental usage associated with defrosting.

Time of Day

Time of day correction factors should not be used since the temperature correction already incorporates virtually all of the time of day effect.

Unplug / Re-plug

Most refrigerators require several minutes (typically 5-8) after being unplugged before the compressor will restart. If an auditor waits this time period before metering, then metered results will be biased high because of the added heat gain into the unit while the compressor is off. This time period should be included in the metering to reflect the fact that loads continue while the compressor is off and a longer cycle will occur at restart to remove the extra load.

Maximum Usage Threshold

A maximum usage threshold may be worth employing to avoid some large overestimates of usage that can occur when metering units that are running all of the time in cool spaces. A usage limit could be established as either somewhere between 160% and 200% of label rated usage (to reflect 100% duty cycle) or as a full year at the running wattage draw (can be recorded during metering when the compressor is on). A high use threshold can be considered optional since it will mostly affect units that qualify for replacement anyway, but should produce better realization rates as it prevents some large over-estimates.

National Grid’s AMPcalc software provides most of the calculation features outlined above – including a temperature correction and usage threshold -- and can be used “as is” provided that PIVs actually use the temperature adjustments whenever needed.

A. Field Data Collection Form

Refrigerator Evaluation Project Field Data Collection Form

Date: **Technician:** HDMC WMECO RCS

Customer Info

Name: ____________ **Address:** ____________ **City:** ____________

Phone#(s): **Account#**

Refrigerator Info

Make: ____________ **Model:** ____________ **Year:** ______(est) **Volume:** ______(est)

AHAM Usage: ____________ **Est. kwh/yr:** ____________

Style: Side-by-Side / Top Freezer / Bottom Freezer / Single Door **Defrost:** Auto / Manual / Partial

Icemaker: None / Thru Door / Internal Icemaker not hooked up **Avocado/Harvest Gold**

General Condition: Like New / Good / Fair / Poor

Door Seal Condition: Like New / Some Wear / Noticeable Gaps / Many Big Gaps

Location: Kitchen / Other:____________ **Unheated** **Recessed into cabinets?** Yes / No / Partial

Clearance (inches): Left Side: ______ Right Side:______ Rear: ______ Top: ______

Other Location Info: High Sun Exposure On Exterior Wall Wood Stove nearby Other:

Refrigerator Settings

Anti-Sweat Switch Position: On ("reduces condensation") / Off ("saves energy") / NA

Thermostat Settings: Fridge:______ Range Cold/Warm:______ Freezer______ Cold/Warm:______

Measured Temperatures: Fridge: °F Freezer: °F

Occupant Interview

Where did Fridge Come From: Bought new / Came with house / Bought or got used

Was Fridge ever repaired: No / Yes minor repair at home / Yes major repair at shop Reconditioned

Anti-Sweat Heater Switch: Never use / Use Seasonally / Other:____________

Fridge Temp. Adjustments: Never adjust / Adjust How adjusted: ____________

Household Occupants: #Adults >=18 ______ Children 6-17 ______ Children <6 ______

People Home During WeekDays 9-2: Never / Rarely / Often / Always Set schedule days per week: ______

Winter Space Thermostat Settings: Typical______ **Setbacks:** Night ___°F ___hrs Day___°F ___hrs

Cooling System: Central Air / Room AC in fridge room / Room AC in adjacent room / None

A/C Usage: Set Point: ______ **How Often Used:** All Summer / Quite a bit / Only a few times / Never

Additional A/C Info: Only use at: Night / Day Other: ____________

Special Conditions that may affect fridge usage: cook a lot rarely cook Fridge runs all the time

Doesn't keep food cold Seasonal residence____________ Fridge room not heated

Other:

Logger Equipment Info

No Metering Reason: pick-up hassle can't wait to replace can't meter Other/Explain:

WattsUp Serial#____________ Time Deployed: ____________

HOBO Serial#____________ Time Deployed: ____________ Location: ____________

WattsUp Zeroed at Start up: ☐ **HOBO LED blinking:** ☐ **Customer Agreement Signed:** ☐

v 1.6 8/06/03

Questions? call Bryon Martinez 508-836-9500 x3252
or Michael Blasnik 617-323-1225

B. Modeling Usage/Temperature Relationships

The effect of room temperature on energy usage for a given refrigerator can be estimated in several ways. A key issue is selecting the proper time period for analysis – should the model be based on one hour data, two hour averages, daily averages? Shorter-term data may lead to bias if occupant loads (e.g., door openings) are correlated with indoor temperature. Longer term data leads to a loss of information from the reduced number of data points and smaller range of temperatures due to averaging. For example, two weeks of metering provides 336 hourly data points covering the full range of temperatures, but daily data would include just 14 points and potentially a much narrower range of temperatures as within-day cycles are averaged out.

Figure 20 shows the profile of usage (expressed as the ratio of hourly usage to the daily average) and temperature deviations over the course of a day averaged across all sites and days in the dataset. The profiles are very similar with the minimum usage occurring at about 4 AM and the minimum temperature at about 5 AM. Usage increases before temperature and more sharply than temperature in the morning due to some occupants waking up and using the refrigerator while preparing breakfast. Clear peaks can be seen at lunch time and especially dinner time. It is not completely clear from the figure how much of the usage changes are due to occupancy effects and how much are due to temperature.

Figure 21 shows the same data except that usage is plotted against temperature and each point is labeled by the hour of the day. The black line in the figure shows an apparently good regression fit, but the slope implies that usage varies by 6.3%/°F. The gray line shows the expected temperature slope of about 2.5% based on prior research and engineering principles. The figure shows a clear pattern with low usage in the very early morning hours helping tilt the slope upward and dinner hours pulling the top of the regression line toward them. The figure confirms that patterns in occupancy are likely to bias the apparent relationship between usage and temperature when looking at data on a less than daily basis.

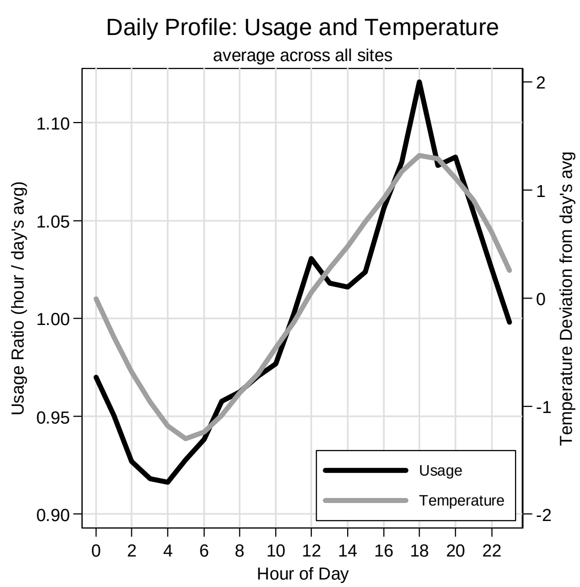


Figure 20. Hourly Profile of Usage and Temperature

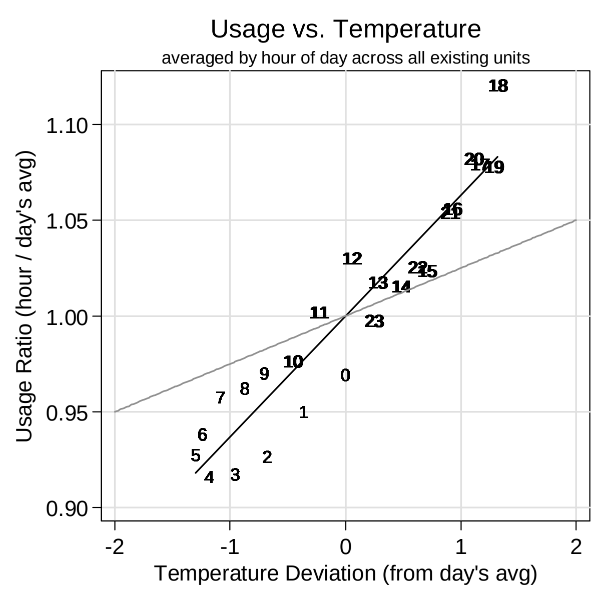


Figure 21. Usage vs. Temperature by hour of day

Further evidence of this bias was found by fitting models of usage against temperature using varying period lengths and finding that daily data had a noticeably smaller slope estimate than shorter periods – 2.6% vs. about 3.3% (for 1, 2, and 4 hour data).

The regression models based on 24 hour data had about twice the uncertainty in estimated annual usage as those based on shorter time periods due to the reduced range of temperatures and reduced number of data points. We believe this drawback is the result of a worthwhile trade-off between bias and variance.

The individual regression slopes varied fairly widely across sites, especially for sites with smaller ranges of temperatures. Figure 22 shows the usage/temperature relationship for a unit that experienced a wide range of indoor temperatures. This unit's temperature slope is well-determined at 3.86%/°F (based on 70°F as the "center"). Daily usage varies by about a two-to-one ratio over the wide range of temperatures. Many sites did not have such a clear relationship either due to greater variability in usage unrelated to temperature and/or a much smaller range of indoor temperatures.

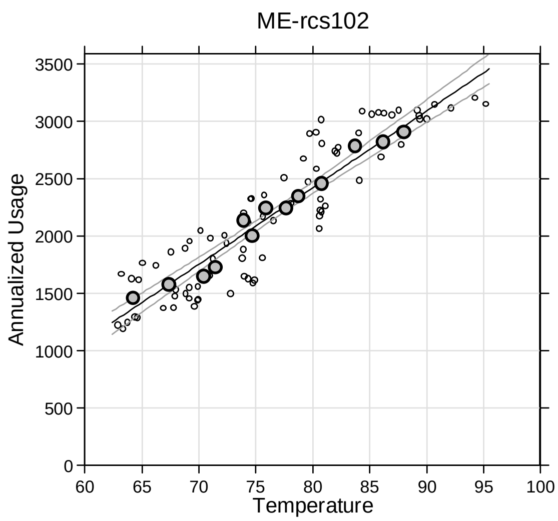


Figure 22. Usage vs. Temperature: site with wide temperature range.

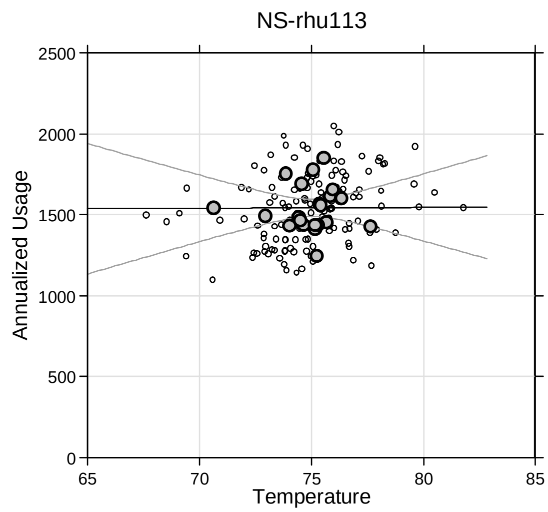


Figure 23. Usage vs. Temperature: site with narrow temperature range

Notes: Usage values are annualized to 8760 hour year. Large gray-filled circles= daily averages. Small circles=4 hour averages. Black line=linear regression fit on daily data. Gray lines=±90% confidence interval.

Figure 23 shows a site with a narrow range of indoor temperatures (note that the horizontal scale only covers half the range as in the graph on the left). It appears that there is little if any relationship between usage and temperature. This lack of relationship could be due in part to the narrow range of temperatures – if the indoor temperatures had varied over a 20°F range, a slope may have been more apparent.

In selecting an analysis approach, we explored ways to avoid the time-of-day bias while mitigating the loss of information that results from using daily data. We also sought ways to estimate the temperature effect for sites with too little data to develop reliable estimates. To address the time of day issue, we tried including time-of-day indicator variables in the regression models, allowing these variables to reflect some of the non-temperature variations. This approach reduced the estimated temperature slopes but they still appeared to be biased high compared to the daily data.

The best approach identified for resolving the main analysis issues is to employ a "random coefficients" regression model[7] using a pooled data set of daily averaged data (referred to as the RC24 model). This type of model is similar to estimating a separate regression for each unit, except the results of the

---

[7] We used the implementation in Stata using the command xtrchh2. For statistical details see Hildreth, C. and C. Houck, 1968. "Some estimators for a linear model with random coefficients," Journal of the American Statistical Association 63: 584-595

individual models are pooled and used to re-estimate the individual models. The random coefficients approach can be thought of as an empirical Bayesian method that employs the average temperature slope across sites as prior information in fitting the site-specific models. Each unit's temperature slope is essentially estimated as a weighted average of the individual unit's regression result and the overall average slope across all units. The weighting depends on the strength of each unit's relationship. For example, a unit where the usage/temperature relationship is weak (perhaps due to a narrow range of temperatures) has model results that rely more heavily on the average of all units, while units with strong relationships are relatively unaffected by the other units' results.

The specific model that we fit used the ratio of each day's usage to the average usage for that site as the dependent variable and the average daily temperature minus 70° as the independent variable. The advantage to this specification is that it estimates temperature effects as having a percentage impact instead of a kWh impact and the intercept provides a direct estimate of annual usage at 70°F (actually it is the ratio of average annual usage at 70°F to the average usage metered, but simple multiplication recovers the kWh estimate). The model excluded the one site where the unit apparently malfunctioned during the metering, but included the sites with "flat" usage (which reduced the overall slope estimate by about 0.2%).

The RC24 model provided an overall estimated temperature slope across all existing units of 2.65%/°F, consistent with engineering principles and prior research. Results for the two sample sites shown previously help illustrate how the model works. The unit with a wide temperature range and high quality fit has essentially the same slope under the RC24 model as the unit-specific model (RC24 estimate was 3.84% vs. 3.86% for the unit-specific model). The unit with the narrow temperature range and "messy" usage has an RC24 slope estimate of 1.0% (vs. 0.0% for the unit-specific regression).

We used the site-specific slope estimates from the RC24 model to estimate the temperature slope for 179 of the 186 sites with usage data. We used the overall RC24 slope for the five sites without temperature data (and imputed the average indoor temperature from a regression model relating indoor and outdoor temperatures). We also used the overall RC24 slope for one "flat" usage site with a limited temperature range and apparent negative slope. For the one site with a clear shift in usage, we used the metered data "as is" with no temperature adjustment.

Anti-sweat heater switch adjustment

The results of the regression analysis provide an estimate of the annual usage for each unit normalized to a room temperature of 70°F -- *assuming that the anti-sweat heater switch position stays the same*. However, some occupants claimed to use the anti-sweat heater switch seasonally. If an occupant claimed to use the switch and it was set appropriately for the season (on during the summer or off during the winter), then we assumed that they did use it – 17 units met this criteria. If an occupant claimed to use the switch seasonally but it was off in the summer or on in the winter, then we assumed that they didn't use it. For the 17 unit where we decided that they may have used the switch, we adjusted the usage estimate. This adjustment first involved estimating the incremental usage associated with turning on the switch. For most units, we found that rated anti-sweat switch use in the CEC database (as the difference between the high and low usage figures). Based on all units where we found this data, we calculated the median percent change in usage from the switch and found it to be 15% and used this value for cases where we didn't have direct data. We then adjusted each of the 17 unit's usage estimates by one half of the switch use -- increasing the usage if the switch was off in the winter when we metered and decreasing the usage if the switch was on in the summer when we metered.

Adjusting to Annual Average Indoor Temperature

The result from this adjustment was our best estimate of annual usage at a room temperature of 70°F. We used the temperature slopes from the RC24 analysis to adjust these usage estimates to the average annual indoor temperature was estimated by the temperature model. This temperature adjusted usage was considered the best estimate of the annual usage of each unit in its environment.