PACIFIC GAS & ELECTRIC COMPANY 1996 NON-RESIDENTIAL NEW CONSTRUCTION EVALUATION

REVISED FINAL REPORT

Study ID Number 389

July 29, 1998

Prepared for :
Valerie Richardson
Pacific Gas & Electric Company
123 Mission Street
San Francisco, CA

Prepared by:
RLW Analytics, Inc.
1055 Broadway
Sonoma, CA 95476

Project PG&E-08

Table of Contents

Executive Summary

Introduction

This is the final report of the 1996 Non-Residential New Construction (NRNC) Program evaluation. The evaluation was conducted by RLW Analytics and Architectural Energy Corporation from August 1997 through January 1998.

This report details findings of energy and demand savings at the whole building level and for lighting, HVAC, refrigeration, motors, and shell end-uses. Both net and gross savings are presented.

The evaluation relied on the use of model-based statistical sampling, on-site engineering surveys, DOE –2.1 building simulation models, and econometric analysis to develop the findings presented. A sample of 138 participant buildings and 138 non-participant buildings were surveyed and modeled to estimate gross energy savings relative to a baseline level. An additional telephone survey was conducted with decision-makers to collect data to estimate free-ridership and spillover. Net savings were developed using logistic and linear regression modeling to predict efficiency choice in the absence of the program.

The 1996 evaluation benefited greatly from the project team's experience with the 1994 PG&E / SCE NRNC evaluation. Valuable lessons were learned during the 1994 evaluation that helped to refine the methodology used in this study. Four key refinements to the 1996 study were:

- An improved sample design stratified by the estimated energy savings of participants
- The use of DOE modelers to conduct the on-site surveys
- The development of the initial model shortly after the survey visit
- The introduction of scaled variables in the econometric analysis

A brief overview of the 1996 evaluation methodology appears below.

Study Design

The goal of this evaluation was to estimate the net and gross energy and demand savings of the 1996 nonresidential new construction program.

The primary deliverables of this evaluation were:

1. Gross savings estimates of annual energy and summer peak demand
2. Net savings estimates of annual energy and summer peak demand
3. Parametric runs to isolate the influences of various measures and end-uses. These parametric runs were for the lighting, HVAC, refrigeration, motors, and shell end-uses.

The RLW Analytics/AEC team used a methodology similar to the 1994 NRNC study, with important modifications to reflect what was learned from that study. The basic approach relied on engineering models to develop gross savings estimates and econometrics to determine the net-to-gross. This methodology conforms to the CADMAC protocols with the important exception that statistical sampling was used in the place of an attempted census of program participants.

Data Collection

A major portion of this project was the collection of the building and decision-maker data necessary to determine the program impacts. Overall, the data collection process ran quite smoothly - no problems were encountered that had an adverse impact on the overall quality of the data. The data collection process used in this study yielded significantly better information than the process used in the 1994 NRNC study.

The data collection process was designed to collect the highest quality data in the most efficient manner possible. This process relied on several people working together to ensure a seamless information flow. Figure 1 shows a graphical representation of the data collection process.

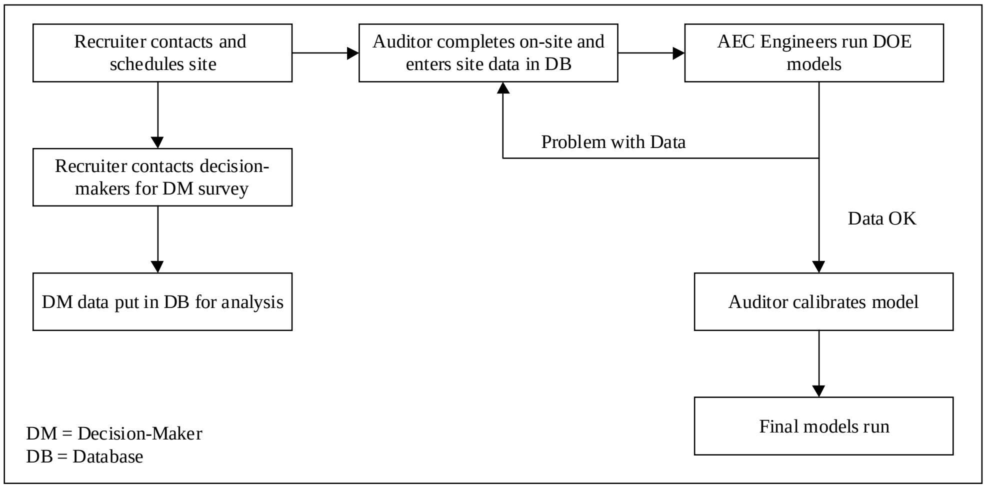

Figure 1: Data Collection Process

The recruiter was responsible for making contact with the site representative and securing its participation in the study. Once that was accomplished, the recruiter scheduled the on-site visit and provided the information to the field surveyors from RLW Analytics and AEC. The recruiter then completed the decision-maker survey with the initial site contact and any additional contacts that were necessary to answer the decision-maker questions.

The on-site surveyor collected building description and operation information from the site and entered the data into a database. Automated modeling software was used to create DOE-2 input files. The surveyors were responsible for checking the models created from the field data, and correcting the data if necessary. The on-site surveyor was also responsible for calibrating the model to billing data or short-term meter data, if available for the site. Senior staff engineers of AEC and RLW checked the final model results for reasonableness.

The calibrated models were delivered to AEC, who produced all of the required parametric runs of the engineering models.

Engineering Models

Engineering models were developed for each building in the on-site survey sample using the DOE-2.1E building simulation program. A series of models were developed for each sample site, including:

- A "baseline" model representing the building with minimally compliant equipment and envelope efficiencies.
- An as-built model representing the building as found by the surveyors.
- A series of parametric runs to isolate the impact of motors, refrigeration, HVAC, lighting, and shell end-uses.

The models were developed using an automated BDL[1] generator, developed by AEC and RLW Analytics. This method ensured that all of the models were consistent, thus eliminating a potential source of bias in the results.

Analysis Baseline and Gross Savings Calculations

The estimates of gross program savings were made by comparing the as-built simulated building energy consumption to a baseline level of energy consumption[2]. The baseline energy consumption for all buildings was defined to be the energy consumption of the building as if all of the equipment was specified to be minimally compliant with Title 24 and the building was operated on the schedule found during the on-site survey.

A gross savings estimate was calculated for each building in the sample. The savings estimated were projected to the population of participants using model-based statistical sampling procedures. Gross savings estimates were developed for both the participant and the non-participant population.

Net Savings Methodologies

Net program savings estimates are the savings that directly result from program participation. Effects of free-ridership, or what the customer would have done in the absence of the program, have been factored out. Two net savings methodologies were used in this evaluation, a "difference-of-differences" approach and an econometric approach. Net-to-gross ratios from both methods are presented in this report. The reported savings is based on the econometric estimate of net-to-gross since the econometric estimate was more statistically precise than the difference-of-differences estimate.

Difference of Differences

A simple "difference-of-differences" estimation approach to net savings was done for this study. This method estimated net savings by comparing the savings of the participants in the sample to a "matched" sample of non-participants. The savings of the non-participant group is assumed to be the savings of the

[1] BDL is DOE-2's Building Description Language

[2] Because the default Title 24 operating schedules were not used to develop the baseline and because the area category method was used for each building regardless of the Title 24 compliance path actually elected, the savings calculated relative to the baseline in this study cannot be interpreted as the degree of compliance with Title 24

participants in the absence of the program. In this methodology, spillover among the non-participants is assumed to be offset by free-ridership among the participants but no attempt is made to measure either spillover or free-ridership

Econometric Modeling

An econometric approach to estimating net savings was also used in this study. The econometric approach appeared to provide a more unbiased and statistically reliable estimate of net savings than the difference of difference approach because it explicitly measured both free-ridership and spillover and controlled for self-selection and other decision-making factors affecting the efficiency choice of each sample site. Figure 2 shows the overall flow of data for the econometric modeling. In this methodology, a logistic regression was performed to create a participation model. This model estimated Mills' ratios for correcting self-selection bias. A second model was built, a linear regression, to estimate the savings of participants in the absence of the program.

The econometric approach also incorporated the relationship between PG&E's influence on the design of projects and the energy efficiency of the current project. This component of the model was used to estimate the spillover effect, i.e., the effect of the program on non-participant savings.

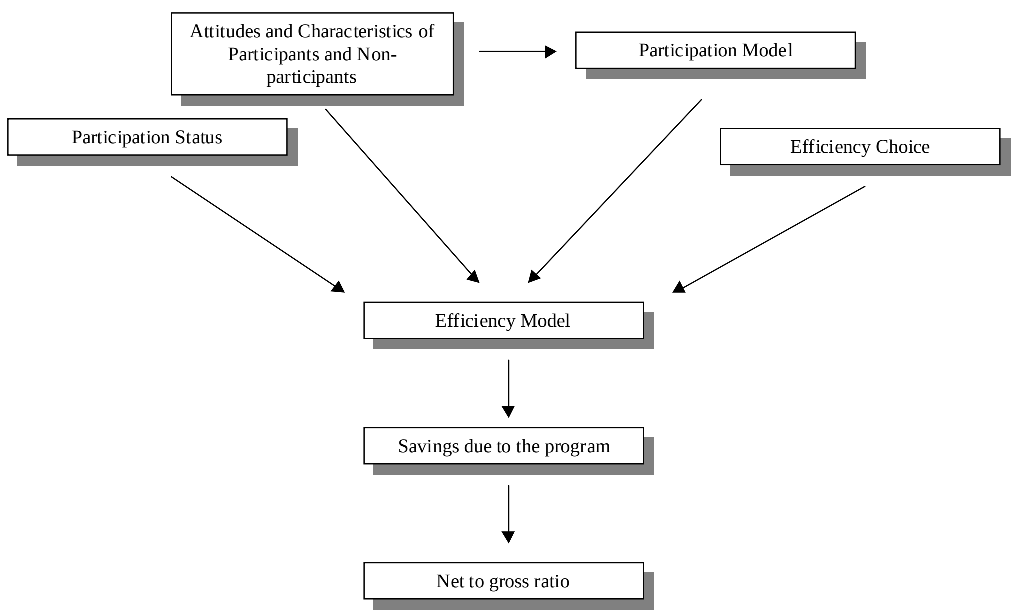

Figure 2: Econometric Modeling Overview

Findings

This section presents gross and net savings estimates for the population of program participants. Table 1 summarizes the overall evaluation findings. These findings are described in greater detail below and later in the report. The net savings reported in Table 1 are based on the econometric net-to-gross analysis.

Spillover is excluded from the net savings results reported in the table below, but was found to total 20,400,582 kWH.

			Net-to-Gross				
Whole Building	**Gross Savings**	**Gross Realization Rate**	**1-FR**	**SO**	**Net-to-Gross Ratio**	**Net Savings**	**Net Realization Rate**
ex ante							
kW	19,110	100.0%	75.0%	-	75.0%	14,333	75.0%
kWh	80,398,024	100.0%	75.0%	-	75.0%	60,298,518	75.0%
Therms	n/a	n/a	n/a	n/a	n/a	n/a	n/a
ex post							
kW	20,000	104.6%	69.8%	--	69.8%	13,951	73.0%
kWh	83,970,000	104.4%	69.8%	-	69.8%	58,569,164	72.8%
Therms	n/a	n/a	n/a	n/a	n/a	n/a	n/a

Table 1: Summary of Evaluation Findings

Gross Savings

Program participants saved 83,970 MWh of energy in their first year of operation. This is a realization rate of 104.4% of the verified savings estimate previously filed by PG&E. The relative precision of the estimate is ±6.0% at the 90% confidence level, meaning that the gross program savings is estimated to be between 78,932 MWh and 89,008 MWh.

The summer on-peak demand savings is 20.0 MW. The realization rate is 104.6% of the verified program savings. The relative precision is ±7.1% at the 90% confidence level, meaning that the gross program demand savings is between 18.6 MW and 21.4 MW. Table 2 below shows the energy and demand savings by PG&E costing period. The winter costing periods have greater energy savings because they consist of more hours than the summer periods.

Period	Energy Savings (MWh)	Energy Rel. precision	Demand Savings (MW)	Demand Rel. precision
Annual	83,970	± 6.0%	--	--
Summer On-Peak	7,697	± 6.3%	20.0	± 7.1%
Summer Part-peak	9,503	± 5.7%	19.3	± 7.4%
Summer Off-Peak	13,840	± 6.7%	19.2	± 6.9%
Winter Part-peak	26,530	± 6.5%	19.7	± 7.1%
Winter Off-Peak	26,410	± 6.9%	18.5	± 7.6%

Table 2: Participant Energy and Demand Savings by Costing Period

To compare participants and non-participants, the savings of each group relative to their own baseline is plotted in Figure 3. The figure clearly shows much higher levels of energy efficiency among participants than among non-participants. The participants' energy use was 19.2% better than baseline, while the non-participants' energy use was only 10.3% better than baseline. "Better than baseline" means that the buildings are more energy efficient than the baseline efficiency levels established for this study. Numerically, a building that is 20% better than baseline uses 20% less energy than it would have used if built to baseline efficiency levels. For summer on-peak demand, the participant group was 19.3% better than baseline while the non-participant group was 11.3% better than baseline.

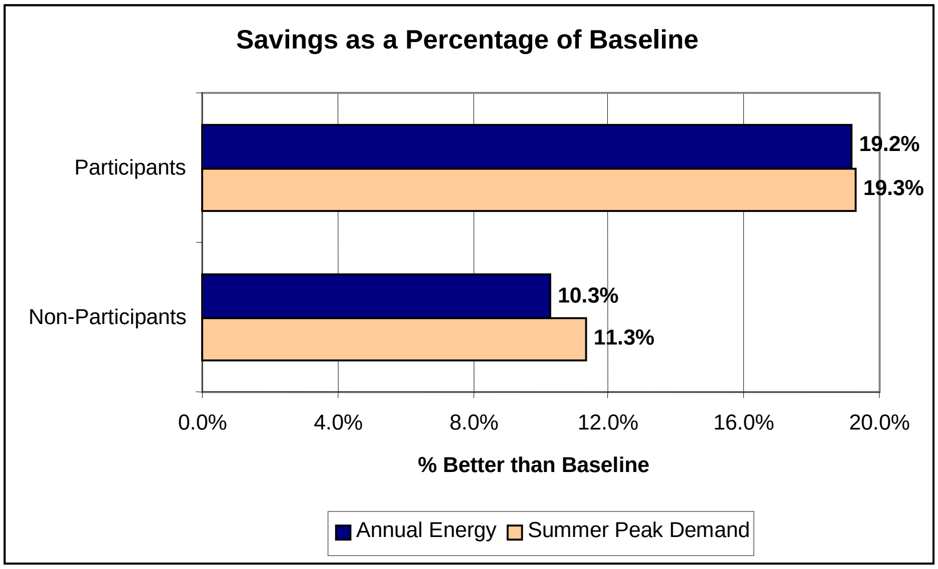

Figure 3: Gross Energy and Demand Savings Relative to Baseline

Energy and demand savings were also estimated for lighting, HVAC, refrigeration, motor, and shell end-uses. Figure 4 and Figure 5 show the composition of the annual energy savings and the summer on-peak demand savings for program participants, respectively. The shell measures did not produce any statistically significant savings. As expected, HVAC savings contributed more to the summer peak demand savings than to annual energy due to the seasonal nature of the end-use.

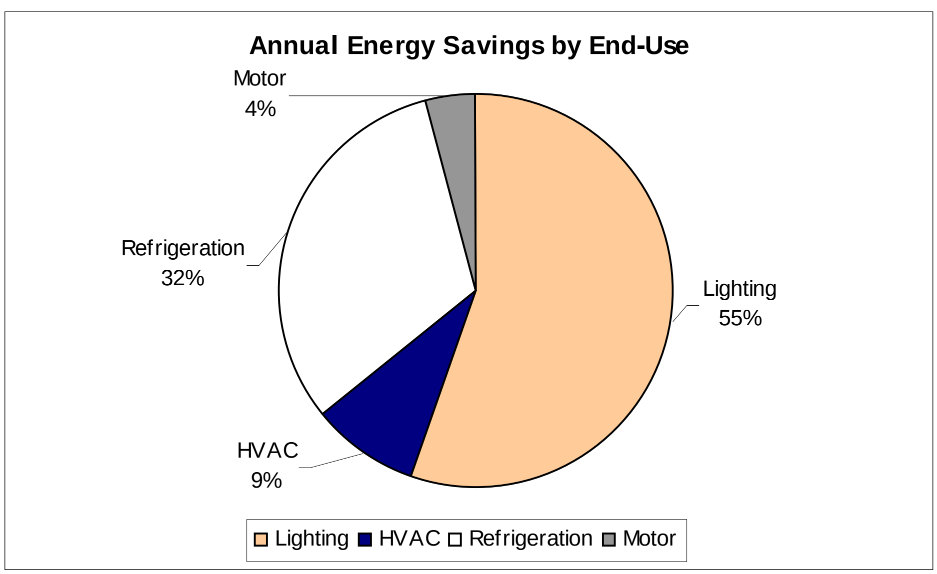

Figure 4: Annual Energy Savings by End-Use

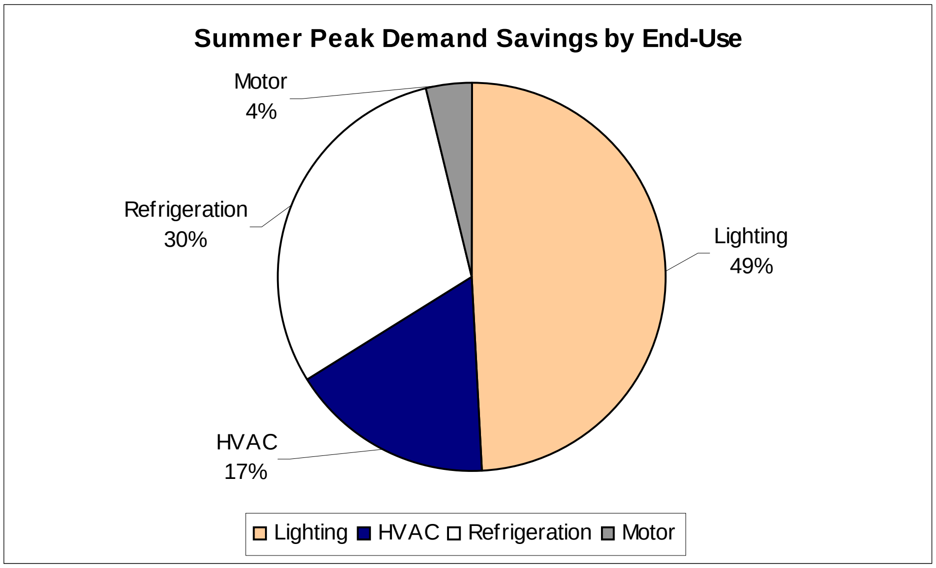

Figure 5: Summer Peak Demand Savings by End-Use

Table 3 shows the energy savings by end-use for each of the costing periods. Table 4 shows the summer on-peak demand savings for each end-use. Shell savings are not included because no statistically significant savings were found.

	Lighting (MWh)	Refrigeration (MWh)	Motors (MWh)	HVAC (MWh)
Annual	46,400	26,490	3,539	7,547
Summer on-peak	4,006	2,276	292	1,027
Summer part-peak	4,917	3,156	342	1,072
Summer off-peak	6,726	5,191	557	1,313
Winter part-peak	16,290	6,858	1,132	2,399
Winter off-peak	14,460	9,009	1,216	1,736

Table 3: End-Use Gross Energy Savings by Costing period (MWh)

.

	Lighting (MW)	Refrigeration (MW)	Motors (MW)	HVAC (MW)
Summer on-peak	9.8	6.0	0.8	3.4
Summer part-peak	9.7	5.7	0.8	3.2
Summer off-peak	9.4	5.8	0.9	3.3
Winter part-peak	9.7	5.6	0.7	3.2
Winter off-peak	9.5	5.5	0.8	2.8

Table 4: End-Use Gross Demand Savings by Costing period (MW)

Net Savings

Net savings is that part of the observed energy savings that can be attributed to the efforts of PG&E. As discussed in a prior section, two different methodologies were followed in the net-to-gross analysis: a relatively simple difference-of-differences approach and a more complex econometric approach. In the difference-of-differences methodology, the net-to-gross ratio was calculated by comparing (a) the gross savings relative to baseline of the program participants and (b) the gross savings relative to baseline of the non-participants. In the econometric approach, the net-to-gross ratio was calculated by using regression modeling techniques to estimate the net savings due to the program for each of the program participants and non-participants.

Difference of Differences Net-to-Gross

Table 5 summarizes the findings from the difference-of-differences analysis. The table shows the estimated net savings and net-to-gross ratio for both annual energy and summer peak demand savings.

	Net Savings	**Net-to-gross Ratio**	**Relative Precision**
Annual Energy	39,054 MWh	46.5%	± 25.5%
Summer Peak Demand	8.2 MW	41.0%	± 25.1%

Table 5: Difference of Differences Net-to-gross Ratio

The table also shows the relative precision of each estimate.[3] For example, in the case of annual energy, the net-to-gross ratio was estimated to be 46.5% with a relative precision of ±25.5%. The error bound for the 90% confidence interval for the true net-to-gross ratio is equal to 25.5% of the estimate, i.e. to ±11.9%. The 90% confidence interval for the true net-to-gross ratio can be calculated using the equation:

$$0.465 \pm (0.465 * 0.255) = (0.346, 0.584)$$

We can be quite confident that this interval contains the true net-to-gross ratio that would have been obtained by developing on-site surveys and building engineering simulation models for all program participants and a very large sample of non-participants using the methodology of this study and then analyzing the resulting data using the difference of differences methodology. The confidence interval reflects sampling variability and random measurement error but does not reflect any possible systematic measurement error that might be repeated throughout the data collection and engineering simulation or that might arise by neglecting explicit estimation of free ridership and spillover.

Econometric Net-to-Gross

Table 6 summarizes some of the findings from the econometric analysis. The table shows the estimated net savings (excluding spillover) and net-to-gross ratio for both annual energy and summer peak demand savings.

[3] Some definitions: The standard error reflects the standard deviation of an estimate in repeated sampling. The error bound at the 90% level of confidence is 1.645 times the standard error. The confidence interval is the estimate plus or minus the error bound. The relative precision is the error bound divided by the estimate itself.

	Net Savings	**Net-to-gross Ratio**	**Relative Precision**
Annual Energy	58,569MWh	69.8%	±10.5%
Summer Peak Demand	13.9 MW	69.8%	±12.1%

Table 6: Econometric Savings and Net-to-gross Ratios

The table also shows the relative precision of each estimate. For example, in the case of annual energy, the net-to-gross ratio was estimated to be 69.8% with a relative precision of ±10.5%. The error bound for the 90% confidence interval for the true net-to-gross ratio is equal to 10.5% of the estimate i.e. to ±7.3%. The 90% confidence interval for the true net-to-gross ratio is

$$0.698 \pm (0.698 * 0.105) = (0.625, 0.771)$$

The confidence interval for annual savings can be calculated in a similar way.

There is a 90% probability that these confidence intervals include the true values that would have been obtained by developing onsite surveys and building engineering simulation models for all program participants and a very large sample of non-participants using the methodology of this study and then analyzing the resulting data using the econometric methodology. These estimates reflect sampling variability, random measurement error, and explicit estimation of free-ridership and spillover. But these estimates do not reflect any possible systematic measurement error that might be repeated throughout the data collection and engineering simulation or any possible bias arising from inaccuracy in the assumed econometric model.

Sample Design

Introduction

The key to effective sample design is to take advantage of the association between the target variables to be measured in the study and any supporting variables already known from the sampling frame. For example, the savings of each program participant measured in this project can be associated with the estimate of savings recorded in the program tracking system. Stratified sampling is used to ensure that the sample has the best mix of small and large sites. Ratio estimation is used to expand the sample data to the target population, taking advantage of the supporting information. Both stratified sampling and ratio estimation are well known and widely used in load research and DSM evaluation.

The principal questions addressed in sample design are:

- How big should the sample be, both overall and within different subsets of the target population?
- How much statistical precision can we expect from the sample?
- How should the sample be stratified to get the best statistical precision?

The usual approach is to estimate the variance of the estimated savings in the program tracking system. This approach is not appropriate for stratified ratio estimation since the statistical precision depends not on the variance of estimated savings but on the strength of the association between the measured savings and the tracking estimate of savings. The Model-Based Statistical Sampling (MBSS) approach is to develop a statistical model describing the relationship between these variables, and then use the parameters of this model to develop the sample design. In this project the parameters of the MBSS model were estimated in our prior evaluation of the 1994 program.

Using this approach, RLW Analytics designed the participant sample to achieve ±10 percent precision at the 90 percent confidence level for the participants' annual measured energy savings. This analysis indicated that the participant sample size should be 138 sites, stratified by the tracking estimate of savings. The non-participant sample was matched to the participant population in terms of square footage and building type. A sample of 138 non-participant sites was selected from F.W. Dodge New Construction data.

Participants

RLW Analytics used the sites that received incentive checks dated in 1996 as a participant sample frame. A sample of 138 sites was drawn from a population of 405. The sample was stratified into 5 sampling strata and one certainty strata for a total of 6 strata by estimated annual energy savings. Sample size, population size, and stratum cutpoints are indicated in the Table 7 below.

Stratum	Maximum Energy Savings (kWh)	Population Size	Population Energy Savings (kWh)	Sample Size
1	53,465.	183.	3,470,614.	25.
2	130,705.	83.	6,990,246.	25.
3	321,374.	54.	11,544,704.	25.
4	474,417.	40.	16,088,200.	25.
5	958,205.	32.	20,826,142.	25.
6	4,000,000.	13.	19,802,836.	13.
Total		**405.**	**78,722,742.**	**138.**

Table 7: Stratified Sampling Plan for Participants

The total tracking savings for the 405 program participants was 78,722 MWH[4]. The anticipated precision from this sample design was ± 9.8 percent at 90 percent confidence. The estimated precision for participants was based on the model parameters used in the sample design, which are shown in Table 8.

Model Parameter	Value
error ratio	0.99
γ	0.47

Table 8: Model-Based Sampling Parameters for Participant Sample

The error ratio and γ were taken from the actual model parameters found in the 1994 NRNC study. The analysis variable is the actual energy saved and the explanatory variable is the tracking estimate of energy saved. The error ratio is a measure of the spread of the data around the trend line. It is analogous to the coefficient of variation. γ is a measure of the heteroskedastisity of the data. Heteroskedastisity is the tendency for the variation around the trend line to increase as the value of the stratification variable increases.

Non-participants

For the non-participant sample design, the *participant* population was re-stratified on building type and square footage. This two-way stratification defined the cells in the sample design, which was then filled with non-participant sites from the Dodge database. This procedure ensured that the non-participant sample could be well matched to the participant sample. Later in this section, a comparison between the participant and non-participant population is shown.

The sample frame for the non-participants was taken from the F.W. Dodge new construction database of projects started in 1995. The database was screened to eliminate out-of-scope and out-of-territory projects. The Dodge project was considered in scope if the building type was eligible for NRNC incentives.

[4] Later investigation revealed that only 392 of the participant population were legitimate participants. The savings calculations later in this report are based on 392 participants and the verified savings estimate as presented in PG&E's e-tables

The non-participant sample was developed using the method outlined in the flowchart below. This led to a non-participant sampling frame of 2,438 sites.

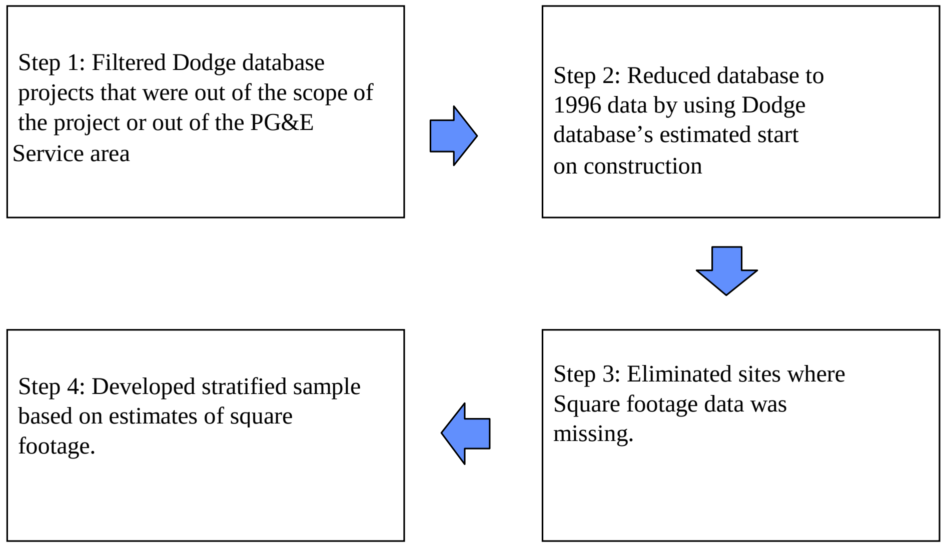

Figure 6: Non-Participant Sample Frame Development

The non-participant sample size was chosen to be 138 sites to match the participant sample size. The non-participant sample was stratified by building type and by square footage. Table 9 below summarizes the sample design used to select the 138 non-participants. For example, in the case of food stores, 4 sites were selected from each of 3 size strata. The number of sites from each building type and the allocation of the sample to the size strata was selected to match the participant population. In Table 9 and Table 10, a dash in the cell indicates that the data element is not applicable to that building type. For example, there were only 2 restaurant strata, therefore there was no strata 3 or strata 4 sample (Table 9) and there were no strata 2 or strata 3 cutpoints (Table 10).

Building Type	Stratum 1	Stratum 2	Stratum 3	Stratum 4	Total
Food Store	4	4	4	--	12
Medical	3	3	3	--	9
Manufacturing	6	6	6	--	18
Miscellaneous	5	5	5	--	15
Office	7	7	7	7	28
R Warehouse	2	2	--	--	4
Restaurant	2	2	--	--	4
Retail	7	7	7	--	21
School	6	6	6	--	18
Warehouse	3	3	3	--	9
Total					**138**

Table 9: Stratified Sampling Plan for Non-Participants

The square footage cutpoints for the non-participant strata are shown in Table 10. For example, in the medical category, stratum 1 consists of sites with square footage less than 75,073 square feet, and stratum 2 of sites between 75,074 and 194,104 square feet.

Building Type	Stratum 1 Max. Square Footage	Stratum 2 Max. Square Footage	Stratum 3 Max. Square Footage
Food Store	43,863	59,270	--
Medical	75,073	194,104	--
Manufacturing	60,663	145,000	--
Miscellaneous	43,602	344,979	--
Office	44,990	118,939	218,117
R Warehouse	67,640	--	--
Restaurant	5,895	--	--
Retail	29,000	144,222	--
School	32,862	51,177	--
Warehouse	177,363	335,794	--

Table 10: Strata Cutpoints

Sample design vs. actual sample

Table 11 shows a summary of the study population, sample design, and achieved sample. Although metered sites are shown in Table 11, they were not part of a nested sample. The metering was done to provide usage data for calibration of sites where billing data was unavailable or unreliable. See the Short-term metering section later in this report for more information.

		Sample Design			Actual Final Sample		
	Population	Phone	On-Site	Meter	Phone	On-site	Meter
Participants	392	138	138	25	141	141	10
Non-Participants	2,438	138	138	25	136	136	10

Table 11: Sample Summary

Table 12 shows the participant sample design and the actual participant sample. As the table shows, fewer than desired large customers (higher strata numbers) were successfully recruited. Overall, more participants were surveyed than called for in the original sample design.

Stratum	Design	Actual
1	25	29
2	25	24
3	25	24
4	25	30
5	25	26
6	13	8
Total	**138**	**141**

Table 12: Participant Sample Design and Actual Sample

There was no stratification of the participant sample by building type. Figure 7 shows the expected distribution of the participant sample by building type and the actual distribution of the participant sample. The distributions have been weighted by their inclusion probability, to reflect the fact that a particular building type would have tended to be in larger or smaller strata.

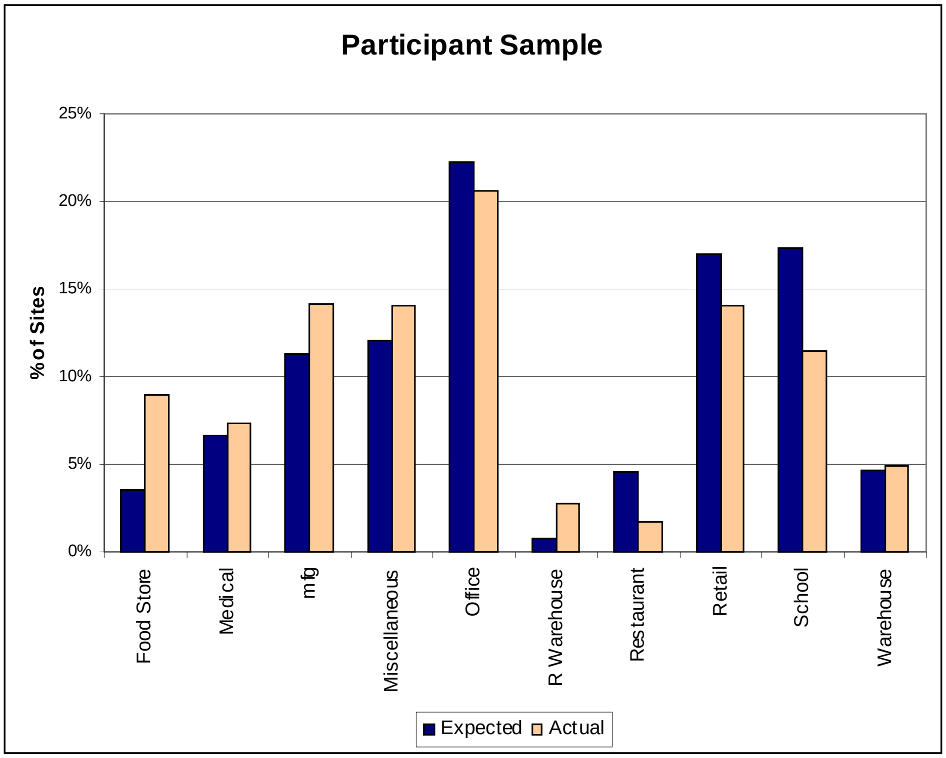

Figure 7: Participant Sample by Building Type

Figure 7 shows that the representativeness of the participant sample is reasonable. In categories such as food stores and restaurants, a better match would have been desirable, but since they represent a rather small portion of the sample, no significant impact on the results is expected from the sample distribution.

The non-participant sample was designed to be comparable to the participant population in composition. The participant population was stratified by building type and square footage. Non-participant sites were selected from the Dodge new construction database to fill that sample design. Table 13 shows the sample design and the actual non-participant sample by building type and size (square footage) strata. Stratum 1 consists of the smallest buildings. Each successive stratum consists of progressively larger buildings. The specific cutpoints differ by building category, as shown previously in Table 10.

In the table, the first number is the actual number of sites surveyed and the second number is the design for the cell. For example, in the food store stratum 1 cell, 5 sites were surveyed and the original sample design called for 4 sites.

The primary reason for the lack of larger buildings is due to the very few non-participant buildings with comparable square footage available. There were differences in the participant and non-participant populations. The program targeted large one-of-a-kind projects for participation, leaving very few large buildings that did not participate.

Category	Stratum 1	Stratum 2	Stratum 3	Stratum 4	Total
Food Store	5 of 4	0 of 4	0 of 4	--	5 of 12
Medical	9 of 3	0 of 3	0 of 3	--	9 of 9
Mfg	14 of 6	4 of 6	0 of 6	--	18 of 18
Miscellaneous	14 of 5	5 of 5	0 of 5	--	19 of 15
Office	12 of 7	11 of 7	1 of 7	3 of 7	27 of 28
R Warehouse	2 of 2	2 of 2	--	--	4 of 4
Restaurant	3 of 2	1 of 2	--	--	4 of 4
Retail	12 of 7	9 of 7	4 of 7	--	25 of 21
School	6 of 6	7 of 6	6 of 6	--	19 of 18
Warehouse	5 of 3	0 of 3	1 of 3	--	6 of 9
Total					**136 of 138**

Table 13: Non-participant Sample by Building Type and Size Strata

Figure 8 shows the non-participant sample design and the actual non-participant sample by building type.

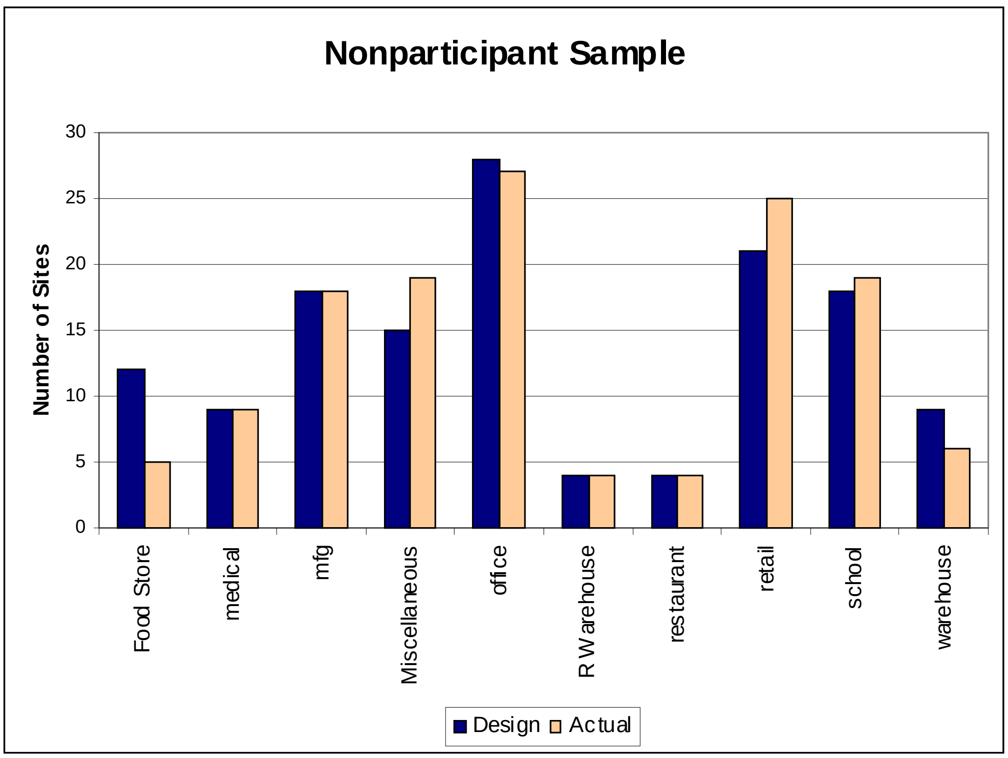

Figure 8: Non-participant Sample by Building Type

Data Collection

Overview

A major portion of this project was the collection of the building and decision-maker data necessary to determine the program impacts. This section discusses the effectiveness of the data collection effort.

Overall, the data collection process ran quite smoothly. This was due to the use of highly qualified staff for recruiting, surveying, and modeling. The data collection process used in this study represented a significant improvement over the process used in the 1994 NRNC study.

The data collection process was designed to collect the highest quality data in the most efficient manner possible. This process relied on several people working together to ensure a seamless information flow. Figure 9 shows a graphical representation of the data collection process.

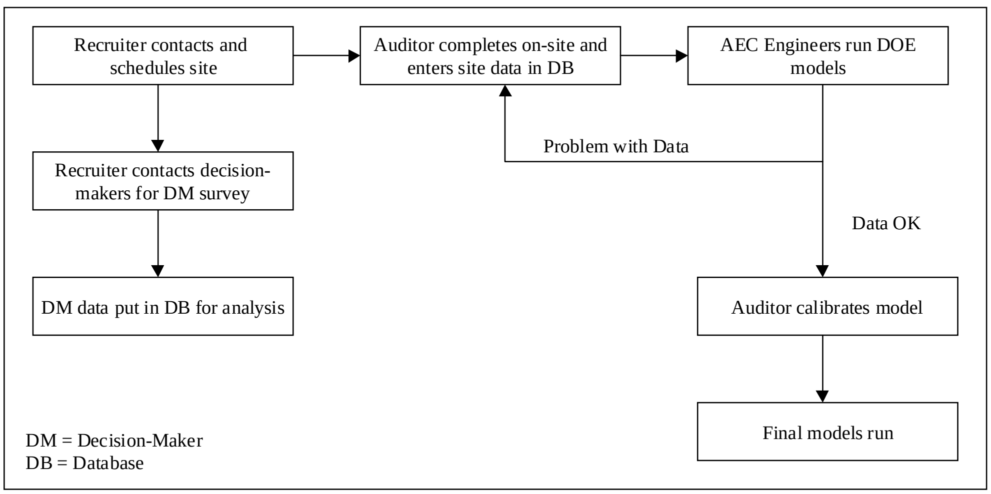

Figure 9: Data Collection Process

The recruiter was responsible for making contact with the site representative and securing their participation in the study. Once that was accomplished, the recruiter scheduled the on-site visit and provided the information to the field auditor. The recruiter then completed the decision-maker survey with the initial site contact and any additional contacts that were necessary to answer the decision-maker questions.

The on-site auditor collected building description and operation information from the site and entered the data into a database. Automated modeling software was used to create DOE-2 input files. The auditors were responsible for checking the models created from the field data, and correcting the data if necessary. The on-site auditor was also responsible for calibrating the model to billing data or short-term meter data, if available for the site. AEC and RLW senior staff checked the final model results.

The calibrated models were returned to AEC, who produced all of the required parametric model runs.

Recruiting

The recruiting process included the use of staff experienced in construction and development. This ensured that the professionals being contacted did not feel that they were speaking with someone who did not understand the basic issues in the field.

Table 14 summarizes the recruiting effort. A conversion rate of 53% was achieved. Only 5% refused to participate in the study.

In the table, "completed" means that the site was successfully recruited and audited. "No contact" means that attempts to contact a decision-maker at the site failed. "Dropped" indicates that the site was eliminated because it was found to be outside the scope of the study or the strata that a particular building fell into was filled before the recruiting process could be completed for a building. Participant buildings were typically dropped for the latter reason. Buildings found to be outside the scope of the project – typically non-participants – were those buildings that were not completed in 1996 or performed work that would not have been eligible for participation in the program (e.g. cosmetic renovations).

Disposition	Participants	Non-Participants	Total
Completed	141	136	277
Refused	6	20	26
No Contact	0	1	1
Dropped	33	182	215

Table 14: Recruiting Disposition

Decision-Maker Surveys

The recruiters completed decision-maker surveys for each audited site. Recruiters made an average of 3.5 calls to 1.1 different individual decision-makers to complete each survey. Figure 10 shows the distribution of the number of calls necessary to complete each survey and the number of individual decision-makers contacted.

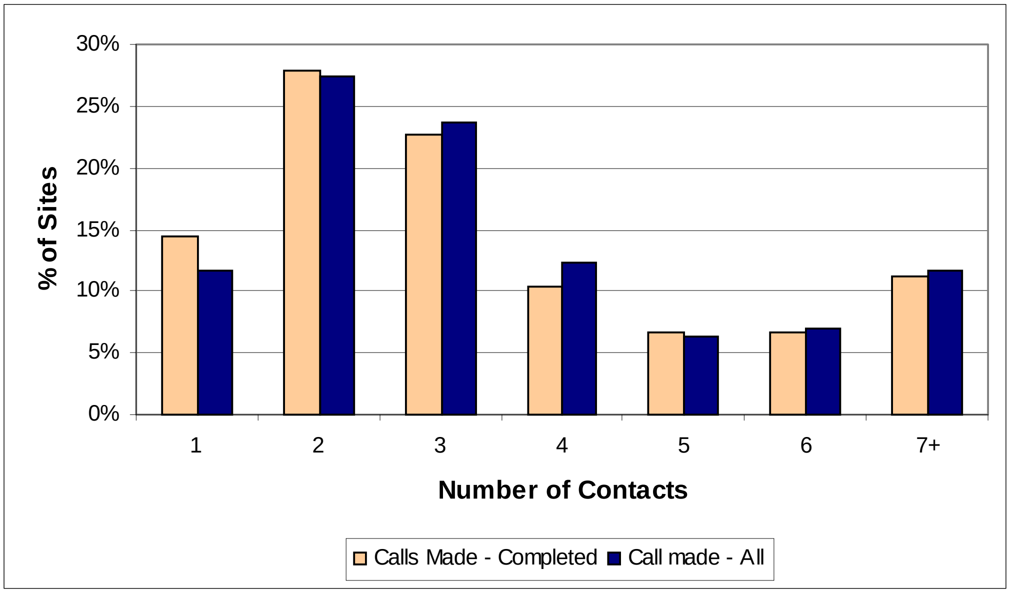

Figure 10: Number of Calls to Complete Each Decision-maker Survey

Figure 10 shows the number of calls made to all sites, including all dropped sites, and to only the sites that were ultimately surveyed. Figure 11 shows the number of individual decision-makers who were needed to complete each survey.

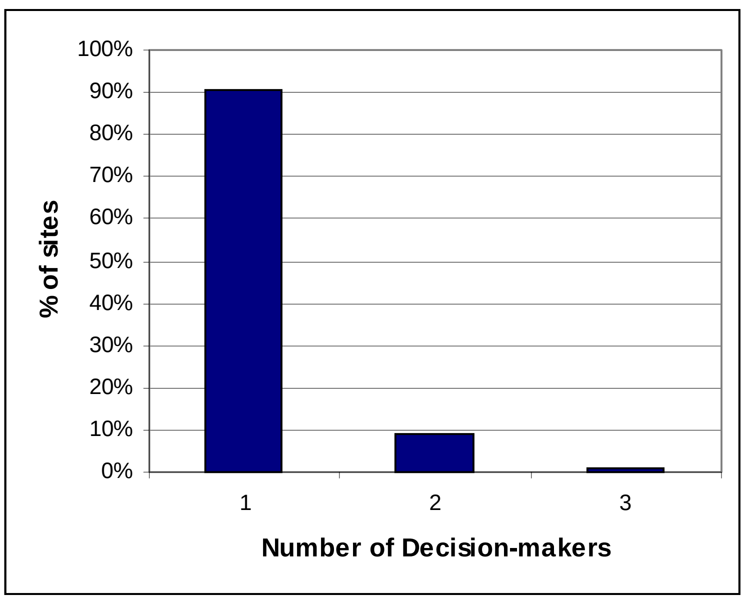

Figure 11: Number of Decision-Makers Needed to Complete Survey

Table 15 summarizes the minimum, maximum, median, and average number of people contacted and calls made.

	Decision-Makers	Calls – Surveyed Sites	Calls – All Sites
Average	1.1	3.54	2.39
Median	1	3	3
Minimum	1	1	1
Maximum	3	16	16

Table 15: Summary of Telephone Contacts

On-Site Surveys

The primary data source for the DOE-2 models was the on-site survey. The survey form was designed so that key modeling decisions on model zoning and equipment/space association were made by the surveyors in the field. The form was designed to follow the logical progression of an on-site survey process. The form started out with a series of interview questions. Conducting the interview first helped orient the surveyor to the building and allowed time for the surveyor to establish a rapport with the customer. Once the interview was completed, an inventory of building equipment was conducted. The survey started with the HVAC systems, and progressed from the roof and/or other mechanical spaces into the conditioned spaces. This progression allowed the surveyor to establish the linkages between the HVAC equipment and the spaces served by the equipment. The incented measures were identified during the on-site audit.

Interview Questions

The surveyor used the interview questions to identify building characteristics and operating parameters that were not observable during the course of the on-site survey. The interview questions covered the following topics:

Building functional areas. Functional areas were defined on the basis of operating schedules. Subsequent questions regarding occupancy, lighting, and equipment schedules, were repeated for each functional area.

Occupancy history. The occupancy history questions were used to establish the vacancy rate of the building during 1996. The questions covered occupancy, as a percent of total surveyed floor space, and HVAC operation during the tenant finish and occupancy of the space. Responses to these questions were used to understand building start-up behavior during the model calibration process.

Building Occupancy schedules. For each functional area in the building, a set of questions were asked to establish the building occupancy schedules. First, the surveyor assigned each day of the week to one of three daytypes: full occupancy, partial occupancy, and unoccupied. This was done to cover buildings that did not operate on a normal Monday through Friday workweek. Holidays and monthly variability in occupancy schedules were identified.

Daily schedules for occupants, interior lighting, and equipment/plug loads. A set of questions was used to establish hourly occupancy, interior lighting, and miscellaneous equipment and plug load schedules for each functional area in the building. During the on-site survey, the surveyor defined hourly schedules for each daytype. A value, which represents the fraction of the maximum occupancy and/or connected load was entered for each hour of the day. The entry of the schedule onto the form was done graphically.

Daily schedules of kitchen equipment. A set of questions were asked to establish hourly kitchen equipment schedules for each functional area in the building for each daytype. A value which represented the equipment-operating mode (off, idle, or low, medium or high volume production) was entered for each hour of the day. The entry of the schedule onto the form was done graphically.

Operation of other miscellaneous systems. General questions on the operation of exterior lighting systems, interior lighting controls, window shading, swimming pools, and spas were covered in this section.

Operation of the HVAC systems. A series of questions were asked to construct operating schedules for the HVAC systems serving each area. The surveyors entered fan operating schedules and heating and cooling setpoints. A series of questions were used to define the HVAC system controls. These questions were intended to be answered by someone familiar with the operation of the building mechanical systems. The questions covered operation of the outdoor air ventilation system, supply air temperature controls, VAV system terminal box type, chiller and chilled water temperature controls, cooling tower controls, and water-side economizers.

Building-wide water use. A series of questions were used to help calculate the service hot water requirements for the building.

Refrigeration system. The operation of refrigeration systems utilizing remote condensers, which are common in groceries and restaurants, was covered in this section. Surveyors divided the systems into three temperature classes, (low, medium and high) depending on the compressor suction temperature. For each system temperature, the refrigerant, and predominant defrost mechanism was identified. Overall system controls strategies were also covered.

Building Characteristics

The next sections of the on-site survey covered observations on building equipment inventories and other physical characteristics. Observable information on HVAC systems, building shell, lighting, plug loads, and other building characteristics were entered, as described below:

Built-up HVAC systems. Make, model number, and other nameplate data were collected on the chillers, cooling towers, heating systems, air handlers, and pumps in the building. Air distribution system type, outdoor air controls, and fan volume controls were also identified.

Packaged HVAC systems. Equipment type, make, model number, and other nameplate data were collected on the packaged HVAC systems in the building.

Zones. Based on an understanding of the building layout and the HVAC equipment inventory, basic zoning decisions were made by the surveyors according to the following criteria:

- ***Unusual internal gain conditions***. Spaces with unusual internal gain conditions, such as computer rooms, kitchens, laboratories were defined as separate zones.

- ***Operating schedules***. Occupant behavior varies within spaces of nominally equivalent use. For example, retail establishments in a strip retail store may have different operating hours. Office tenants may also have different office hours.
- ***HVAC system type and zoning***. When the HVAC systems serving a particular space were different, the surveyors sub-divided the spaces according to HVAC system type. If the space was zoned by exposure, the space was surveyed as a single zone, and a "zone by exposure" option was selected on the survey form.

For each zone defined, the surveyor recorded the floor area and occupancy type. Enclosing surfaces were surveyed, in terms of surface area, construction type code, orientation, and observed insulation levels. Window areas were surveyed by orientation. The surveyor also identified and inventoried basic window properties, interior and exterior shading devices, lighting fixtures and controls, and miscellaneous equipment and plug loads. Finally, the surveyor identified and entered zone-level HVAC equipment, such as baseboard heaters, fan coils, and VAV terminals.

Refrigeration systems. The surveyor inventoried the refrigeration equipment separately, and associated the equipment with a particular zone in the building. Refrigerated cases and stand-alone refrigerators were identified by case type, size, product stored, and manufacturer. Remote compressor systems were inventoried by make, model number, and compressor system type. Each compressor or compressor rack was associated with a refrigerated case temperature loop and heat rejection equipment such as a remote condenser, cooling tower, and/or HVAC system air handler. Remote condensers were inventoried by make, model number, and type. Nameplate data on fan and pump hp were recorded. Observations on condenser fan speed controls were also recorded.

Cooking Equipment. The surveyor recorded the cooking equipment separately and associated with a particular zone in the building. Major equipment was inventoried by equipment type (broiler, fryer, oven, and so on), size, and fuel type. Kitchen ventilation hoods were inventoried by type and size. Nameplate data on exhaust flowrate and fan hp were recorded and each piece of kitchen equipment was associated with a particular ventilation hood.

Hot Water/Pools. Water heating equipment was inventoried by system type, capacity, and fuel type. The surveyor recorded observations on delivery temperature, heat recovery, and circulation pump horsepower. Solar water heating equipment was inventoried by system type, collector area, and collector tilt and storage capacity. The surveyor inventoried pools and spas by surface area and location (indoor or outdoor). The filter pump motor horsepower was recorded, along with the surface area, collector type, and collector tilt angle data for solar equipment serving pools and/or spas.

Miscellaneous exterior loads. Connected load, capacity, and other descriptive data on elevators, escalators, interior transformers, exterior lighting, and other miscellaneous equipment were recorded.

Meter Numbers. Additional data were collected in the field to assist in the billing data account matching and model calibration process. This section served as the primary link between the on-site survey and billing data for non-participants. The surveyor recorded meter numbers for each meter serving the surveyed space. If the meter served space in addition to the surveyed space, the surveyor made a judgment on the ratio of the surveyed space to the space served by the meter.

Establishing Component Relationships

In order to create a DOE-2 model of the building from the various information sources contained in the on-site survey, relationships between the information contained in the various parts of the survey needed to be established. In the interview portion of the form, schedule and operations data were cataloged by building functional area. In the equipment inventory section, individual pieces of HVAC equipment: boilers, chillers, air handlers, pumps, packaged equipment and so on were inventoried. In the zone section of the survey, building envelope data, lighting and plug load data, and zone-level HVAC data were collected. The following forms provided the information needed by the software to associate the schedule, equipment, and zone information.

System/Zone Association Checklist. The system/zone association checklist provided a link between each building zone and the HVAC equipment serving that zone. Systems were defined in terms of a collection of packaged equipment, air handlers, chillers, towers, heating systems, and pumps. Each system was assigned to the appropriate thermal zones in accordance with the observed building design.

Interview "Area" / Audit "Zone" Association Checklist. Schedule and operations data gathered during the interview phase of the survey were linked to the appropriate building zone. These data were gathered according to the building functional areas defined previously. Each building functional area could contain multiple zones. This table facilitated the association of the functional areas to the zones, and thereby the assignment of the appropriate schedule to each zone.

Refrigerated Warehouses

Models of each facility were constructed from a combination of program documents and on-site surveys. Hard-copy program documents were also obtained from PG&E for each participant. The required program documentation included application forms, facility plans, building load calculations, equipment specification sheets, system operations manuals and proof of purchase documents. The refrigerated warehouse on-site survey was used to obtain the following information:

1. ***Verify facility design information***. Facility physical dimensions, equipment nameplate data, and other design parameters provided in the program file were field-verified. Additional facility description data required to develop the engineering model was collected.
2. ***Verify the installation of incented measures***. The surveyor identified all incented measures using the program files. The surveyor then physically counted the measures and compared nameplate data to program records.

3. ***Determine facility operation***. The facility operations data necessary to construct the engineering model was also collected. Interview questions identified facility operations parameters such as:

- Current operating hours
- Current operating months
- Future production and/or construction plans
- Product types received, receiving schedule, and product receiving temperature
- Product shipping schedule
- Process water flow schedules, temperature, and source (when heat recovery is used)
- Number and size of forklifts or other vehicles used, operating schedules
- Vehicle recharging schedules

During the facility walk-through portion of the on-site survey, additional equipment and facility operating parameters were observed. Such as:

- Space temperatures for coolers, freezers, loading vestibules, etc.
- Defrost schedules
- Suction pressures
- Minimum head pressure setpoints

These data were combined with the program information to construct a description of the design and operation of each participating refrigerated warehouse facility. Once the on-site surveys were conducted, an as-built TRNSYS model of each facility was constructed.

Short-term Metering

As a part of the overall modeling process, the DOE-2 simulations were calibrated to billing data. In order for a comparison between simulated electricity consumption and billing data to be meaningful, there needs to be a good match between the surveyed space and the space served by the PG&E meter. At selected sites where the surveyed space and the metered space did not match, short-term metering equipment was installed. An example of such a mismatch is a major tenant improvement or tenant finish in a multi-tenant building, where the PG&E revenue meter serves the entire space. Short-term metering equipment was installed on the circuits feeding the surveyed space only, thus serving as a temporary "proxy" meter for the surveyed and modeled space. These data were then used to calibrate the DOE-2 model for the site, instead of billing data.

During the on-site survey, the surveyors collected meter number information, and assessed the match between the space served by the meter(s) and the surveyed space. In situations where a poor match was evident, the surveyors assessed the feasibility of installing short-term metering equipment. The electrical panels serving the surveyed space were identified during the on-site survey. Sites with fairly "clean" circuitry, allowing metering with one or two watt transducers at the whole-panel or switchgear level were identified. If the site appeared to be a reasonable candidate, the surveyor recruited the site contact for short-term metering.

An electrical contractor was dispatched to install the watt transducers on the circuits or panels identified by the surveyor soon after the completion of the on-site survey. The data loggers collected "whole-building" or "whole-space" 5-minute interval demand data, which were averaged to hourly data. This data was then used to calibrate the DOE-2 models.

A total of 39 sites were identified as having a poor match between the surveyed space and the metered space. Of these, 14 sites were determined to be unsuitable by the surveyors, leaving 25 sites as short-term metering candidates. Ten sites were successfully recruited for short-term metering. The remaining sites were dropped from short-term metering because of:

- Customer refusal
- No response to further recruiting efforts
- Building circuiting not amenable to short-term metering of surveyed space
- Sub-metered data obtained from the customer

Of the ten sites scheduled, installation was successful at seven sites. Reasons for dropping sites during installation were:

- Unsafe installation (2 sites)
- Mixed circuits (1 site)

Of the 7 sites metered, 4 were refrigerated warehouses.

Data gathered for the monitored sites were used to calibrate the models. Lighting and occupancy schedules were inferred from the time-series profiles of the metered data. An example of a time-series plot is shown in Figure 11.

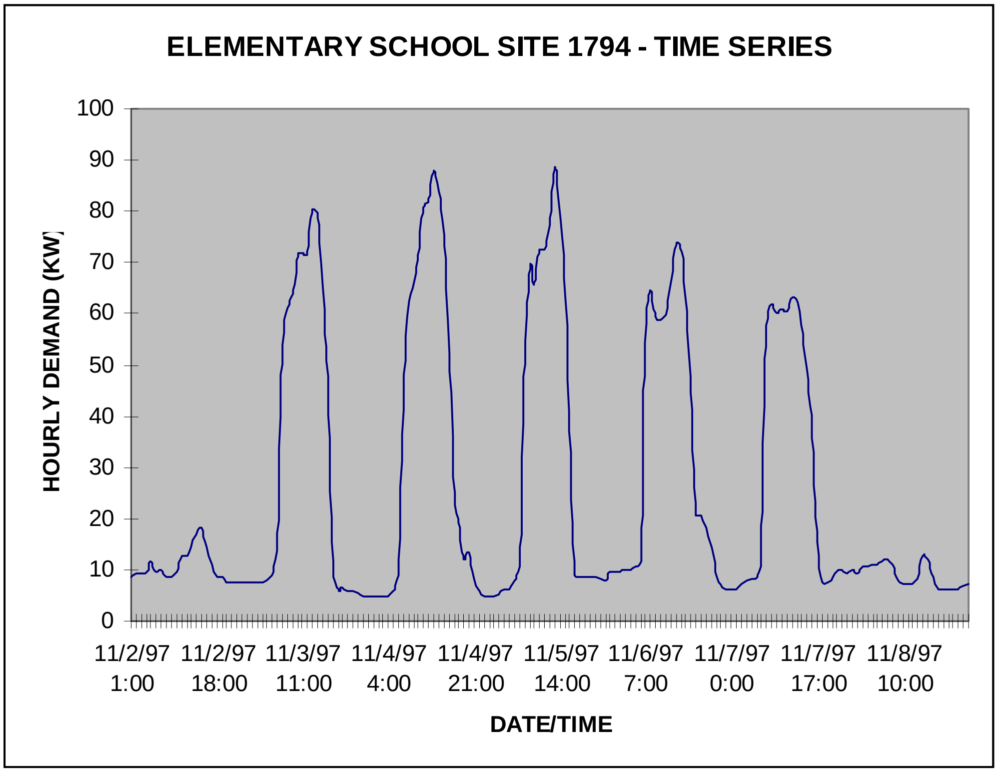

Figure 12: Short Term Monitored Time Series Data

The short-term metered data were also compared to the models on a temperature response basis. Daily average electricity consumption was plotted against daily average temperature for the monitored and simulated data. The models were calibrated to match the metered data to within ±10%, as shown in Figure 12 and Figure 13.

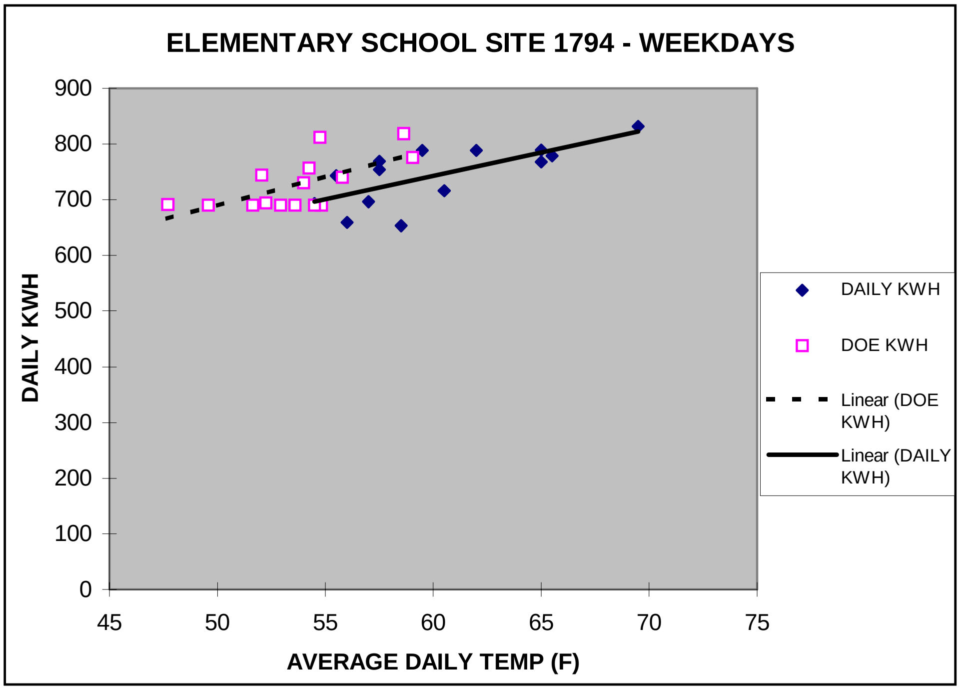

Figure 13: Short Term Temperature Response Comparison – Workdays

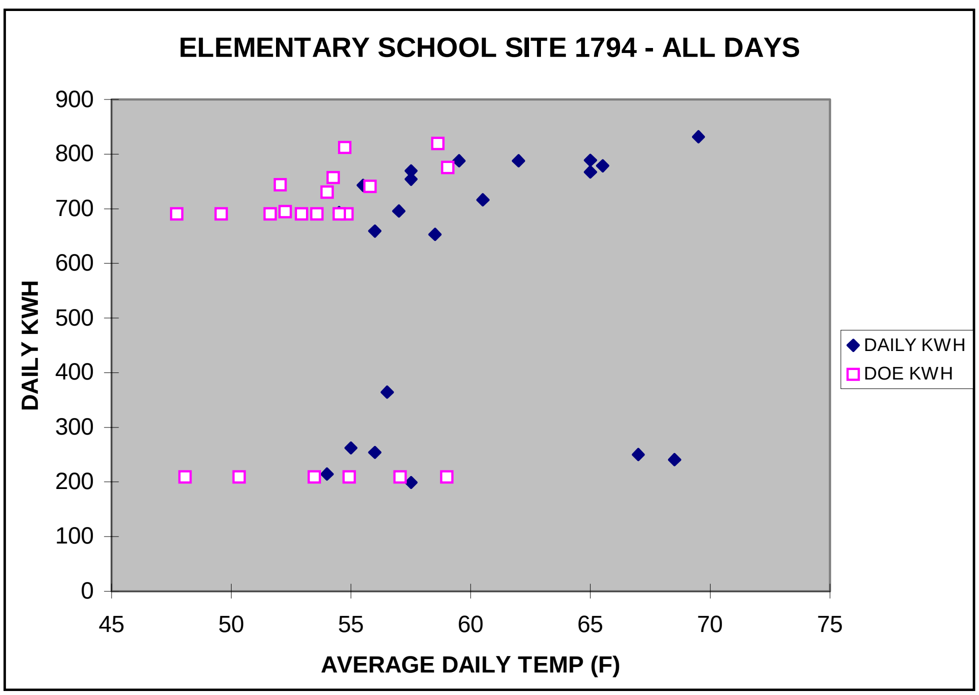

Figure 14: Short Term Temperature Response Comparison – All days

Engineering Models

An automated process was used to develop basic DOE-2 models from data contained in the on-site surveys, Title 24 compliance forms, program information and other engineering data. The modeling software took information from these data sources and created a DOE-2 model. The data elements used, default assumptions, and engineering calculations are described for the Loads, Systems, and Plant portions of the DOE-2 input file as follows.

Loads

Schedules were created for each zone in the model by associating the zones defined in the on-site survey with the appropriate functional area, and assigning the schedule defined for each functional area to the appropriate zone. Hourly schedules were created by the software on a zone-by-zone basis for:

- Occupancy
- Lighting
- Electric equipment
- Gas equipment (primarily kitchen equipment)
- Solar glare
- Window shading
- Infiltration

Occupancy, lighting, and equipment schedules. Each day of the week was assigned to a particular daytype, as reported by the surveyor. Hourly values for each day of the week were extracted from the on-site database according to the appropriate daytype. These values were modified on a monthly basis, according to the monthly building occupancy history.

Solar and shading schedules. The use of blinds by the occupants was simulated by the use of solar and shading schedules. The glass shading coefficient values were modified to account for the use of interior shading devices.

Infiltration schedule. The infiltration schedule was established from the fan system schedule. Infiltration was scheduled “off” during fan system operation, and was scheduled “on” when the fan system was off.

Shell materials. A single-layer, homogeneous material was described which contains the conductance and heat capacity properties of the composite wall used in the building. The thermal conductance and heat capacity of each wall and roof assembly was taken from the Title 24 documents, when available. If the Title 24 documents were not available, default values for the conductance and heat capacity were assigned from the wall and roof types specified in the on-site survey, and the observed R-values. If the R-values were not observed during the on-site survey and the Title 24 documents were not available, an “energy-neutral” approach was taken by assigning the same U-value and heat capacity for the as-built and Title 24 simulation runs.

Windows. Window thermal and optical properties from the building drawings or Title 24 documents (when available) were used to develop the DOE-2 inputs. If these documents were not available, default values for the glass conductance were assigned according to the glass type specified in the on-site survey. If the glass type was not observed during the on-site survey and the Title 24 documents were not available, an "energy-neutral" approach was taken by assigning the same U-value and shading coefficient for the as-built and Title 24 simulation runs.

Lighting kW. Installed lighting power was calculated from the lighting fixture inventory reported on the survey. A standard fixture wattage was assigned to each fixture type identified by the surveyors. Lighting fixtures were identified by lamp type, number of lamps per fixture, and ballast type as appropriate.

Lighting controls. The presence of lighting controls was identified in the on-site survey. For occupancy sensor and lumen maintenance controls, the impact of these controls on lighting consumption was simulated as a reduction in connected load, according to the Title 24 lighting control credits. Daylighting controls were simulated using the "functions" utility in the load portion of DOE-2. Since the interior walls of the zones were not surveyed, it was not possible to use the standard DOE-2 algorithms for simulating the daylighting illuminance in the space. A daylight factor, defined as the ratio of the interior illuminance at the daylighting control point to the global horizontal illuminance was estimated for each zone subject to daylighting control. Typical values for sidelighting applications were used as default values. The daylight factor was entered into the function portion of the DOE-2 input file. Standard DOE-2 inputs for daylighting control specifications were used to simulate the impacts of daylighting controls on lighting schedules. The default daylight factors were adjusted during model calibration.

Equipment kW. Connected loads for equipment located in the conditioned space, including miscellaneous equipment and plug loads, kitchen equipment and refrigeration systems with integral condensers were calculated. Input data were based on the "nameplate" or total connected load. The nameplate data were adjusted using a "rated-load factor," which is the ratio of the average operating load to the nameplate load during the definition of the equipment schedules. This adjusted value represented the hourly running load of all equipment surveyed. Equipment diversity was also accounted for in the schedule definition.

For the miscellaneous equipment and plug loads, equipment counts and connected loads were taken from the on-site survey. When the connected loads were not observed, default values based on equipment type were used.

For the kitchen equipment, equipment counts and connected loads were taken from the on-site survey. Where the connected loads were not observed, default values based on equipment type and "trade size" were used. Unlike the miscellaneous plug load schedules, the kitchen equipment schedules were defined by operating regime. An hourly value corresponding to "off", "idle", or "low," "medium," or "high" production rates were assigned by the surveyor. The hourly schedule was developed from the reported hourly operating status and the ratio of the hourly average running load to the connected load for each of the operating regimes.

For the refrigeration equipment, refrigerator type, count, and size were taken from the on-site survey. Equipment observed to have an "integral" compressor/condenser that is, equipment that rejects heat to the conditioned space, were assigned a connected load per unit size.

Source input energy. Source input energy represented all non-electric equipment in the conditioned space. In the model, the source type was set to natural gas, and a total input energy was specified in terms of Btu/hr. Sources of internal heat gains to the space that were not electrically powered include kitchen equipment, dryers, and other miscellaneous process loads. The input rating of the equipment was entered by the surveyors. As with the electrical equipment, the ratio of the rated input energy to the actual hourly consumption was calculated by the rated load factor assigned by equipment type and operating regime.

Heat gains to space. The heat gains to space were calculated based on the actual running loads and an assessment of the proportion of the input energy that contributed to sensible and latent heat gains. This in turn depended on whether or not the equipment was located under a ventilation hood.

Spaces. Each space in the DOE-2 model corresponded to a zone defined in the on-site survey. In the instance where the "zoned by exposure" option was selected by the surveyor, additional DOE-2 zones were created. The space conditions parameters developed on a zone by zone basis were included in the description of each space. Enclosing surfaces, as defined by the on-site surveyors, were also defined.

Systems

This section describes the methodology used to develop DOE-2 input for the systems simulation. Principal data sources include the on-site survey, Title 24 documents, manufacturers' data, and other engineering references as listed in this section.

Fan schedules. Each day of the week was assigned to a particular daytype, as reported by the surveyor. The fan system on and off times from the on-site survey was assigned to a schedule according to daytype. These values were modified on a monthly basis, according to the monthly HVAC operating hour adjustment. The on and off times were adjusted equally until the required adjustment percentage was achieved. For example, if the original schedule was "on" at 6:00 hours and "off" at 18:00 hours, and the monthly HVAC adjustment indicated that HVAC operated at 50% of normal in June, then the operating hours were reduced by 50% by moving the "on" time up to 9:00 hours and the "off" time back to 15:00 hours.

Setback schedules. Similarly, thermostat setback schedules were created based on the responses to the on-site survey. Each day of the week was assigned to a particular daytype. The thermostat setpoints for heating and cooling, and the setback temperatures and times were defined according to the responses. The return from setback and go to setback time was modified on a monthly basis in the same manner as the fan-operating schedule.

Exterior lighting schedule. The exterior lighting schedule was developed from the responses to the on-site survey. If the exterior lighting was controlled by a time clock, the schedule was used as entered by the surveyor. If the exterior lighting was controlled by a photocell, a schedule, which follows the annual variation in daylength, was used.

System type. The HVAC system type was defined from the system description from the on-site survey. The following DOE-2 system types were employed:

- Packaged single zone (PSZ)
- Packaged VAV (PVAVS)
- Packaged terminal air conditioner (PTAC)
- Water loop heat pump (HP)
- Evaporative cooling system (EVAP-COOL)
- Central constant volume system (RHFS)
- Central VAV system (VAVS)
- Central VAV with fan-powered terminal boxes (PIU)
- Dual duct system (DDS)
- Multi-zone system (MZS)
- Unit heater (UHT)
- Four-pipe fan coil (FPFC)

Packaged HVAC system efficiency. Manufacturers' data were gathered for the equipment surveyed based on the observed make and model number. A database of equipment efficiency and capacity data was developed from an electronic version of the ARI rating catalog. Additional data were obtained directly from manufacturers' catalogs, or the on-line catalog available on the ARI website (www.ari.org). Manufacturers' data on packaged system efficiency is a net efficiency, which considers both fan and compressor energy. DOE-2 requires a specification of packaged system efficiency that considers the compressor and fan power separately. Thus, the manufacturers' data were adjusted to prevent "double-accounting" of fan energy, according to the procedures described in the 1995 Alternate Compliance Method (ACM) manual.

Pumps and fans. Input power for pumps, fans and other motor-driven equipment was calculated from motor nameplate hp data. Motor efficiencies as observed by the surveyors were used to calculate input power. In the absence of motor efficiency observations, standard motor efficiencies were assigned as a function of the motor hp, RPM and frame type. A rated load factor was used to adjust the nameplate input rating to the actual running load. For VAV system fans, custom curves were used to calculate fan power requirements as a function of flow rate in lieu of the standard curves used in DOE-2, as described in the 1995 ACM manual.

Refrigeration systems. Refrigeration display cases and/or walk-ins were grouped into three systems defined by their evaporator temperatures. Ice cream cases were assigned to the lowest temperature circuit, followed by frozen food cases, and all other cases. Case refrigeration loads per lineal foot were taken from manufacturers' catalog data for typical cases. Auxiliary energy requirement data for evaporator fans, anti-sweat heaters, and lighting were also compiled from manufacturers' catalog data. Model inputs were calculated based on the survey responses. For example, if the display lighting was surveyed with T-8 lamps, lighting energy requirements appropriate for T-8 lamps were used to derive the case auxiliary energy input to DOE-2.

Compressor EER data were obtained from manufacturers' catalogs as a function of the suction temperatures corresponding to each of the three systems defined above. These data were used to create default efficiencies for each compressor system. Custom part-load curves were used to simulate the performance of parallel-unequal rack systems.

Total heat of rejection (THR) data at design conditions were obtained for refrigeration system condensers from manufacturers' data. These data were used to calculate hourly approach temperatures and fan energy using the enhanced refrigeration condenser algorithms in DOE-2.1 E version 119.

Service hot water. Service hot water consumption was calculated based on average daily values from the 1995 ACM for various occupancy types. Equipment capacity and efficiency were assigned based on survey responses.

Exterior lighting. Exterior lighting input parameters were developed similarly to those for interior lighting. The exterior lighting connected load was calculated from a fixture count, fixture identification code and the input wattage value associated with each fixture code.

Plant

This section describes the methodology used to develop DOE-2 input for the plant simulation. Principal data sources included the on-site survey, Title 24 documents, manufacturers' data, program data, and other engineering references.

Chillers. The DOE-2 input parameters required to model chiller performance included chiller type, full-load efficiency and capacity at rated conditions, and performance curves to adjust chiller performance for temperature and loading conditions different from the rated conditions. Chiller type was assigned based on the type code selected during the on-site survey. Surveyors also gathered chiller make, model number, and serial number data. These data were used to develop performance data specific to the chiller installed in the building. Program data and/or manufacturers' data were used to develop the input specifications for chiller efficiency.

Cooling towers. Cooling tower fan and pump energy was defined based on the nameplate data gathered during the on-site survey. Condenser water temperature and fan volume control specifications were derived from the on-site survey responses.

Refrigerated Warehouses

A combination of engineering techniques was used to calculate the energy performance and energy savings of refrigerated warehouses. The DOE-2.2 and TRNSYS transient simulation programs were used in tandem to create the engineering models. The DOE-2.2 program was used to calculate hourly facility refrigeration loads. The TRNSYS program was used to simulate the performance of specialized refrigeration equipment such as industrial refrigeration evaporators, defrost systems, evaporative condensers, and industrial refrigeration compressor systems.

Models of each facility were constructed from a combination of design documents and on-site surveys. Hard-copy program documents were obtained from PG&E, which included application forms, facility plans, building load calculations, equipment specification sheets, system operations manuals and proof of purchase documents. Similar documentation was obtained for non-participants during the on-site survey and subsequent interviews with facility designers.

An on-site survey was conducted for each sampled site. The on-site survey was used to obtain the following information:

1. ***Verify facility design information***. Facility physical dimensions, equipment nameplate data, and other design parameters were field-verified. Additional facility description data required to develop the engineering model were collected.
2. ***Verify the installation of incented measures***. All incented measures were identified, and the physical count and nameplate data were compared to program records.
3. ***Determine facility operation***. The facility operations data necessary to construct the engineering model were collected. Interview questions identified facility operations parameters such as:
    - Current operating hours
    - Current operating months
    - Future production and/or construction plans
    - Product types received, receiving schedule, and product receiving temperature
    - Product shipping schedule
    - Process water flow schedules, temperature, and source (when heat recovery is used)
    - Number and size of forklifts or other vehicles used, operating schedules
    - Vehicle recharging schedules

      During the facility walk-through portion of the on-site survey, additional equipment and facility operating parameters were observed such as:
    - Space temperatures for coolers, freezers, loading vestibules, etc.
    - Defrost schedules
    - Suction pressures
    - Minimum head pressure setpoints

These data were used to construct a description of the design and operation of each refrigerated warehouse facility. Once the on-site surveys were conducted, an engineering model of each facility was constructed.

Gross savings calculations

The as-built performance of the facility was calculated from the facility characteristics verified during the on-site survey. Since there are no energy standards for refrigerated warehouses, the PG&E program baseline equipment specifications as reported in the Advice Filings served as the baseline or reference point for the gross impact calculations. Gross savings for each participant and non-participant warehouse were calculated from the difference in the energy consumption between the facility modeled with the baseline specifications and the facility modeled with the as-built efficiency specifications. The refrigerated warehouse baseline specifications are summarized in Table 16. The PG&E program minimum requirements for pipe insulation were less stringent than the baseline level established for the study. In other words, a refrigerated warehouse who only installed the minimum required pipe insulation would have negative savings.

Attribute	Application	Baseline Characteristics	Program Minimum	Incentive Levels	Comments	Reference
Lighting	All refrigerated space	Not addressed	0.6 W/SF	none	Since no incentives paid, installed lighting will be held energy-neutral.	
Roof Insulation	Cooler	R-30	R-30	R-40 - R-50	Baseline = program minimum	Advice Filing NRNC- A - A7
	Freezer	R-45	R-45	R-50 - R-100	Baseline = program minimum	Advice Filing NRNC- A - A7
Wall Insulation	Cooler	R-25	R-25	R-35 - R-45	Baseline = program minimum	Advice Filing NRNC- A - A7
	Freezer	R-35	R-35	R-40 - R-60	Baseline = program minimum	Advice Filing NRNC- A - A7
Vessel insulation	Cooler	R-10	R-11	R-16		Advice Filing NRNC- A - 40
	Freezer	R-17	R-14	R-24	Baseline higher than program minimum	Advice Filing NRNC- A - 41
Pipe insulation	Cooler - pipe dia .5 - 1.5 in. pipe dia 2 - 5 in. pipe dia 6 - 12 in.	R-6 R-9 R-10	R-3.5 R-5.5 R-5.5	R-5 R-8 R-11	Baseline higher than incentive levels Baseline higher than incentive levels Baseline higher than program minimum	Advice Filing NRNC- A - 40
	Freezer - pipe dia .5 - 1.5 in. pipe dia 2 - 5 in. pipe dia 6 - 12 in.	R-9 R-14 R-15	R-5 R-8 R-8	R-8 R-11 R-16	Baseline higher than incentive levels Baseline higher than incentive levels Baseline higher than program minimum	Advice Filing NRNC- A - 40
Doors	Forklift doors - open to ambient	Slow-closing automatic door, 14 second cycle time.	None	Quick-close door		Advice Filing NRNC- A - 42
	Forklift doors - open to adjacent space	Open door with strip curtain	None	Quick-close door		Pers comm, Stan Tory
	Material pass-through doors	Open door with strip curtain	None	Quick-close door	50% reduction in door use and infiltration	Pers comm, Stan Tory
Evaporators	Fan control	One-speed	None	Two speed, VSD		Advice Filing NRNC- A - 44
	Fan power	0.39 hp/ton	None	0.3 hp/ton		Advice Filing NRNC- A - 44
	Motor efficiency	Standard efficiency	None	High efficiency		Advice Filing NRNC- A - 44
	Approach temperature	20 °F	None	8 °F		Advice Filing NRNC- A - 44

Table 16: Refrigerated Warehouse Baseline Specifications

Attribute	Application	Baseline Characteristics	Program Minimum	Incentive Levels	Comments	Reference
Low temperature piping design	Systems with loads at different temperatures	Lowest value for all evaporators	None	Separate low temp suction line	Second system < -25°F SST, > 10°F below initial system	
Pipe sizing	Suction line pressure drop	0.5 psi/100 ft, max of 2.0 total	None	Upsize one pipe diameter		Advice Filing NRNC-A-F12
	Discharge line pressure drop	1.5 psi/100 ft, max of 3 total	None	Upsize one pipe diameter		Advice Filing NRNC-A-F12
Liquid sub-cooling	High pressure liquid	No sub-cooling	None	5 °F difference between refrigerant and cooling water		
Evaporative condensers	Approach temperature	20 °F	10 °F	Same as program minimum		Advice Filing NRNC - A56
	Minimum condensing temperature	75 °F	60 °F	Same as program minimum		Advice Filing NRNC - A56
	Condensing temperature control	Pressure control	Wet-bulb control for systems > 300 T	Same as program minimum	Program minimum and incentive level is press control for systems < 300 T	
	Motor efficiency	Standard	Energy-efficient	Same as program minimum		Advice Filing NRNC - A55
	Fan control	One-speed	Two speed	Same as program minimum		Advice Filing NRNC - A55
	Fan and pump power	0.09 hp/ton	0.11 hp/ton	Same as program minimum	Lower condensing temp makes up for higher fan hp	Pers comm., Stan Tory
Compressors	Efficiency	Stock compressor bhp/ton from manufacturer.	None	4% improvement over stock compressor efficiency		Manufacturers' data, program documents
	Motor efficiency	Standard efficiency	None	Premium efficiency		Advice Filing NRNC - A-54
	Oil cooling	Liquid-injection	Thermo-syphon oil cooling > 300 T	Thermo-syphon oil cooling all sizes T	Must use thermosyphon oil cooling to get compressor incentive	Advice Filing NRNC - A-54
Battery chargers		Ferro-resonant battery charger with manual timer	None	Select from list of qualifying models		

Table 18 (con't): Refrigerated Warehouse Baseline Specifications

Model Calibration

An integral part of DOE-2 model development was the model calibration process. Monthly energy consumption and demand from the DOE-2 models was compared to billing data for the same period to assess the reasonableness of the models. Changes were made to a fixed set of calibration parameters until the models matched the billing data. The goal of the calibration process was to match billing demand and energy data within ± 10 percent on a monthly basis. The overall model calibration process consisted of the following steps:

1. Review and format billing data. Billing data as received from PG&E were reformatted as required by the model calibration software.
2. Select relevant accounts. For many of the sites, a number of accounts were provided. Account information such as customer name, address, business type, and meter number was compared to the on-site survey information. The list of accounts that seemed to best match the surveyed space was selected.
3. Assign surveyed to metered space percentage. During the on-site survey, the surveyors were asked to assess the ratio of the space surveyed to the space served by the building meter(s). Billing data records were adjusted to reflect the portion of the metered data that applied to the modeled space.
4. Run model. The as-built model was run with actual 1996 and 1997 weather data applicable to the particular site, using the occupancy as reported by the surveyors. Annual simulations for both years were done, and the modeled consumption and demand were aggregated to correspond to the meter read dates from the billing data. The 1997 calibration covered billing data and simulated energy consumption for the first six months of the year.
5. Review kWh and kW comparison. The modeled and metered consumption and demand for each billing period were compared using a graphical data visualization tool. An example output screen from the calibration tool is shown in Figure 14.
6. Reject unreasonable or faulty billing data. Some of the billing data received was incomplete or not well matched to the modeled space. In these cases, the billing data were rejected, and the models were not calibrated.
7. Make adjustments to calibration variables. A fixed set of calibration variables was provided to the modeling calibration team. The calibration parameters, and the range of acceptable adjustments are shown in Table 17. The modelers adjusted the calibration parameters until the modeled results matched the metered results within ± 10 percent for each billing period. This was an iterative process, involving changing the model inputs, repeating the simulation, and reviewing the results. At each iteration, the changes made to the model and the impacts of the change on the model vs. billing data comparison were entered into a calibration log file.

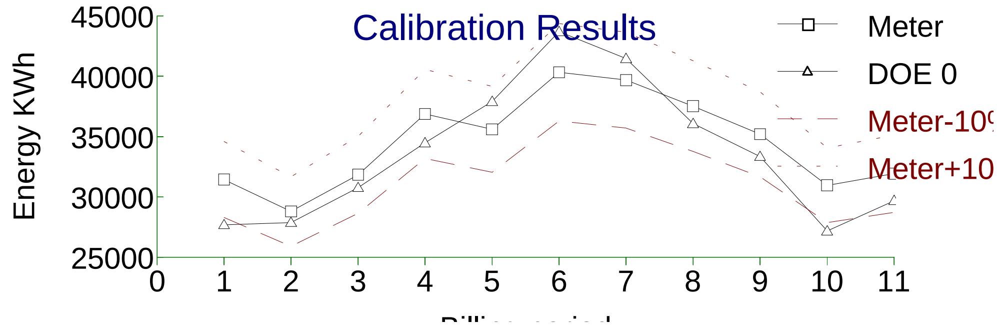

Figure 15: Example Calibration Tool Screen

Calibration Parameter	Adjustment range
Monthly schedule multiplier	.2 – 2
Lighting diversity multiplier	.2 – 2
Plug load diversity multiplier	.2 – 5
Plug load internal heat gains multiplier	.2 – 5
Heating thermostat setpoint	± 5°F
Cooling thermostat setpoint	± 5°F
DHW water use multiplier	.1 – 10
Minimum outside air ratio	.1 - .7, if no additional information
Refrigeration compressor efficiency	± 20%
Heating supply air temp control	discrete choices
Direct evaporative system effectiveness	0.2 - 0.8
Indirect evaporative system effectiveness	0.2 - .07
Heat pump defrost control	discrete choices
Daylight factor	look at hourly reports to verify correct operation
Building azimuth	± 45 degrees

Table 17: Model Calibration Parameters and Acceptable Adjustment Range

In some cases, it was not possible to calibrate the models. When billing or short-term metering data were not available, the modeled results were examined for reasonableness, in terms of annual energy consumption (kWh/SF) by building type and end-use percentage of total consumption. Even when billing data were available, some of the models resisted reasonable attempts to achieve calibration. Rather than making unreasonable adjustment to the models, the models were left un-calibrated or partially calibrated. During calibration, the models were run with actual year weather data provided by PG&E from 32 local weather stations located throughout their service territory.

The results of the model calibration process are shown in Figure 15. The modelers were able to successfully calibrate 58% of the models. We were unable obtain billing data for 20% of the sites. A total of 22% of the models resisted reasonable attempts at calibration. In other words, for 22% of the sites, billing data were available but the model could not be brought into agreement with the data by making reasonable modifications to the model.

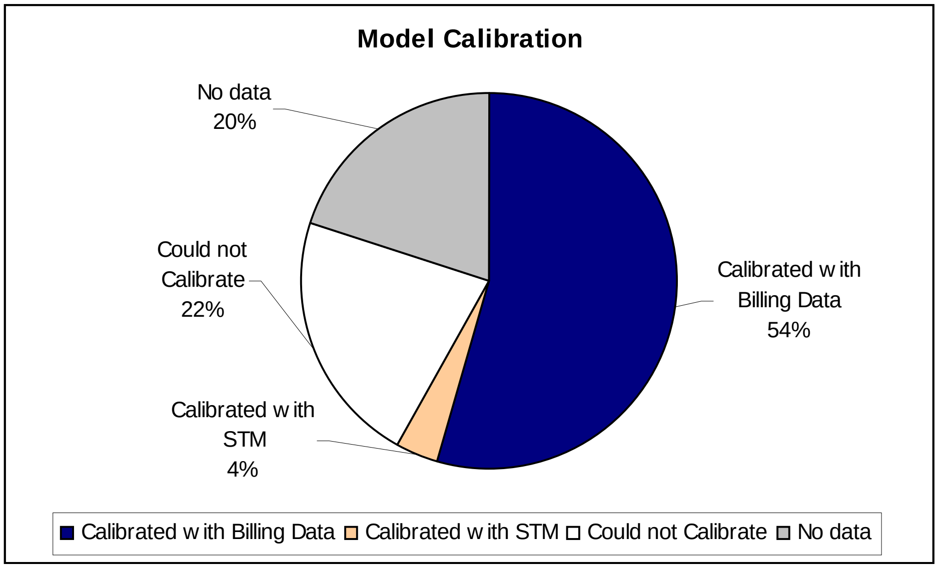

Figure 16: Model Calibration Results

Effects of Model Calibration

To understand the effect of calibrating the models to available billing or short-term metering data, models that were successfully calibrated were projected to the population and compared. That is, only the models that were ultimately calibrated were used in this test. Overall, model calibration had the effect of changing the measured savings by 2.8%.

The frequency of calibration actions taken by the modelers is shown in Figure 16. Note that plug load diversity multiplier adjustments were the most common changes made during the calibration process.

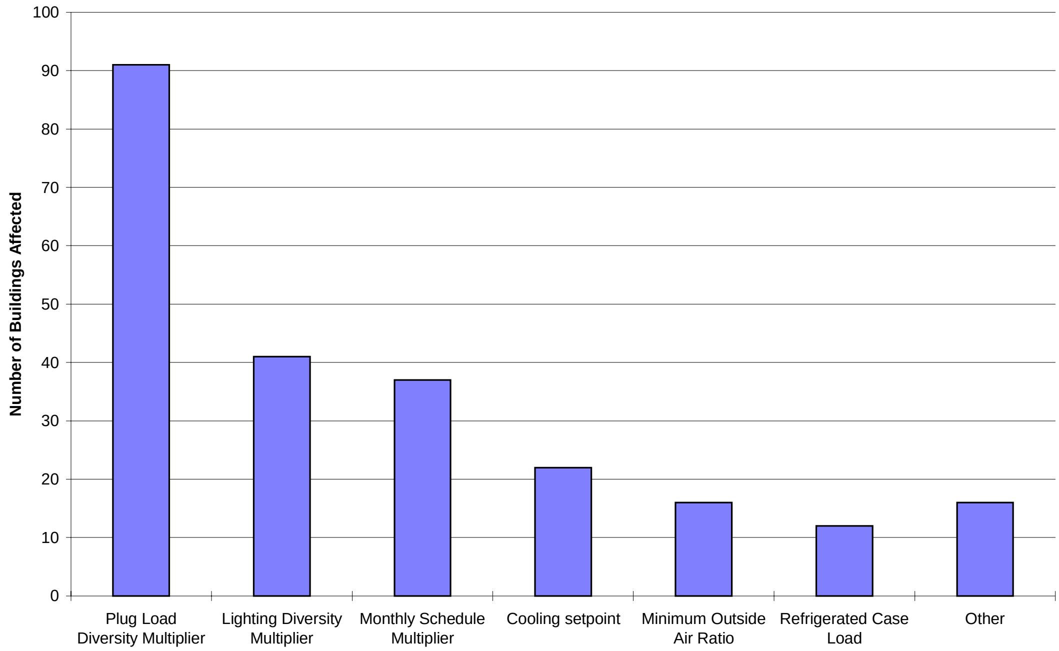

Figure 17: Frequency of Calibration Actions

The average initial and final values for the most common calibration variables are shown in Table 18.

Calibration Variable	Average Initial Value	Average Final Value
Plug load diversity multiplier	1	2.76
Lighting diversity multiplier	1	1.35
Monthly schedule multiplier	1	1.19
Cooling setpoint	75.2°F	73.3°F
Outdoor air fraction	0.12	0.43
Refrigerated case load adjustment	1	1.42

Table 18: Initial and Final Calibration Variables

The plug load diversity multiplier showed the largest average change (276%) of the set of most common calibration variables. Plug loads were not extensively surveyed, since plug load energy consumption was not addressed by the program or Title 24. The uncertainty in the calculated plug load density and schedule diversity was high, as was the influence of plug loads on total building consumption and demand. However, the impact of plug loads on calculated energy savings was minor.

Model Review and Quality Control

The on-site survey data entry program contained numerous quality control (QC) checks designed to identify invalid building characteristics data during data entry. Once the data were entered, the models were run and the results were reviewed by the surveyor/modeler and senior engineering staff. A building characteristics and model results summary report was created for each site. The model results were compared to a set of QC criteria as shown in Table 19. Data falling outside of the QC range were validated during the QC process.

Building Parameter	Range	Definition
Lighting Power Density	0.9 - 1.9	building wide average
Equipment Power Density	0.1 - 5	building wide average
Cooling Ratio	95 - 200%	capacity from annual run / capacity from sizing run
Cooling EER	8 - 14	capacity weighted cooling efficiency
Wall-U	0.5 - 0.033	area weighted average, includes air film
Roof-U	0.5 - 0.033	area weighted average, includes air film
Win-U	0.3 - 0.88	area weighted average, includes air film
Win-Shading Coefficient	0.35 - 0.88	area weighted average
Win Area	0 - 70%	Percentage of gross wall area associated w/windows, expressed as a true percentage 0 –100
Sky-U	0.3 - 0.9	area weighted average of glazing contained in roof
Sky-Shading Coefficient	0.35 - 0.88	area weighted SC for all horizontal glazing
Sky-Area	0 - 10%	Percentage of gross roof area associated with sky light, expressed as a true percentage 0 –100
LTG Occ	0 - 50%	Percentage of lighting watts controlled by occupancy sensors, expressed as a true percentage 0 –100
LTG DayL	0 - 50%	Percentage of lighting watts controlled by daylighting sensors, expressed as a true percentage 0 –100
Measures only savings relative to program expectations (participants only)	50% - 150%	measures-only savings / program expectations
Total savings (all sites)	0% - 50%	Savings expressed as a percentage of baseline energy consumption

Table 19: Model Quality Control Criteria

Parametrics

Once the models were calibrated and quality checked, a batch process was used to create a series of parametric simulation runs. These runs were used to simulate gross savings for participants and non-participants on a whole-building and measure-class basis. The parametric runs performed for this study are listed below:

As-Built Parametric Run.

Once the models were completed, checked for reasonableness, and/or calibrated, the as-built parametric runs were done. Monthly schedule variations resulting from partial occupancy and building startup were eliminated, and the models were run using long-term average weather data from the National Weather Service.

Baseline Parametric Run.

Key building performance parameters were reset to a baseline condition to calculate gross energy savings for participants and non-participants. The California Building Energy Efficiency Standard (Title 24) was the primary reference for establishing baseline performance parameters. Title 24 specifies minimum specifications for building attributes such as:

- Opaque shell conductance
- Window conductance
- Window shading coefficient
- HVAC equipment efficiency
- Lighting power density

Title 24 applied to most of the building types covered in the programs covered under this evaluation, with the exception of:

- Hospitals
- Unconditioned space (including warehouses)

Incentives were also offered by the programs for building attributes not addressed by Title 24. In situations where Title 24 does not address building types or equipment covered under the program, baseline parameters equivalent to those used for the program baseline efficiencies were used.

Envelope

Opaque shell U-values were assigned based on Title 24 requirements as a function of climate zone and heat capacity of the observed construction. For windows, Title 24 specifications for maximum relative solar heat gain were used to establish baseline glazing shading coefficients. Fixed overhangs were removed from the baseline building. Glass conductance values as a function of climate zone were applied. For skylights, shading coefficients and overall conductance were also assigned according to climate zone.

Mechanical

Baseline specifications for HVAC equipment efficiency were derived from the Title 24 requirements as a function of equipment type and capacity. Maximum power specifications for fans were established based on Title 24 requirements, which address fan systems larger than 25 hp. Specific fan power was held energy neutral (as built W/CFM = baseline W/CFM) for fan systems under 25 hp. Additionally, all systems larger than 2500 CFM (except for hospitals) were simulated with economizers in the baseline run. All VAV fan systems larger than 50 hp were simulated with inlet vane control. All variable-volume pumps were simulated with throttling valve control.

HVAC system sizing

HVAC system sizing for the as-built case was determined by direct observation of the nameplate capacities of the HVAC equipment. The installed HVAC system capacity was compared to the design loads imposed on the system to determine a sizing ratio for the as-built building. Once established, the sizing ratio was held constant for each subsequent DOE-2 run. A separate sizing run was done prior to the baseline and parametric runs. The peak cooling system size was calculated using the equipment sizing algorithms in DOE-2. The system capacity was reset using the calculated peak cooling capacity, and the as-built sizing ratio. A new system size was calculated for the baseline run and each parametric run.

Lighting

The Title 24 area category method was used to set the baseline lighting power for each zone as a function of the observed occupancy. Task lighting was not included in the baseline lighting calculation, and exit signs were reset to the program baseline (40 W/exit sign). A lighting power density appropriate for corridor/restroom/support areas was assigned according to the portion of each space allocated to these areas. All lighting controls were turned off for the baseline simulation.

Additional Parametric Runs

Once the as-built and baseline building models were defined, an additional set of parametric runs were done to estimate the program impact on the lighting, HVAC, and shell / daylighting end-uses. The baseline model was returned to the as-built design in a series of steps outlined as follows:

1. Shell, measures only – Baseline envelope properties (glazing U-value and shading coefficient; and opaque surface insulation) for incented measures only were returned to their as-built condition.
2. All Shell – All baseline envelope properties were returned to their as-built condition.
3. Lighting, measures only – Run 2 above, plus baseline lighting power densities and controls for spaces in the building that received incentives were returned to their as-built condition.
4. All Lighting – Run 2 above, plus all baseline lighting power densities and controls were returned to their as-built condition.

5. Motors and Air Distribution, measures only – Run 4 above, plus baseline motor efficiency and fan power indices (W/CFM) for incented measures only returned to their as-built condition.
6. All Motors and Air Distribution – Run 4 above, plus all baseline motor efficiency and fan power indices (W/CFM) returned to their as-built condition.
7. HVAC, measures only. Run 6 above, plus HVAC parameters for incented measures only returned to their as-built condition.
8. All HVAC – Run 6 above, plus all HVAC parameters returned to their as-built condition. This run is equivalent to the full as-built run.
9. Refrigeration, measures only – Run 8 above, plus refrigeration parameters for incented measures in buildings eligible for the grocery store refrigeration and refrigerated warehouse programs only returned to their as-built condition.
10. All Refrigeration – Run 8 above, plus all refrigeration parameters in buildings eligible for the grocery store refrigeration and refrigerated warehouse programs returned to their as-built condition. This run is equivalent to the full as-built run. Note: refrigeration parameters in buildings not eligible for the grocery store refrigeration and refrigerated warehouse programs remained at the as-built level for all parametric runs.

Gross Savings

This section presents the gross energy and demand savings estimates of participants. Savings findings for the whole building as well as for shell, lighting, motors, HVAC, and refrigeration end-uses are reported.

Some definitions would be helpful to clarify the discussion.

Baseline	A consistent standard of energy efficiency against which all buildings were measured. This was defined as the output of a DOE-2.1E simulation of a building using Title 24 required equipment efficiencies (where applicable) run using the operating schedule found by the on-site surveyor. Where Title 24 did not apply (e.g. hospitals), the baseline that was defined by the program for estimating the program savings was used.
As Built	A DOE-2.1E simulation of a building using all equipment and operating parameters as found by an on-site surveyor.
Savings	The difference between baseline and as built. Positive savings indicate that the building was more efficient – used less energy -- than its base case.
"Better than baseline"	The as built simulation showed less energy consumption than the baseline simulation – more efficient than the base case. Positive savings.
"Worse than baseline"	The as built simulation showed more energy consumption than the baseline simulation – less efficient than the base case. Negative savings.
Costing period	PG&E defined time periods for reporting energy usage. See Table 20 for description of each period.

Period	Dates	Days / Times
Summer On-peak	May 1to October 31	Weekdays 12 pm to 6 pm
Summer Part-peak	May 1 to October 31	Weekdays 8:30 am to 12 pm and 6 pm to 9:30 pm
Summer off-peak	May 1 to October 31	Weekdays 9:30 pm to 8:30 am. All day weekends and holidays
Winter part-peak	November 1 to April 30	Weekdays 8:30 am to 9:30 pm
Winter Off-peak	November 1 to April 30	Weekdays 9:30 pm to 8:30 am. All day weekends and holidays.

Table 20: Costing Periods

Methodology

This project used a statistical methodology called Model-Based Statistical Sampling or MBSS™. MBSS has been used for many evaluation studies to select the sites or projects to be studied and to extrapolate the results to the target population. MBSS has been used for NEES, Northeast Utilities, Consolidated

Edison, The New York Power Authority, Wisconsin Electric, Sierra Pacific Power Company, and Washington Power and Light among others. MBSS was used in the end-use metering component of the 1992 evaluation of PG&E's CIA program and the 1994 NRNC evaluations for PG&E and Southern California Edison. A complete description of MBSS methodology is available if further discussion of the methodology is required.[5]

The Sample design discussion in an earlier chapter described the sample designs used in this study. Therefore this section will describe the methods used to extrapolate the results to the target population. Three topics will be described:

- Case weights
- Balanced stratification to calculate case weights
- Stratified ratio estimation using case weights.

Case Weights

We will use the following example problem to develop the idea of case weights[6]. Given observations of a variable y in a stratified sample, estimate the population total Y.

Note that the population total of y is the sum across the H strata of the subtotals of y in each stratum. Moreover each subtotal can be written as the number of cases in the stratum times the mean of y in the stratum. This gives the equation:

$$Y = \sum_{h=1}^{H} N_h \mu_h$$

Motivated by the preceding equation, we estimate the population mean in each stratum using the corresponding sample mean. This gives the conventional form of the stratified-sampling estimator, denoted $\hat{Y}$, of the population total Y:

$$\hat{Y} = \sum_{h=1}^{H} N_h \bar{y}_h$$

With a little algebra, the right-hand side of this equation can be rewritten in a different form:

[5] *Methods and Tools of Load Research, The MBSS System, Version V*. Roger L. Wright, RLW Analytics, Inc. Sonoma CA, 1996.

[6] This example is provided only to demonstrate the statistical concepts used in the study. The numbers presented have no relevance to the 1996 NRNC study findings.

$$\begin{aligned}\hat{Y} &= \sum_{h=1}^{H} N_h \bar{y}_h \\ &= \sum_{h=1}^{H} N_h \left(\frac{1}{n_h} \sum_{k \in s_h} y_k \right) \\ &= \sum_{k=1}^{n} \left(\frac{N_h}{n_h} \right) y_k\end{aligned}$$

Motivated by the last expression, we define the ***case weight*** of each unit in the sample to be $w_k = \dfrac{N_h}{n_h}$. Then the conventional estimate of the population total can be written as a simple weighted sum of the sample observations:

$$\hat{Y} = \sum_{k=1}^{n} w_k y_k$$

The case weight w_k can be thought of as the number of units in the population represented by unit *k* in the sample. The conventional sample estimate of the population total can be obtained by calculating the weighted sum of the values observed in the sample.

Table 21 shows an example. In this example, the population of program participants has been stratified into five strata based on the annual savings of each project shown in the tracking system. For example, the first stratum consists of all projects with annual savings less than 101,978 kWh. The maximum kWh in each stratum is called the stratum cut point. There are 339 projects in this stratum and they have a total tracking savings of 8,038,527 kWh. The estimate of gross impact was obtained from the measured savings found in a sample of 85 projects. Column 5 of Table 21 shows that the sample contains 62 projects from the first stratum. Each of these 62 projects can be given a case weight of 339 / 62 = 5.47.

	Max	**Population**	**Total**	**Sample**	**Case**
Stratum	**kWh**	**Size**	**KWh**	**Size**	**Weight**
1	101,978	339	8,038,527	62	5.47
2	278,668	61	10,949,421	9	6.78
3	441,916	35	12,598,315	8	4.38
4	816,615	22	13,654,171	3	7.33
5	4,000,000	12	17,469,244	3	4.00
Total		469	62,709,678	85	

Table 21: Stratification Example

Balanced Stratification

Balanced stratification is another way to calculate case weights. In this approach, the sample sites are sorted by the stratification variable, tracking kWh, and then divided equally among the strata. Then the first stratum cutpoint is determined midway between the values of the stratification variable for the last sample case in the first stratum and the first sample case in the second stratum. The remaining strata cutpoints are determined in a similar fashion. Then the population sizes are tabulated within each stratum. Finally the case weights are calculated in the usual way.

Table 22 shows an example. In this case the sample of 85 sites has been equally divided among five strata, so there are 17 sites per stratum. Then the stratum cutpoints shown in column two were calculated from the tracking estimates of kWh for the sample sites. Next the population sizes shown in column three were calculated from the stratum cutpoints. The final step was to calculate the case weights shown in the last column. For example, the case weight for the 17 sites in the first stratum is 136 / 17 = 8.

Stratum	Max kWh	Population Size	Total KWh	Sample Size	Case Weight
1	7,948	136	417,368	17	8.00
2	22,361	84	1,211,832	17	4.94
3	63,859	84	3,605,867	17	4.94
4	202,862	73	8,146,886	17	4.29
5	2,883,355	92	49,327,725	17	5.41
Total		469	62,709,678	85	

Table 22: Balanced Stratification

Stratified Ratio Estimation

Ratio estimation is used to estimate the population total *Y* of the target variable *y* taking advantage of the known population total *X* of a suitable explanatory variable *x*. The ratio estimate of the population total is denoted $\hat{Y}_{ra}$ to distinguish it from the ordinary stratified sampling estimate of the population total, which is denoted as $\hat{Y}$.

Motivated by the identity $Y = B\,X$, we estimate the population total *Y* by first estimating the population ratio *B* using the sample ratio $b = \bar{y}/\bar{x}$, and then estimating the population total as the product of the sample ratio and the known population total *X*. Here the sample means are calculated using the appropriate case weights. This procedure can be summarized as follows:

$$
\begin{aligned}
\hat{Y}_{ra} &= b\,X \quad \text{where} \\
b &= \frac{\bar{y}}{\bar{x}} \\
\bar{y} &= \frac{1}{\hat{N}} \sum_{k=1}^{n} w_k\, y_k \\
\bar{x} &= \frac{1}{\hat{N}} \sum_{k=1}^{n} w_k\, x_k \\
\hat{N} &= \sum_{k=1}^{n} w_k
\end{aligned}
$$

The conventional 90 percent confidence interval for the ratio estimate of the population total is usually written as

$$
\begin{aligned}
\hat{Y}_{ra} &\pm 1.645\sqrt{V\left(\hat{Y}_{ra}\right)} \quad \text{where} \\
V\left(\hat{Y}_{ra}\right) &= \sum_{h=1}^{H} N_h^2 \left(1-\frac{n_h}{N_h}\right)\frac{s_h^2(e)}{n_h} \\
s_h^2(e) &= \frac{1}{n_h-1}\sum_{k\in s_h}\left(e_k-\bar{e}_h\right)^2 \\
e_k &= y_k - b\,x_k
\end{aligned}
$$

We can calculate the relative precision of the estimate $\hat{Y}_{ra}$ using the equation

$$rp = \frac{1.645\sqrt{V\left(\hat{Y}_{ra}\right)}}{\hat{Y}_{ra}}$$

MBSS theory has led to an alternative procedure to calculate confidence intervals for ratio estimation, called model-based domains estimation. This method yields the same estimate as the conventional approach described above, but gives slightly different error bounds. This approach has many advantages, especially for small samples, and has been used throughout this study.

Under model-based domains estimation, the ratio estimator of the population total is calculated as usual. However, the variance of the ratio estimator is estimated from the case weights using the equation

$$V\left(\hat{Y}_{ra}\right) = \sum_{k=1}^{n} w_k\left(w_k-1\right)e_k^2$$

Here w_k is the case weight discussed above and e_k is the sample residual $e_k = y_k - b\,x_k$. Then, as usual, the confidence interval is calculated as

$$\hat{Y}_{ra} \pm 1.645\sqrt{V\left(\hat{Y}_{ra}\right)}$$

and the achieved relative precision is calculated as

$$rp = \frac{1.645\sqrt{V\left(\hat{Y}_{ra}\right)}}{\hat{Y}_{ra}}$$

The model-based domains estimation approach is often much easier to calculate than the conventional approach since it is not necessary to group the sample into strata. In large samples, there is generally not much difference between the case-weight approach and the conventional approach. In small samples the case-weight approach seems to perform better. For consistency, we have come to use model-based domains estimation in most work.

This methodology generally gives error bounds similar to the conventional approach. Equally, the model-based domains estimation approach can be derived from the conventional approach by making the substitutions:

$$\begin{aligned} \bar{e}_h &\approx 0 \\ s_h^2(e) &\approx \frac{1}{n_h}\sum_{k \in s_h} e_k^2 \end{aligned}$$

In the first of these substitutions, we are assuming that the within-stratum mean of the residuals is close to zero in each stratum. In the second substitution, we have replaced the within-stratum variance of the sample residual *e*, calculated with $n_h - 1$ degrees of freedom, with the mean of the squared residuals, calculated with n_h degrees of freedom.

Model-based domains estimation is appropriate as long as the expected value of the residuals can be assumed to be close to zero. This assumption is checked by examining the scatter plot of *y* versus *x*. It is important to note that the assumption affects only the error bound, not the estimate itself. $\hat{Y}_{ra}$ will be essentially unbiased as long as the case weights are accurate.

Gross Savings Expansions

Baseline, as-built, and savings estimates were developed for each building in the sample. The sample of baseline, as built, and savings estimates was projected to the participant population using model-based statistical methods described above.

Energy Impact Findings

Whole Building

PG&E's whole building gross energy savings was 83,970 MWh. The relative precision of the estimate was ±6.0%. This represents a gross realization rate of 104.4% of verified annual savings. Table 23 shows the estimated energy savings by costing period.

	Energy Savings (MWh)	Error Bound (MWh)		Relative Precision
Annual	83,970	±	5,010	± 6.0%
Summer On-Peak	7,697	±	482	± 6.3%
Summer Part-Peak	9,503	±	545	± 5.7%
Summer Off-Peak	13,840	±	924	± 6.7%
Winter Part-Peak	26,530	±	1,736	± 6.5%
Winter Off-Peak	26,410	±	1,813	± 6.9%

Table 23: Whole Building Energy Savings by Costing period

The participant group was more energy efficient than the non-participant comparison group. Figure 17 shows the savings of participants and non-participants expressed as a percentage of each group's whole-building baseline usage.

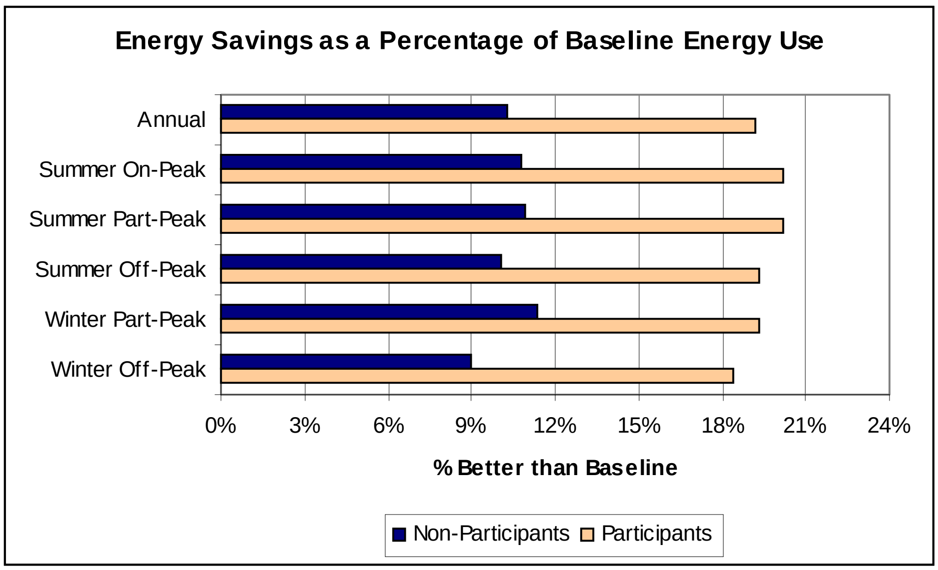

Figure 18: Participant and Non-participant Energy Savings as a Percentage of Baseline

As Figure 17 shows, the participant group was 19% better than baseline on average. The non-participant comparison group was 10% better than baseline. The level of efficiency relative to the baseline remains fairly constant throughout the year.

End-Use Savings

Five end-uses were examined as part of this study:

- Lighting – Lamps, ballasts, controls
- HVAC – Compressor efficiency, VSDs, oversized cooling towers
- Refrigeration – Commercial refrigeration systems (condensers, compressors, cases)

- Motors – All energy efficient motors, including HVAC fans. Also overall air distribution system design measures such as efficient cooling coils and oversized ducts.
- Shell – High performance glass

Those sites that had savings in a particular end-use were projected to the population to arrive at the total savings estimate. Note that the sum of the end-use savings may not add exactly to 1 due to rounding. In each of the figures describing end-use savings, the percentages are of the whole building baseline. The percentage scale in the figures is an indicator of the contribution to overall savings of each end-use.

Figure 18 shows the breakdown of annual energy savings by end-use. The savings associated with the shell end-use was not statistically significant and is omitted in Figure 18. The shell end-use will not be discussed further in this section.

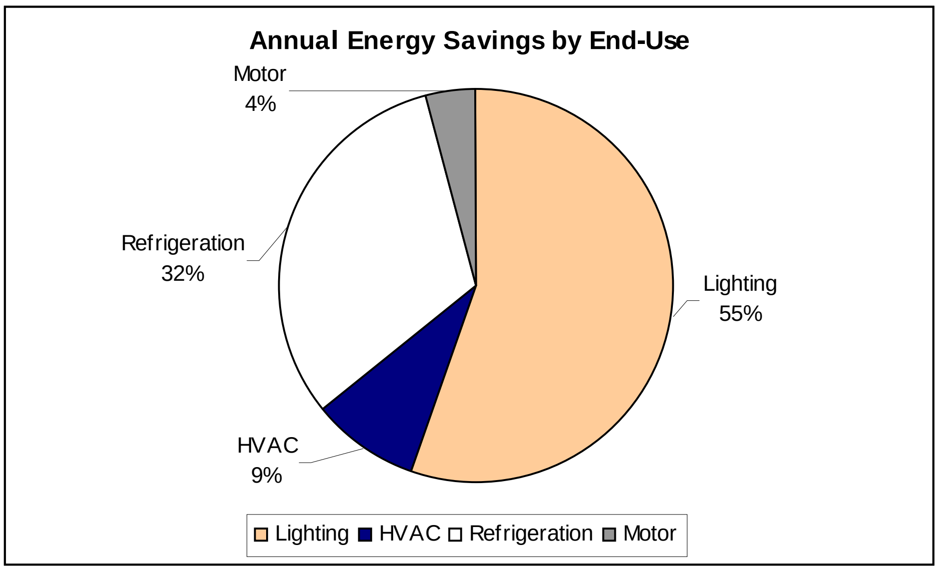

Figure 19: Composition of Annual Energy Savings as a Percentage of Whole Building Savings

Lighting

The lighting end-use accounted for 46,400 MWh of annual energy savings. This was 55% of the total annual energy savings. Table 24 shows the savings and relative precision by costing period.

	Energy Savings (MWh)	**Error Bound (MWh)**	**Relative Precision**
Annual	46,400	± 5,101	± 11.0%
Summer On-Peak	4,006	± 459	± 11.5%
Summer Part-Peak	4,917	± 530	± 10.8%
Summer Off-Peak	6,726	± 764	± 11.4%
Winter Part-Peak	16,290	± 1,868	± 11.5%
Winter Off-Peak	14,460	± 1,620	± 11.2%

Table 24: Lighting Energy Savings by Costing period

Figure 19 shows the participant and non-participant lighting savings relative to baseline consumption by costing period. The lighting energy efficiency of participants was 83% greater than the non-participants.

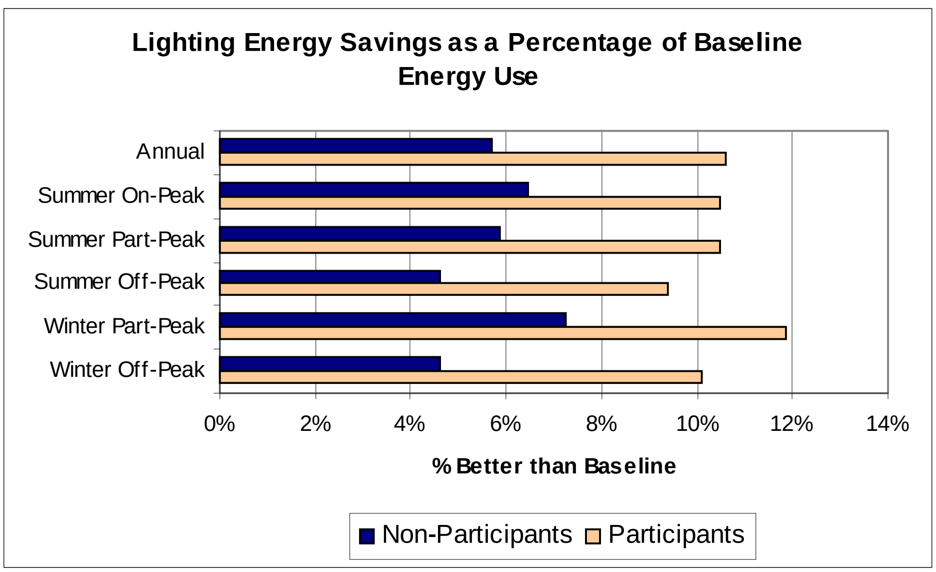

Figure 20: Lighting Energy Savings as a Percentage of Baseline Energy Use

HVAC

The HVAC end-use accounted for 7,547 MWh of energy savings, or 9% of annual energy savings. Table 25 shows the savings and relative precision by costing period.

	Energy Savings (MWh)	Error Bound (MWh)	Relative Precision
Annual	7,547	± 1,237	± 16.4%
Summer On-Peak	1,027	± 175	± 17.0%
Summer Part-Peak	1,072	± 179	± 16.7%
Summer Off-Peak	1,313	± 285	± 21.7%
Winter Part-Peak	2,399	± 354	± 14.8%
Winter Off-Peak	1,736	± 377	± 21.7%

Table 25: HVAC Energy Savings by Costing period

Figure 20 shows the participant and non-participant HVAC savings relative to baseline consumption by costing period. The HVAC end-use savings for participants was 243% of the savings for non-participants. Note that in the winter off-peak period, non-participant HVAC energy savings was not statistically different from the baseline.

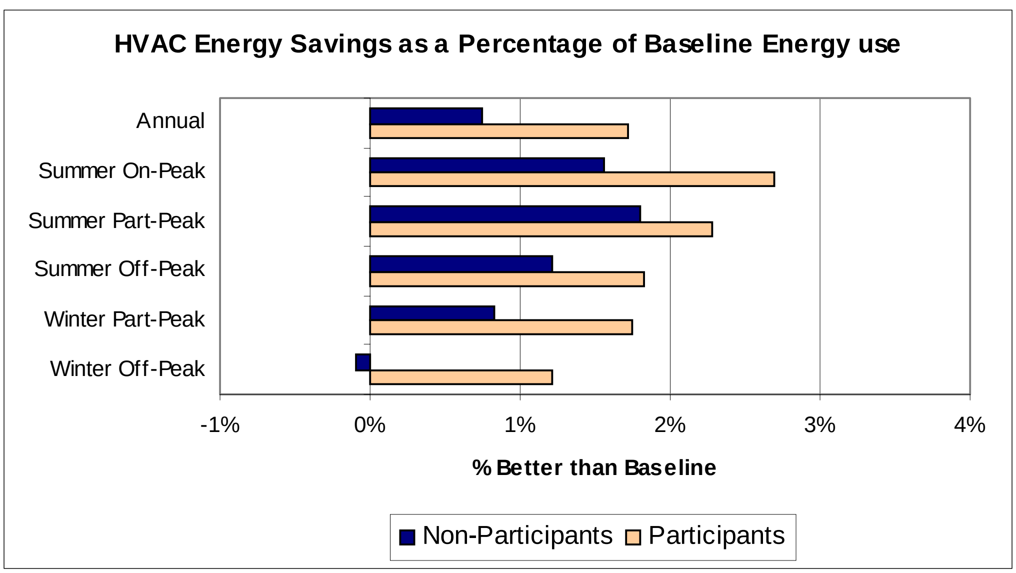

Figure 21: HVAC Energy Savings as a Percentage of Baseline

Refrigeration

The refrigeration end-use accounted for 26,490 MWh, or 32%, of the participant group energy savings. Table 26 shows the savings and relative precision by costing period.

	Energy Savings (MWh)	**Error Bound (MWh)**	**Relative Precision**
Annual	26,490	± 3,485	± 13.2%
Summer On-Peak	2,276	± 364	± 16.0%
Summer Part-Peak	3,156	± 493	± 15.6%
Summer Off-Peak	5,191	± 751	± 14.5%
Winter Part-Peak	6,858	± 960	± 14.0%
Winter Off-Peak	9,009	± 1,207	± 13.4%

Table 26: Refrigeration Energy Savings by Costing period

Figure 21 shows the participant and non-participant refrigeration savings relative to baseline consumption by costing period. The participants' refrigeration savings was 300% of the non-participants' refrigeration savings.

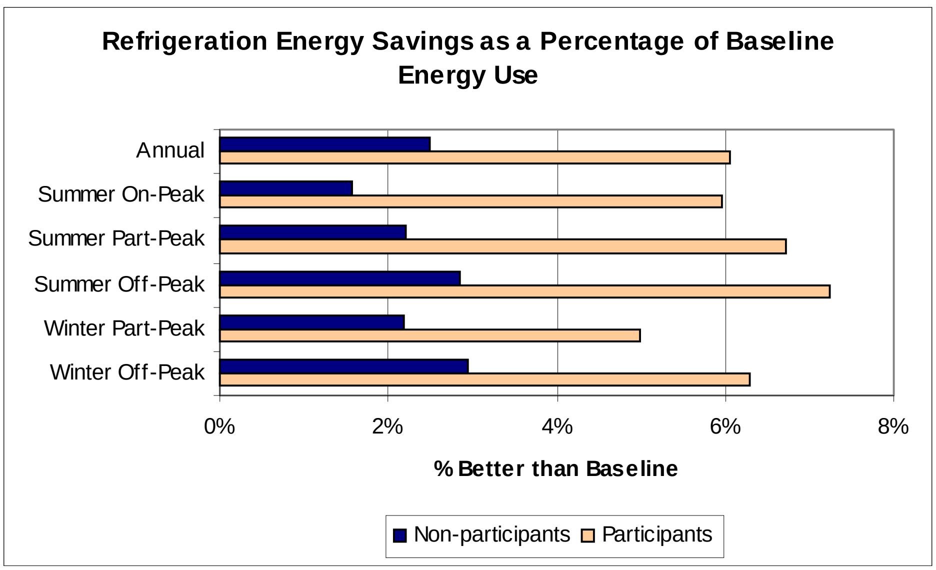

Figure 22: Refrigeration Energy Savings as a Percentage of Baseline Energy

Motors

The motor end-use made the smallest contribution to savings at 3,539 MWh. This was 4% of total savings. Table 27 shows the motor energy savings by costing period.

	Energy Savings (MWh)	Error Bound (MWh)		Relative Precision
Annual	3,539	±	1,005	± 28.4%
Summer On-Peak	292	±	81	± 27.8%
Summer Part-Peak	342	±	93	± 27.2%
Summer Off-Peak	557	±	165	± 29.6%
Winter Part-Peak	1,132	±	323	± 28.6%
Winter Off-Peak	1,216	±	371	± 30.5%

Table 27: Motor Energy Savings by Costing Period

Figure 22 shows the participant and non-participant savings relative to the baseline. Although the figure shows that the non-participants had greater savings than the participants, the difference between the two groups is not statistically significant for any costing period.

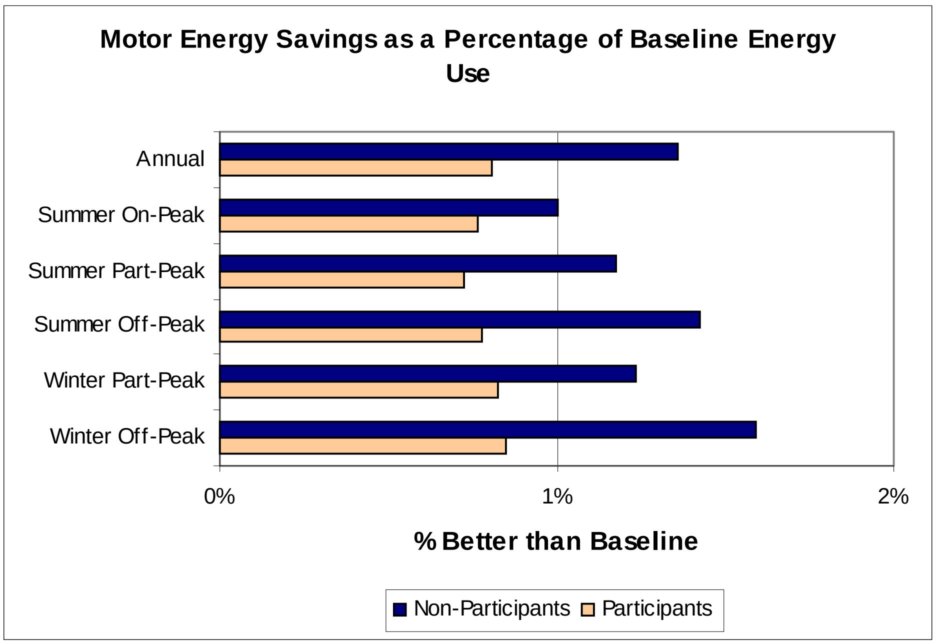

Figure 23: Motor Energy Savings as a Percentage of Baseline Energy

Demand Impact Findings

Whole Building

PG&E's whole building gross demand savings was 20.0 MW. The relative precision of the estimate was ±7.1%. This represents a gross realization rate of 104.6% of verified summer on-peak demand savings. Table 28 shows the estimated savings by costing period.

	Peak Demand Savings (MW)	Error Bound (MW)	Relative Precision
Summer On-Peak	20.0	± 1.4	± 7.1%
Summer Part-Peak	19.3	± 1.4	± 7.4%
Summer Off-Peak	19.2	± 1.3	± 6.9%
Winter Part-Peak	19.7	± 1.4	± 7.1%
Winter Off-Peak	18.5	± 1.4	± 7.6%

Table 28: Whole Building Demand Savings by Costing period

The participant group was more energy efficient than the non-participant comparison group. Figure 23 shows the savings of participants and non-participants expressed as a percentage of each group's baseline demand.

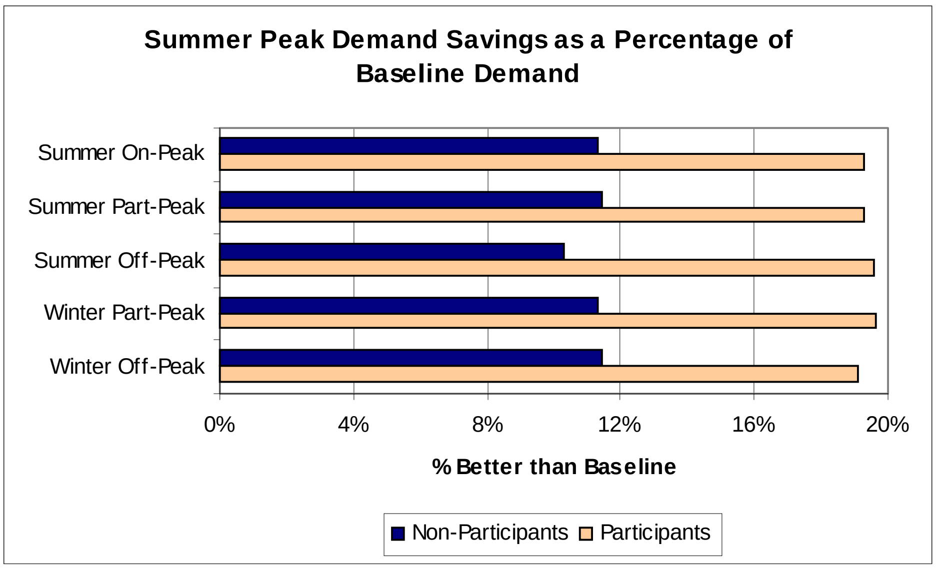

Figure 24: Participant and Non-participant Savings as a Percentage of Baseline Demand

As Figure 23 shows, the summer on-peak demand of the participant group was 19.3% better than baseline. The non-participant comparison group was 11.3% better than baseline. The participnat group summer on-peak demand savings was 71% greater than the non-participant group savings. The level of efficiency relative to the baseline remains fairly constant throughout the year.

End-Use Demand Savings

Figure 24 shows the breakdown of annual energy savings by end-use. The HVAC end-use has a larger impact on summer peak demand savings than it does on annual energy because of its seasonal nature.

The Shell end-use did not produce statistically significant savings and is not discussed in this section.

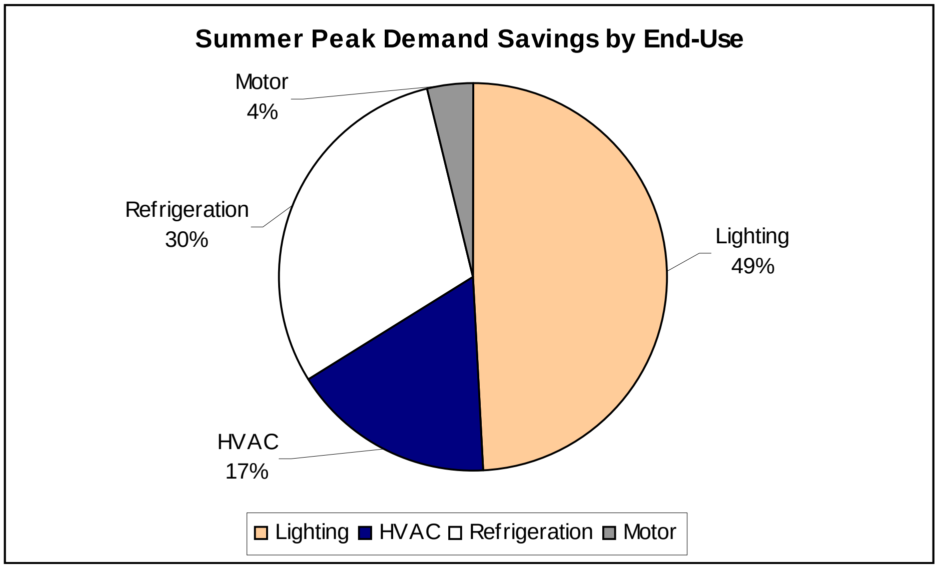

Figure 25: Summer Peak Demand Savings by End-Use

Lighting

PG&E's lighting end-use summer on-peak gross demand savings was 9.8 MW. The relative precision of the estimate was ±12.3%. Table 29 shows the estimated savings by costing period.

	Peak Demand Savings (MW)	Error Bound (MW)	Relative Precision
Summer On-Peak	9.8	± 1.2	± 12.3%
Summer Part-Peak	9.7	± 1.2	± 12.2%
Summer Off-Peak	9.4	± 1.1	± 11.3%
Winter Part-Peak	9.7	± 1.2	± 12.1%
Winter Off-Peak	9.5	± 1.3	± 13.2%

Table 29: Lighting Demand Savings by Costing Period

The participant group was more energy efficient than the non-participant comparison group. Figure 25 shows the savings of participants and non-participants expressed as a percentage of each group's baseline demand. The lighting end-use participants saved 48% more than the non-participants relative to baseline.

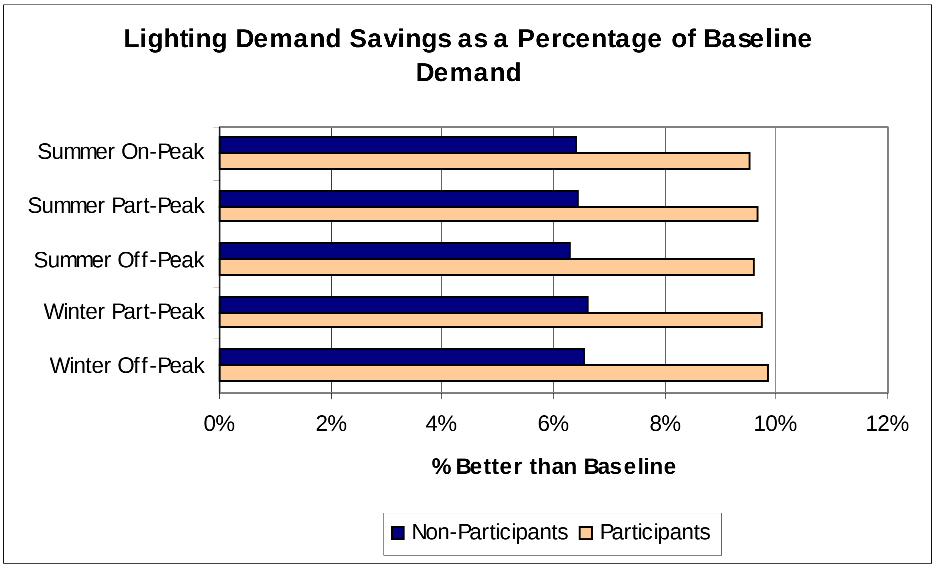

Figure 26: Lighting Demand Savings as a Percentage of Baseline Demand

HVAC

PG&E's HVAC end-use gross demand savings was 3.4 MW. The relative precision of the estimate was ±16.5%. Table 30 shows the estimated savings by costing period.

	Peak Demand Savings (MW)	Error Bound (MW)		Relative Precision
Summer On-Peak	3.4	±	0.6	± 16.5%
Summer Part-Peak	3.2	±	0.5	± 16.0%
Summer Off-Peak	3.3	±	0.5	± 15.8%
Winter Part-Peak	3.2	±	0.5	± 15.9%
Winter Off-Peak	2.8	±	0.4	± 15.9%

Table 30: HVAC Demand Savings by Costing Period

The participant group was more energy efficient than the non-participant comparison group. Figure 26 shows the savings of participants and non-participants expressed as a percentage of each group's baseline demand. The participant group savings was 37% larger than the non-participant group savings for this end use.

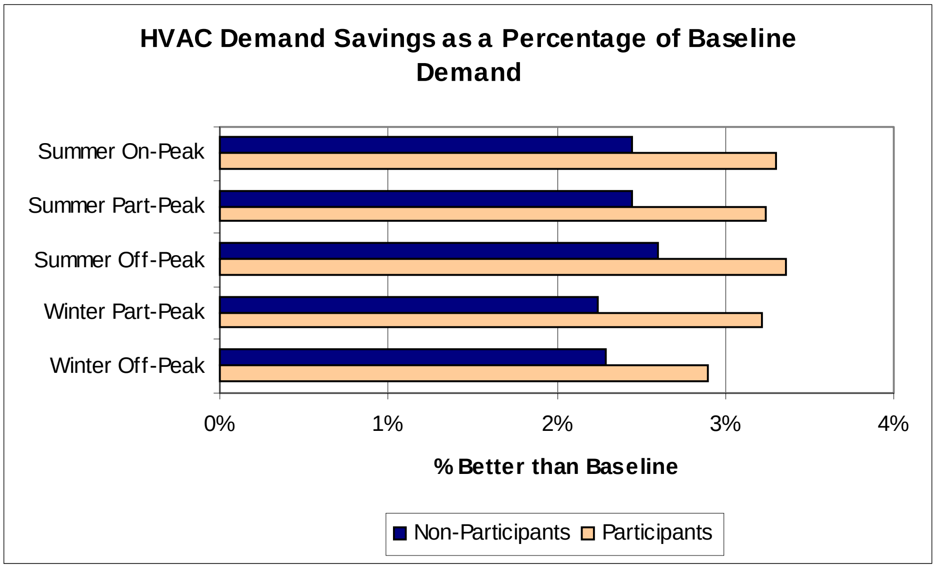

Figure 27: HVAC Demand Savings as a Percentage of Baseline Demand

Refrigeration

PG&E's refrigeration end-use gross demand savings was 6.0 MW. The relative precision of the estimate was ±19.0%. Table 31 shows the estimated savings by costing period.

	Peak Demand Savings (MW)	Error Bound (MW)	Relative Precision
Summer On-Peak	6.0	± 1.1	± 19.0%
Summer Part-Peak	5.7	± 1.1	± 19.1%
Summer Off-Peak	5.8	± 1.1	± 19.3%
Winter Part-Peak	5.6	± 1.1	± 18.7%
Winter Off-Peak	5.5	± 1.0	± 19.1%

Table 31: Refrigeration Demand Savings by Costing Period

The participant group was more energy efficient than the non-participant comparison group. Figure 27 shows the savings of participants and non-participants expressed as a percentage of each group's baseline demand. The participant group saved more than 6 times what the non-participant group saved during the summer on-peak period for this end-use.

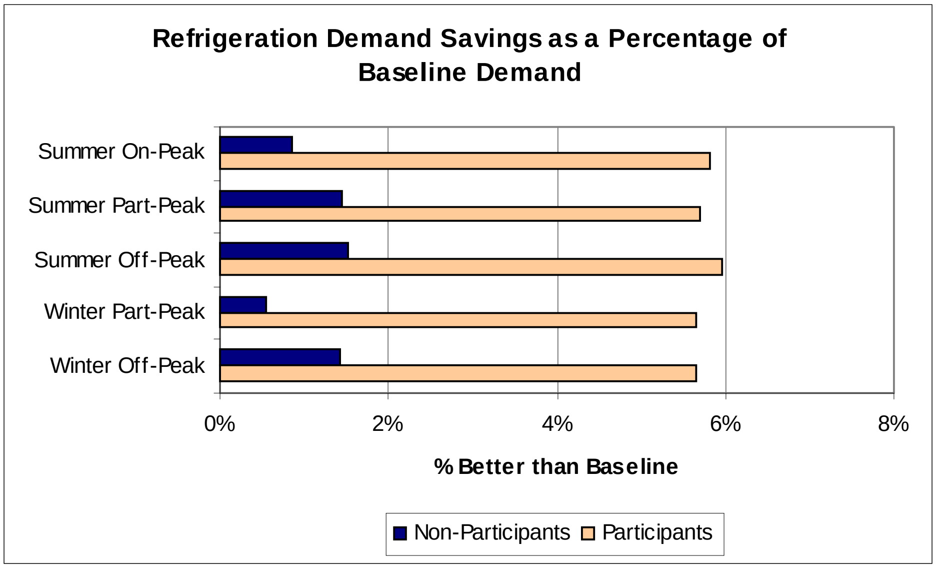

Figure 28: Refrigeration Demand Savings as a Percentage of Baseline Demand

Motors

The motor end use had a summer on-peak demand savings of 0.8 MW, or 4% of the total demand savings. Table 32 shows the savings by costing period.

	Peak Demand Savings (MW)	Error Bound (MW)	Relative Precsion
Summer On-Peak	0.8	± 0.2	± 23.5%
Summer Part-Peak	0.8	± 0.2	± 24.4%
Summer Off-Peak	0.9	± 0.2	± 24.6%
Winter Part-Peak	0.7	± 0.2	± 23.6%
Winter Off-Peak	0.8	± 0.2	± 25.1%

Table 32: Motor Demand Savings by Costing Period

Figure 28 shows the savings for participants and non-participants as a percentage of baseline demand. As with the energy results, the participant / non-participant differences were not statistically significant.

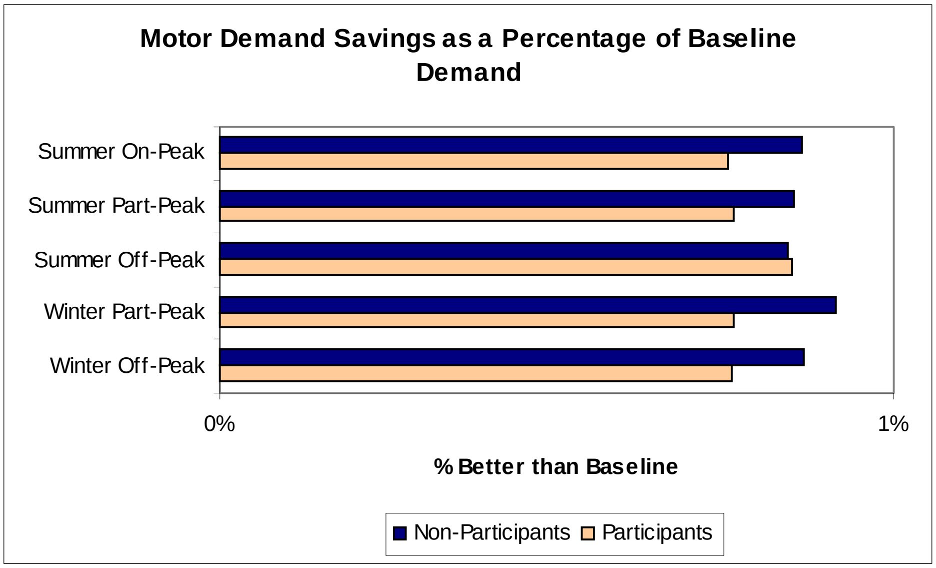

Figure 29: Motor Demand Savings as a Percentage of Baseline Demand

Net Savings

Two different methodologies were followed in the net-to-gross analysis: a relatively simple difference of differences approach and a more complex econometric approach. In the difference-of-differences methodology, the net-to-gross ratio was calculated by comparing the gross savings relative to baseline of the program participants to the gross savings relative to baseline of the non-participants. In the econometric approach, the net-to-gross ratio was calculated by using regression modeling techniques to estimate the net savings due to the program for each of the program participants.

Difference of Differences Net-to-Gross

The following table summarizes the findings from the difference-of-differences analysis. The table shows the estimated net-to-gross ratio for both annual energy and summer peak demand savings.

	Net Savings	**Net-to-gross Ratio**	**Relative Precision**
Annual Energy	39,054 MWh	46.5%	±25.5%
Summer Peak Demand	8.2 MW	41.0%	±25.1%

Table 33: Difference of Differences Savings and Net-to-gross Ratios

The table also shows the relative precision of each estimate.[7] For example, in the case of annual energy, the net-to-gross ratio was estimated to be 46.5% with a relative precision of ±25.5%. The error bound for the 90% confidence interval for the true net-to-gross ratio is equal to 25.5% of the estimate, i.e. to ±11.9%. The 90% confidence interval for the true net-to-gross ratio can be calculated using the equation:

$$0.465 \pm (0.465 * 0.255) = (0.346, 0.584)$$

We can be quite confident that this interval contains the true net-to-gross ratio that would have been obtained by developing on-site surveys and building engineering simulation models for all program participants and a very large sample of non-participants using the methodology of this study and then analyzing the resulting data using the difference of differences methodology. The confidence interval reflects sampling variability and random measurement error but does not reflect any possible systematic measurement error that might be repeated throughout the data collection and engineering simulation or that might arise by neglecting explicit estimation of free ridership and spillover.

Econometric Net-to-Gross

The following table summarizes the findings from the econometric analysis. The table shows the estimated net savings and net-to-gross ratio for both annual

[7] The standard error reflects the standard deviation of the estimate in repeated sampling. The error bound at the 90% level of confidence is 1.645 times the standard error. The confidence interval is the estimate plus or minus the error bound. The relative precision is the error bound divided by the estimate itself.

energy and summer peak demand savings. These are the total net savings excluding non-participant spillover savings.

	Net Savings	Net-to-gross Ratio	Relative Precision
Annual Energy	58,569 MWh	69.8%	±10.5%
Summer Peak Demand	13.9 MW	69.8%	±12.1%

Table 34: Econometric Estimates of Saving and Net-to-gross Ratios

The table also shows the relative precision of each estimate. For example, in the case of annual energy, the net-to-gross ratio was estimated to be 69.8% with a relative precision of ±10.5%. The error bound for the 90% confidence interval for the true net-to-gross ratio is equal to 10.5% of the estimate i.e. to ±7.3%. The 90% confidence interval for the true net-to-gross ratio is

$$0.698 \pm (0.698 * 0.105) = (0.625, 0.771)$$

The confidence interval for annual savings can be calculated in a similar way.

There is a 90% probability that these confidence intervals include the true values that would have been obtained by developing onsite surveys and building engineering simulation models for all program participants and a very large sample of non-participants using the methodology of this study and then analyzing the resulting data using the econometric methodology. These estimates reflect sampling variability, random measurement error, and explicit estimation of free-ridership and spillover. But these estimates do not reflect any possible systematic measurement error that might be repeated throughout the data collection and engineering simulation or any possible bias arising from inaccuracy in the assumed econometric model.

Difference of Differences Methodology

This section describes the difference-of-differences methodology. For simplicity we will discuss the methodology used to analyze annual energy savings. An analogous approach was used to analyze summer peak demand savings.

Table 34 summarizes the derivation of the net-to-gross ratio for annual energy. The analysis starts with the baseline and as-built energy consumption of the participants and non-participants. All of these results are reported in MWh and were obtained by statistically expanding the sample data to the population of 1996 program participants. For example, the table shows that we would estimate that all program participants would have an aggregate annual consumption of 353,830 MWh, based on the as-built simulation runs developed for the sites in the participant sample. By contrast, if we expand the as-built simulation runs of the non-participants to the same participant population, we would expect an aggregate annual consumption of 550,200 MWh.

Considering only the as-built results, the participants would appear to be more energy efficient than the non-participants. However, this fails to control for differences between the two samples. The preceding table shows that the baseline results were 437,800 MWh using the participant sample and 613,100

	Participants	Non-Participants	Net Savings
Baseline (MWh)	437,800.	613,100.	
As-Built (MWh)	353,830.	550,200.	
Savings (MWh)	83,970.	62,900.	39,054
Savings (% of baseline)	19.18%	10.26%	8.92%
Net-to-Gross Ratio			46.5%

Table 35: Summary of Difference of Differences Calculation

MWh using the non-participant sample. Both samples were designed to be representative of the population of 1996 program participants. However we would expect differences in the baseline results from the two samples due to sampling variability. Moreover, difficulty in obtaining large non-participant sample sites to match the large participants in the program may have led to some systematic difference between the participant and non-participant samples. In fact, the observed difference corresponds to a t-statistic of about 2.6,[8] implying that the difference is significant at the 5% level of significance. This suggests that there were some nonrandom differences in the participant and non-participant samples.

For a more meaningful comparison, the as-built energy use should be considered relative to the baseline. The table shows the gross savings, calculated as the difference between the baseline and the as-built energy use. Calculated this way, the gross savings relative to baseline were 83,970 MWh for the participant sample and 62,900 MWh for the non-participant sample. In proportion to the respective baseline energy use of each sample, the gross savings were 19.18% for the participant sample and 10.26% for the non-participant sample.

In the difference-of-differences approach, the net savings can be estimated as the difference between the percentage savings of the participants and non-participants. In this case the net savings is 8.92% of baseline use. Multiplying 437,800 MWh by 8.92%, the net savings of the population of 1996 program participants can be estimated to be 39,054 MWh.

The net savings of the program participants can also be calculated using the following equation.

$$\left(\frac{550{,}200}{613{,}100}\right) \cdot 437{,}800 - 353{,}830 = 39{,}054$$

Here the first factor is the as-built energy use relative to the baseline energy use using the non-participants. This is used to adjust the baseline energy use of the participants. Then the net savings is calculated by subtracting the as-built energy use of the participants. Finally, the net savings is found to be 8.92% of the baseline energy use of the participants. The two approaches for calculating net savings are mathematically equivalent.

[8] The standard errors were about 27,200 MWh for the baseline energy use from the participant sample and about 62,000 MWh for the baseline energy use from the non-participant sample.

Finally the net-to-gross ratio can be calculated by dividing the net savings (8.92%) by the participants' gross savings (19.2%). This gives the difference of differences estimate of 46.5% for the net-to-gross ratio for annual energy.

Error Bound Methodology for the Difference of Differences Estimate

In the preceding section, it was shown that the difference of differences estimate of net savings can be expressed as an adjustment of the participant sample results based on the ratio between the as-built and baseline results observed from the non-participant sample. This is an extension of the technique of ratio estimation that is common in survey sampling.[9] The error bound and relative precision of the difference of differences estimate of net savings can be estimated using techniques similar to the methods of standard ratio estimation. In this section, we will describe the approach.

First some notation. Let $r_{np} = \left(\frac{550{,}200}{613{,}100}\right)$ denote the ratio between the as-built and baseline energy use obtained in the non-participant sample. Let $r_b = \left(\frac{437{,}800}{613{,}100}\right)$ denote the ratio between the baseline energy use of the participants relative to the non-participants. For any sample site, participant or non-participant, let *e* denote the difference between the as-built energy use of the site and the product of r_{np} and the baseline energy use of the site.

Now the error bound of the difference-of-differences estimate of net savings can be estimated in three steps:

1. Calculate *e* for each site in the participant sample and use standard techniques to expand the results to the target population. Let *A* denote the error bound of the result, calculated the usual way.
2. Calculate *e* for each site in the non-participant sample and use standard techniques to expand the results to the target population. Let *B* denote the error bound of the result, calculated the usual way.
3. Estimate the error bound of the difference of differences estimate of net savings using the equation $\sqrt{A^2 + r_b^2 B^2}$.

The preceding methodology can be derived from a standard Taylor's series approximation to the sampling distribution of the difference of differences estimator.

Rationale for the Econometric Net-to-Gross Methodology

The econometric methodology can be regarded as an extension of a simple comparison of the efficiency choice of non-participant and participants through the difference of difference methodology. A coefficient of the participation indicator variable reflects the difference in efficiency choice between a

[9] See, for example, Chapter 6 of *Sampling Techniques,* by W. A. Cochran, Wiley and Sons, third edition, 1977.

participant and a non-participant. Other variables are included in the model to control for other factors that are associated with efficiency choice.

The inclusion of these variables can improve the statistical model in two ways:

1. Reduce potential bias, and
2. Provide improved statistical precision.

The potential bias arises if the model omits an explanatory variable that (a) is related to efficiency choice, and (b) is correlated with participation. For example, suppose a particular type of builder or designer tends to build a more efficient building and also tends to participate in the program. Then the difference of difference approach would tend to overestimate the actual impact of the program. This is sometimes called self-selection bias.

As another example, suppose that some of the non-participants have incorporated efficiency measures into the current building that they learned from participating in the program in prior years. In this case the difference of difference approach would underestimate the actual impact of the program. This can be called bias due to spillover.

Therefore, under most circumstances the difference of difference approach provides a biased estimate of the actual program impact. The size of the bias depends on the balance between any positive bias due to self-selection and related factors versus any negative bias due to spillover and similar factors. The only circumstances under which the difference of difference approach would give an unbiased estimate are either (a) if both self-selection and spillover are negligible, or (b) if they are exactly equal. Both of these assumptions seem unlikely, especially for a program deliberately designed to influence general practice in new construction, so a more powerful methodology is needed to obtain an unbiased estimate of net savings.

The econometric methodology seeks to obtain an unbiased estimate of net savings by including both program variables and other explanatory variables in a multivariate regression model. If the model is accurately specified and if the program variables and other explanatory variables are not multicollinear, then the model will provide an unbiased estimate of the net program savings among the participants as well as the spillover impact among the non-participants. This is the primary motivation for a multivariate regression analysis.

The econometric approach can also improve statistical precision by including explanatory variables that significantly affect efficiency choice. If an explanatory variable has a significant relationship with efficiency, then its inclusion in the model may significantly decrease the residual variance, or unexplained variance, of the model, and in turn, provide more statistically reliable estimates of net savings and spillover impacts.

Conversely, there are reasons for excluding all variables that do not have a significant relationship with efficiency. The inclusion of such variables needlessly tends to reduce the statistical precision of the results and makes the models unnecessarily complex and difficult to interpret. Therefore, we seek to include all truly relevant variables but drop the irrelevant variables. Necessarily, this is an iterative process, but a well-defined and objective procedure can be followed to obtain the final model and resulting estimates of net savings and spillover impacts.

Explanatory Variables

The following table summarizes the data elements used to develop the potential explanatory variables for the econometric analysis. The table shows the source of each data element and gives a brief description of the relevance of each data element to the econometric analysis.

Data Element	*Collection*	*Rationale*
Weather Zone	On-site	Account for geographic differences. Both construction practice and PG&E local office program implementation
Building Type	On-site	Different types of buildings may be built to different efficiency standards. This was seen in the 1994 study.
Project Type	Phone	New construction may be built more efficiently than additions or renovations
Building ownership	Phone	Owner occupants may be more concerned with efficiency than developers / landlords.
Construction circumstances	Phone	Same as above
Owner input	Phone	More owner input makes owner attitudes more important with respect to efficiency choices.
Pre-existing plans	Phone	Standard designs reduce the likelihood of efficiency measures in response to the program.
Investment Criteria	Phone	Investment criteria may affect willingness to install efficiency measures
Signif. Of energy costs	Phone	Significance of energy costs may influence efficiency choice.
Signif. Of energy eff	Phone	Significance of energy efficiency may influence decision to install higher eff. equipment
Awareness of program	Phone	Awareness may lead to spillover.
Interaction with utility on this project	Phone	Interaction with PG&Emay lead to spillover
Influence of utility on this project	Phone	Influence of PG&E may lead to spillover.
Interaction with utility on previous projects	Phone	Previous interaction with PG&E may lead to spillover
Influence of utility on previous projects	Phone	Previous influence of PG&E may lead to spillover

Table 36: Variables Considered for Econometric Analysis

General Methodology for Data Screening and Analysis

A systematic process was followed to specify the final logistic and efficiency choice models. The present section summarizes how each of the following issues were addressed. Additional details will be found in other sections of the report, especially the following sections of this chapter.

- Weather adjustment
- Background variables such as economic activity
- Missing data points
- Missing or unusable billing data
- Missing responses to questions
- Outliers and data screens
- Model specification
- Cross sectional variation
- Time series variation
- Participant self selection
- Omitted factors
- Estimation of net impacts
- Errors in measuring variables
- Autocorrelation
- Heteroscedasticity
- Collinearity
- Influential data points
- Statistical Precision

Weather adjustment

This was handled in the engineering modeling. The model calibration used actual weather concurrent with the available billing data. Then all models were run using typical meteorological weather data. In this way the gross savings determined by the engineering models reflected normal weather conditions expected in each climate zone.

Background variables such as economic activity

This was also handled in the engineering modeling. The schedules used in the models were based on the levels of building use observed in the on-site survey. The schedules were held fixed in calculating the gross savings. Therefore the savings can be regarded as representing the actual savings obtained under the economic activity found at the time of the on-site surveys.

Missing data points

Sites that refused to participate in the study were replaced using a randomly drawn sample of backup sites. The level of refusal was rather low, as discussed earlier in this report.

Missing or unusable billing data

Whenever possible, the engineering models were calibrated to the available billing data. However, many of the projects studied in this evaluation were actually renovations or additions to existing buildings. In many of these cases, the available billing data described the whole building rather than the actual space that was renovated or added. In these cases, when it was practical we installed special metering equipment to collect load data for use in calibration. When this was not practical, the models were used without calibration.

Missing responses to questions

When a decision-maker did not know or refused to answer a particular question, we tried to identify a more appropriate respondent. If this failed, we recorded the response as 'don't know' or 'refused'. In the case of questions with categorical answers, we treated all such answers as a distinct category of response and created a corresponding indicator variable. In the case of the questions that were answered on a seven-point scale, we coded the response as 0 and created a corresponding indicator variable.

Outliers and data screens

The full sample was retained throughout the analysis. Studentized residuals were used to identify outliers. A site was considered to be an outlier if its studentized residual was greater than three in absolute value. A separate indicator variable was used to represent each such outlier in the model. The coefficient of this indicator variable indicated how much the dependent variable deviated from its expected value for the particular outlier. The statistical significance of these indicator variables was used to identify outliers that were statistically significant.

Model specification

A systematic approach was followed so that each model would be properly specified. The primary concern was to avoid bias arising from specification error – omitted variables, outliers, omitted statistical interactions, etc. We also sought to obtain a parsimonious final model that included only statistically significant variables. The following sections trace the approach, indicate some of the tests and graphical displays that were used to examine intermediate models, and compare the models that were examined. The entire process of refining the models is documented in SPSS command files.

Cross sectional variation

Cross-sectional variation was addressed throughout the sample design and experimental approach as well as in the modeling. The sample design was based on the experience of the 1994 evaluation study and sought to represent the full diversity of participants in the program, and a matched sample of non-participants. The sample size and stratification were chosen to yield statistically reliable estimates of the overall savings of the program. The experimental approach was built around engineering auditing and modeling techniques that were designed to capture the full range of actual building equipment types and schedules found in the population. The gross analysis was designed to determine the actual gross savings of each site, controlled for the actual equipment and use of the site. The net-to-gross analysis was designed to control for additional factors affecting the decision making process.

Time series variation

In the gross analysis, time series variation was controlled by the simulation methodology. The gross savings were calculated by simulating the building with and without the energy efficiency measures but holding other equipment and schedules fixed as observed. Time-series variation was not an issue in the net-to-gross regression analysis since all observations reflected the same time period. In other words, the regression modeling addressed variation from one same site to another, but not from one time point to another.

Participant self selection

Self selection was addressed in the net-to-gross analysis by developing a logistics model for the probability of participating, and then using the resulting double inverse Mills ratios as added explanatory variables in the efficiency choice models. The statistical significance and effect of the inverse Mills ratios were estimated and reported.

Omitted factors

Two factors might be discussed: the use of Title 24 documentation and billing data. The study sought to use both Title 24 documentation and billing data to the extent practical. When either Title 24 documentation or billing data was available, it was used to improve the accuracy of the engineering models. This approach allowed us to maintain the full sample even when these data were unavailable.

The evaluation of the 1994 NRNC program clearly demonstrated the difficulty of obtaining Title 24 documentation, especially for the non-participants. In order to avoid high refusal rates and the concomitant risk of nonresponse bias, we only insisted on Title 24 documentation for sites that used the tailored lighting approach or the performance-based approach to Title 24 compliance.

Billing data was used to calibrate each individual engineering model whenever possible. However, as described elsewhere, the available billing data did not always reflect the space affected by the new construction. In some of these cases, we sought to supplement the billing data with our own metering. Nevertheless, some of the sites did not have actual usage data. In such cases, the uncalibrated model was used. As discussed earlier in this report, the uncalibrated models were found to produce results that were quite similar to the calibrated models. This analysis confirmed that the pre-calibration models were very accurate.

Estimation of net impacts

The combination of statistical sampling, on-site surveys, site-specific engineering models, econometric analysis, and statistical expansion was carefully designed to provide an unbiased and statistically reliable estimate of net program savings. In particular, the decision-maker survey was designed to isolate self-selection bias and the long-run impact of the program on design practice. The model was specified to include any observable and statistically significant effects of the program on the energy efficiency of both participants and non-participants.

Errors in measuring variables

In the on-site surveys and engineering modeling we sought to obtain an accurate representation of each individual sample site. Past experience suggested that serious errors could arise from failing to model the space in the building actually affected by the new construction, or by failing to accurately describe some of the equipment and schedules of use. The present study addressed these problems by improved training and communication with the auditors, earlier retrieval and review of program files, having the auditors themselves responsible for the data entry and modeling, and having the auditors develop the model for a site soon after completing its survey. The engineering team met with PG&E's program managers and reviewed the site-specific models in detail. We also redesigned the decision-maker survey, streamlined the process used to recruit each site and complete the decision maker survey. All of these measures resulted in much more accurate data going into the econometric analysis than in the prior study.

Autocorrelation

Autocorrelation was not an issue since, as explained above, the analysis was cross sectional.

Heteroscedasticity

Heteroscedasticity – the tendency of larger projects to have greater variation – was addressed in both the sample design and efficiency-choice regression models.

The MBSS methodology used in the sample design addressed heteroscedasticity by modeling the variation in savings as a function of the tracking estimate of savings or the square footage of each site and then using an efficiently stratified sampling plan to increase the probability of selecting large sites. This ensures that the sample is effectively focused where the savings are greatest, while retaining an unbiased representation of small and large projects alike.

The efficiency-choice regression models were specified to minimize the danger of heteroscedasticity by defining the dependent variable as the gross savings as a fraction of the baseline energy use. This specification is closely related to the weighted-least-square methodology resulting from the assumption that the residual variation in gross savings is proportional to the baseline energy use of each site. Graphical scatter plots of the studentized residuals were examined to confirm the absence of Heteroscedasticity.

Collinearity

Multicollinearity is generally a less serious problem in a cross sectional analysis than in a time series analysis. Our methodology was designed to protect against the type of problem that might arise in a cross sectional analysis. Extreme multicollinearity can cause computational problems. Several of the indicator variables used in the regression models were perfectly collinear. This occurred, for example, if a respondent who failed to answer a given question also failed to answer a second question. In this case the missing-response indicators would be perfectly collinear. The SPSS software used in the analysis identifies and reports these instances and automatically drops one of the variables from the analysis. The software also provides a warning if the multicollinearity is strong enough to affect the numerical accuracy of the estimated coefficients. In practice there was no indication of a serious problem with numerical accuracy.

When explanatory variables have strong but not extreme multicollinearity, it is important to guard against obtaining biased results. Omitted-variable bias can arise if one of the correlated variables is dropped from the model. We guarded against this possibility by systematically comparing the estimated coefficients of our various models and looking for other indicators such as large shifts in statistical significance.

Influential data points

We followed diagnostic procedures recommended by Belsley, Kuh and Welsh.[10] Our key indicator of an influential observation was the studentized residual, which can be related to the t-distribution. We also examined normal probability plots, partial-regression leverage plots for each explanatory variable, and other case-specific measures of influence.

Statistical Precision

In each regression model, we used standard logistics or least-squares techniques to calculate the standard error and statistical precision of each coefficient. We used the standard MBSS statistical techniques described in the Gross Savings chapter to expand to the econometric estimates for each sample site to the population and to measure the statistical precision of the results.

Overview of the Econometric Net-to-Gross Methodology

Under the econometric approach, the net-to-gross ratio was calculated in the following seven steps. For simplicity we will discuss the methodology used to analyze annual energy savings. An analogous approach was used to analyze summer peak demand savings.

1) **Dependent Variable:** For each site in the combined participant/non-participant sample, calculate the efficiency choice of each site; this is the difference between the baseline and as-built energy use as a fraction of the baseline energy use. The efficiency choice was the dependent variable, i.e., the *y*-variable, in the regression analysis.

2) **Analysis Data Base:** For each site in the combined participant-non-participant sample, create an indicator variable for program participation, and indicator variables reflecting the responses to the categorical questions in the decision-maker survey. Create indicator variables to identify missing data to each of the decision-maker questions. Create indicator variables to identify the building-type categories. Include the scale response variables from the decision-maker survey as additional potential explanatory variables.

3) **Logistic Regression Model:** Develop a logistic regression model to estimate the probability that each sample site is a participant. Use the preceding indicator variables as well as the scale response variables as possible explanatory variables in the model. Examine the model for outliers and other violations of the assumptions of logistics regression. Drop explanatory variables that are not statistically significant. Use the simplified logistics model to calculate the predicted probability that each site in the combined sample is a participant. Then use the predicted probabilities to calculate double Inverse Mills ratios in order to correct for possible self-selection bias.

4) **Efficiency choice Regression Model:** Formulate a regression model explaining the variation in efficiency choice as a function of various

[10] D. A. Belsley, E. Kuh and R. E. Welsch, *Regression Diagnostics*, Wiley, 1980.

variables describing the participants and non-participants. The explanatory variables included the following:

(a) The indicator variable for program participation,

(b) Indicators describing the type of building,

(c) Indicators for the decision makers planning process and priorities, concern about energy, etc.

(d) Scale variables measuring the degree of interaction with PG&E and the amount of influence PG&E had on the design of this project and of past projects, and

(e) The inverse Mills and double inverse Mills ratios, and

(f) Indicators for potential outliers.

5) **Model Diagnostics and Simplification:** Examine suitable graphs and statistics to determine the adequacy of the regression model. Simplify the regression model by dropping statistically insignificant variables. Add statistically significant interaction variables.

6) **Net Savings:** Use the simplified regression model to estimate the net savings attributable to the program for each sample participant, after statistically controlling for the efficiency choice of non-participants, any significant differences between participants and non-participants in the other explanatory variables, and self selection via the inverse Mills and double inverse Mills ratios. Then use the statistical sampling methods to expand the net savings attributable to the program for each sample participant to the population of 1996 program participants, as described in the Gross Savings chapter. Finally, calculate the error bound and relative precision of the results using the usual statistical sampling methods.

7) **Spillover:** Use the simplified regression model to estimate the spillover effect of the program for each sample non-participant. Then use standard statistical sampling methods to expand the net savings attributable to the program for each sample non-participant to the population of 1996 non-participants, using Dodge new construction data. Finally, calculate the error bound and relative precision of the results using the usual statistical sampling methods.

Database for the Econometric Analysis

The analysis database consisted of 247 sample observations with twenty variables. Forty-seven additional indicator variables were created to reflect the building types, categorical survey information and missing responses to specific questions. Several additional indicator variables were created to represent individual sample sites that appeared to be outliers in the preliminary residual analysis. Additional variables were created within the analysis for statistical interactions, for the Mills ratios, and for various diagnostic tests.

Logistic Regression Model

As previously indicated, the objective of this task was to develop a logistic regression model to estimate the probability that each sample site is a program participant.

Table 37 summarizes the final logistic model. The column labeled B is the regression coefficient for each explanatory variable. A positive value indicates a higher probability of being a program participant whereas a negative value indicates a lower probability. For example, a new building was more likely to be a program participant, whereas a site in climate zone 2 was relatively unlikely to be a program participant.

Explanatory Variable	B	S.E.	Sig
climate zone 2	-1.21	0.63	0.05
project description missing	-7.19	15.37	0.64
new building	1.02	0.35	0.00
state government owned	-0.88	0.42	0.04
low operating cost	-1.61	0.43	0.00
pge influence on current project	0.57	0.11	0.00
previous interaction missing	8.70	15.33	0.57
previous influence	-0.18	0.10	0.08
participation without rebate missing	-0.81	0.49	0.10
which measures without rebates missing	-1.20	0.50	0.02
Constant	-0.20	0.45	0.66

Table 37: Logistic Regression Model

The preceding model was developed in the following steps.

1. Estimate a logistic regression model relating the dependent variable – the indicator of program participation – to all of the potential explanatory variables. Measure the fit, save the diagnostic statistics, and examine the diagnostic graphs. This analysis indicated that there were no outliers or other observable problems with the model.
2. Use backward stepwise logistic regression to eliminate the statistically insignificant variables from the preceding model. Use a p-value of 0.05 for adding variables and 0.10 for deleting variables.
3. Estimate the simplified model shown above, measure its fit, save its diagnostic statistics, and examine its diagnostic graphs.

The following figure is a sequence plot of the studentized residuals. None of the sample sites has a studentized residual greater than 3 or less than –3 and therefore we can conclude that there are no outliers. Moreover, the residuals appear to be random.

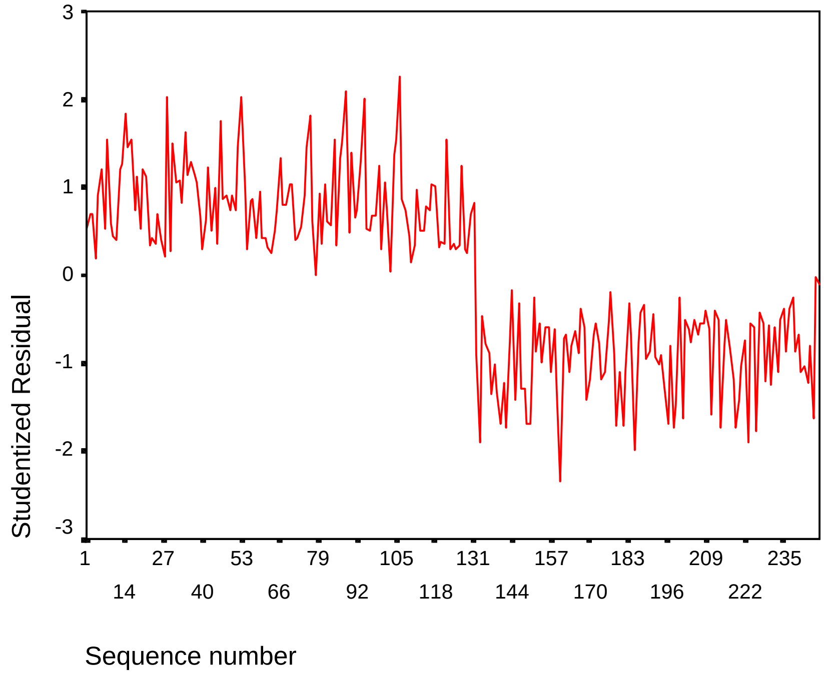

Figure 30: Sequence Plot of Residuals

Figure 30 shows a normal probability plot for the studentized residuals of the model. This is a tool to assess the hypothesis of a normal probability distribution that is the basis of the logistics analysis. If the hypothesis of a normal probability distribution is valid, then the plotted points should lie along the straight line. A failure of the residuals to be normally distributed may be indicated if the plotted points deviate substantially from the line. The figure supports the hypothesis of a normal probability distribution.

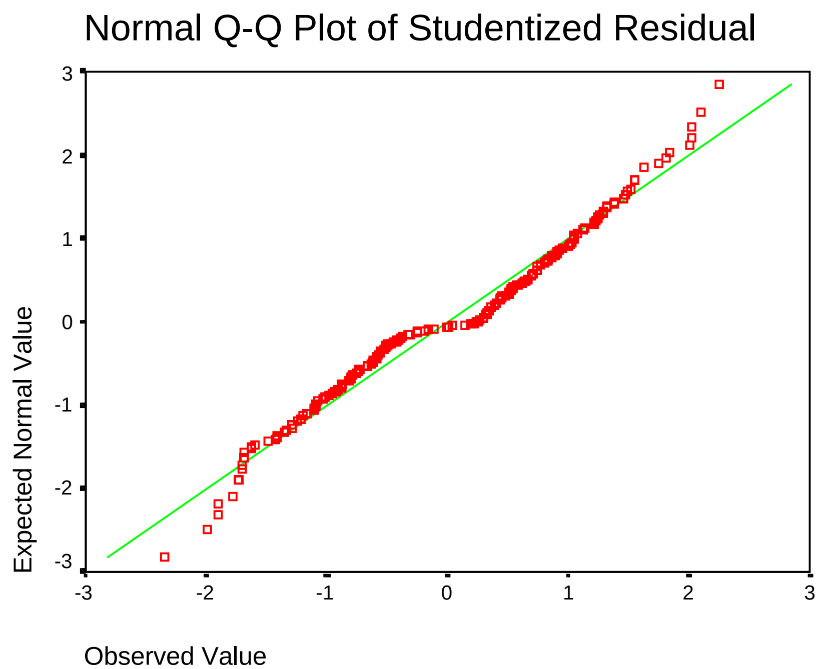

Figure 31: Normal Q-Q Plot of Studentized Residuals

The following table provides a common measure of the goodness of fit of the final model called the classification table. Of the 116 non-participants in the sample, the model correctly predicted that 84 were non-participants, for a score of 72.4% correct. Of the 131 participants, the model correctly predicted that 101 were participants, for a score of 77.1% correct. The overall score was 74.9%.

	Predicted		Percent Correct
Observed	**Non-participant**	**Participant**	
Non-participant	84	32	72.4%
Participant	30	101	77.1%
Overall			74.9%

Table 38: Logistic Model Participation Prediction

Two other measures were calculated reflecting the goodness of fit of the logistics model. The Nagelkerke R-squared statistic was 46% - indicating that the model explained 46% of the total variance. The statistical significance of the model was .0000 – indicating that the model was statistically very significant.

Another way to assess the simplified model is to compare its goodness of fit to the full model developed in step 2 of the analysis. This analysis indicated that the variables that were deleted from the full model were not statistically significant as a group. This suggests that the simplified model is an adequate summary of the relationship between program participation and the variables developed from the decision-maker survey. From all of the preceding analysis, we can conclude that the simplified model is a good predictive model for program participation.

The simplified logistic model was then used to estimate the probability that each site in the sample might have been a participant as a function of the characteristics of the site and the information about the decision-making process. For each site, let *Z* represent the numerical result of substituting the values of the explanatory variables into the logistic equation. Then the estimated probability is calculated using the equation

$$\hat{p} = \frac{e^{Z}}{1+e^{Z}}$$

The inverse Mills ratio was calculated as

$$C = \left[\frac{(1-\hat{p}) \times \ln(1-\hat{p})}{\hat{p}} + \ln(\hat{p}) \right] \qquad \text{for participants, and}$$

$$C = -1 \times \left[\frac{\hat{p} \times \ln(\hat{p})}{1-\hat{p}} + \ln(1-\hat{p}) \right] \qquad \text{for non-participants.}$$

The double inverse Mills ratio was calculated by multiplying *C* by the indicator variable for program participation.[11] These variables were labeled *Mills ratio* and *Double Mills ratio*, respectively.

Annual Energy Regression Model

The objective of this task was to develop a regression model to estimate the efficiency choice of each sample site, participant and non-participant. The efficiency choice of each sample site was measured as the difference between as built and baseline use as a fraction of baseline use.

Table 39 summarizes the final efficiency choice model. The column labeled B is the regression coefficient for each explanatory variable. A positive value indicates a higher efficiency choice whereas a negative value indicates a lower efficiency choice. For example, the model indicates that a program participant tended to have a 0.115 higher efficiency choice than a non-participant. The econometric standard error of this estimate was 0.042 indicating that the error bound at the 90% level of confidence was 1.645 * 0.042 = 0.069. The 90% confidence interval for the true value is 0.115 ± 1.645 * 0.042 = (0.046, 0.184). The program participant coefficient was statistically significant at the .007 level of significance.

Three other explanatory variables based on seven-point scale variables were used in the calculation of energy efficiency. The variable labeled participant awareness measured the decision maker's awareness of the PG&E program, coded 1 (not at all) to 7 (very significant). This variable was equal to 0 for a non-participant. The positive coefficient means that the efficiency choice was greater for a participant who reported a significant awareness of the program compared

[11] Net Savings Estimation: *An Analysis of Regression and Discrete Choice Approaches*, Prepared for the CADMAC Subcommittee on Base Efficiency, Prepared by Xenergy

, Inc. Madison WI, by M. Goldberg and K. Train, Revised March 1996.

to a participant who reported very little awareness of the program. The variable labeled influence on non-participant measured the influence of PG&E on the decisions regarding design and equipment choices for this project, coded 1 (not at all) to 7 (very significant). This variable was equal to 0 for a participant. The positive coefficient means that the efficiency choice was greater for a non-participant who reported that PG&E had a significant influence compared to a non-participant who reported that PG&E had very little influence. This variable measures spillover and was statistically significant at the .010 level of significance. The 'previous influence on participant' variable measured the amount of PG&E influence on decision makers in previous projects, using the same scale as participant awareness. This variable was also equal to 0 for a non-participant. The negative coefficient means that the efficiency choice was lower for a participant who reported significant previous PG&E influence compared to a participant who reported very little previous PG&E influence.

We will discuss the role of these variables in detail in a later section.

The remaining variables represent other factors that were found to have a statistically significant effect on efficiency choice. The model indicates that projects that were built using preexisting plans were less efficient. Two other explanatory variables were 1/0 indicator variables of missing responses to the financial criteria and the previous influence question. The model indicates that Sites 315, 349, and 343 had significantly higher efficiency choice than expected based on other factors and were treated as outliers. The inverse Mills ratio was not statistically significant, indicating that there was no statistically significant correction for self selection. The double Mills variable was also not statistically significant and was not included in the model.[12]

[12] The inclusion or deletion of the two Mills ratio variables had very little effect on the remaining coefficients of the model.

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients		
		B	Std. Error	Beta	t	Sig.
1	(Constant)	.094	.021		4.460	.000
	participant	.115	.042	.380	2.745	.007
	participant awareness	.013	.006	.241	2.215	.028
	previous influence on participant	-.012	.006	-.166	-1.868	.063
	influence on nonparticipant	.020	.008	.211	2.602	.010
	preexisting plans	-.054	.020	-.148	-2.636	.009
	Financial criteria missing	.087	.044	.114	1.972	.050
	previous influence missing	-.129	.055	-.142	-2.363	.019
	outlier site 315	.654	.130	.275	5.016	.000
	outlier site 349	.531	.131	.223	4.062	.000
	outlier site 343	.417	.131	.175	3.190	.002
	MILL	-.003	.013	-.027	-.270	.787

a. Dependent Variable: ENERGY

Table 39: Annual Energy Regression Model

The following table provides several measures of the goodness of fit of the final model. The adjusted R square was .287 indicating that the model explains about 29% of the total variation in efficiency choice. The F-statistic was 9.995, corresponding to a statistical significance of 0.000, indicating that the model as a whole was highly significant.

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.565	.319	.287	.1278529

Table 40: Annual Energy Model Summary

The figure below shows a normal probability plot for the deviancies of the final model. This is a tool to assess the hypothesis of a normal probability distribution that is the basis of the efficiency-choice regression analysis. If the hypothesis of a normal probability distribution is valid, then the plotted points should lie along the straight line. The figure suggests that this assumption is generally valid.

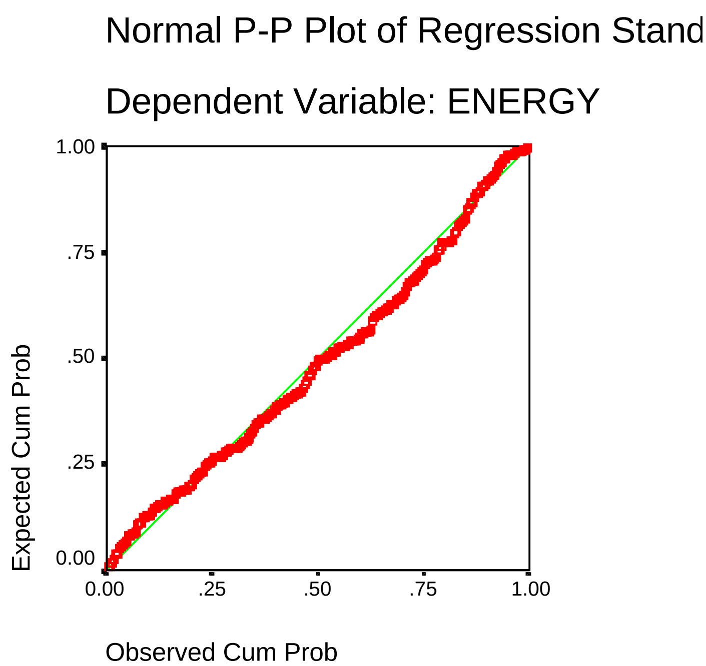

Figure 32: Normal P-P Plot of Annual Energy Residuals

The following figure shows a more conventional histogram of the standardized residuals of the model. Again the assumption of a normal distribution appears to be generally satisfactory. This evidence, together with the relatively large size of the sample, indicates that standard measures of statistical significance should be valid.

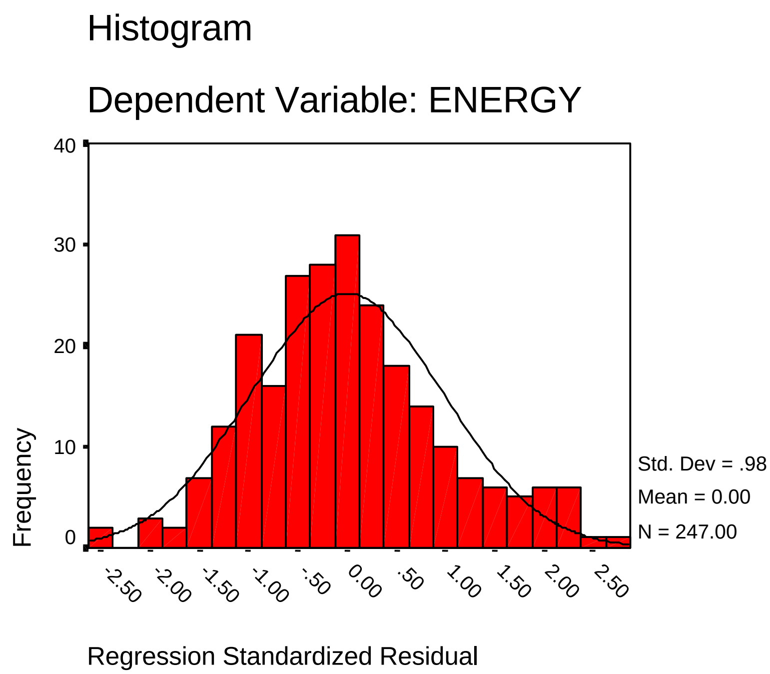

Figure 33: Annual Energy Model Residual Histogram

The following figure is a sequence plot of the studentized residuals of the final model. The residuals appear to be random and to indicate no remaining outliers.

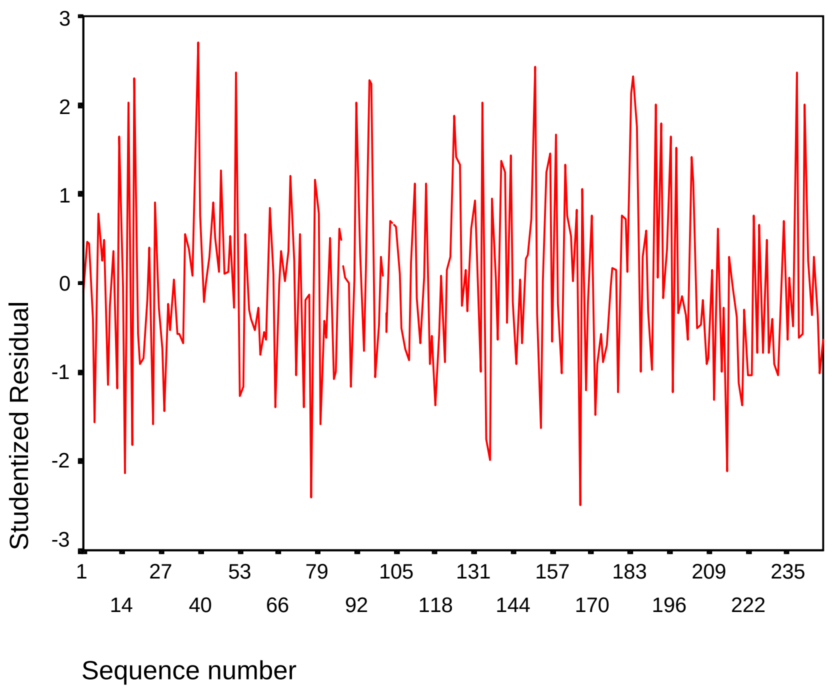

Figure 34: Sequence Plot of Residuals

The remaining graph shows a scatter plot of the residuals compared to predicted values. The important issue is not the range of predicted values on the horizontal axis, but rather the range of the residuals on the vertical axis. Again this graph shows that the residuals are randomly distributed. Moreover, it shows that the residuals are homoscedastic. In other words, the variance of the residuals seems to be independent of the predicted values.

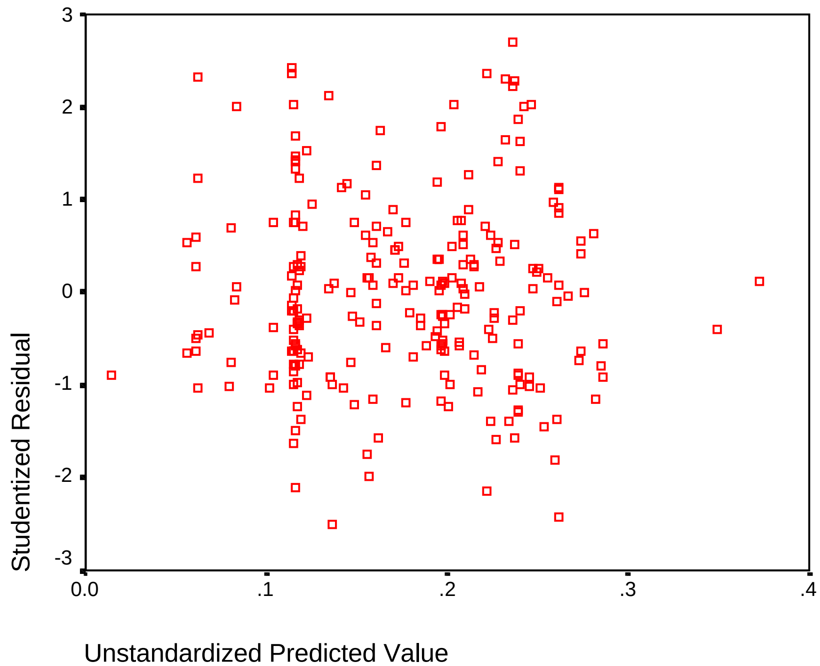

Figure 35: Scatter Plot of Predicted Values and Residuals

Another way to assess the simplified model is to compare its goodness of fit to the full model developed in the first step of the analysis. This analysis indicated that the variables that were deleted from the full model were not statistically significant as a group. This suggests that the simplified model is an adequate summary of the relationship between efficiency choice and the variables developed from the decision-maker survey. From all of the preceding analysis, we can conclude that the simplified model is a good predictive model for efficiency choice.

Comparison of Models

In seeking the most complete and parsimonious model for the energy efficiency choice, a sequence of regression models were examined. The following general steps were followed to obtain the final efficiency-choice model.

1. Estimate a linear regression model relating the dependent variable – the efficiency choice of each site – to all of the potential explanatory variables. Measure the fit, save the diagnostic statistics, and examine the diagnostic graphs. This analysis suggested that sites 315 and 349 might be outliers.

1B. Rerun the preceding model with indicator variables for the outliers. Observe the statistical significance of these indicator variables, Measure the fit, save the diagnostic statistics, and examine the diagnostic graphs. This analysis indicated that the model was well specified.

2. Use backward stepwise regression to eliminate the statistically insignificant variables from the preceding model. Use a p-value of 0.05 for adding variables and 0.10 for deleting variables.
3. Estimate the simplified model from the stepwise regression, adding an indicator variable for program participation adding and dropping two variables that were not statistically significant. Measure its fit, save its diagnostic statistics, and examine its diagnostic graphs. This analysis suggested that site 343 might be an outlier.
4. Use backward stepwise regression to eliminate the statistically insignificant variables from the preceding model with outlier site 343 added in. Use a p-value of 0.05 for adding variables and 0.10 for deleting variables.
5. Estimate the simplified model from the stepwise regression. Measure its fit, save its diagnostic statistics, and examine its diagnostic graphs. The results indicated that the model was well specified. This analysis gave the final regression reported above.
6. Add Mills to the previous simplified model get the final model.
7. Add Mills 2.

The following table shows the coefficients of the program participant and participant interaction variables and their standard errors for each of the models that were estimated. These coefficients are important because they determine the net savings estimated from the regression model. In other words, any bias in estimating these regression coefficients may produce a bias in the final estimate of net savings.

The table traces how the value of the coefficient changed as various variables were added or dropped. All of the models were based on the same underlying data. Models 1 through 5 trace the steps that were taken to obtain the final model. Model 6 is the final model itself.

In models 1-5 we were seeking (a) to identify and deal with outliers that might bias the results, and (b) to simplify the model. The approach was to start with a full model reflecting all candidate explanatory variables, look at the various diagnostic statistics and graphs to check the validity of the model, and introduce corrections to any problems that are indicated. Our objective was to get a good model that passes the diagnostic statistics before working to refine the model.

The first model included all of the candidate explanatory variables. The studentized residuals of this model indicated that two sites were potential outliers, using 3.0 as the critical value. Model 1B was similar to model 1 but included indicator variables for these two added outliers. Both of these models suffered from multicolinearity, as can be seen by the large standard error for the participant variable. The results indicated that the model was well specified. Model 2 was the backward stepwise regression with the two outliers identified in Model 1 added in. Model 3 was the simplified model obtained from the backward stepwise regression. The studentized residuals of Model 3 indicated that one more site was a potential outlier, using 3.0 as the critical value. Model 4 was a second backward stepwise regression with an indicator variable added in for the potential outlier identified in Model 3. Model 5 was the simplified model obtained from the second backward stepwise regression.

Model 6 was obtained by adding the Mills ratio and became the final model. Model 7 was obtained by adding the double Mills ratio to model 6. These models show that the Mills and Double Mills ratios have very little effect on the coefficients.

Model	Participant		Awareness		Previous Influence		Description
	B	SE	B	SE	B	SE	
E1	0.600	0.277	0.005	0.006	-0.014	0.013	Mulitcollinearity and outliers
E1B	0.394	0.262	0.008	0.008	-0.019	0.012	Multicollinearity
E2	0.075	0.032	0.012	0.006	-0.013	0.006	Stepwise Regression Model
E3	0.075	0.032	0.012	0.006	-0.013	0.006	Simplified Model from E2
E4	0.109	0.035	0.013	0.006	-0.012	0.006	2nd Stepwise with added outlier
E5	0.109	0.035	0.013	0.006	-0.012	0.006	Simplified Model from E4
E6	**0.115**	**0.042**	**0.013**	**0.006**	**-0.012**	**0.006**	**New Final Model (Mills Added)**
E7	0.122	0.042	0.013	0.006	-0.013	0.006	Add Mills 2

Table 41: Model Development Summary

Analysis of Program Impact and Spillover, Annual Energy

The final energy efficiency model was described in an earlier section. The first three explanatory variables reflected program participation and the interaction and influence of PG&E with the decision-maker on the current project. The remaining variables all reflected factors other than the program. So the analysis of the program impact and possible spillover is based on the first three variables in the model. By using the multivariate regression model, these results are adjusted for non-program factors that appear to influence the efficiency choice.

The efficiency choice regression model can be written as follows:

Expected efficiency = .115 * Participant + .013 * participant awareness

- .012 * previous influence on participant

+ .020 * influence on non-participant + other factors

Here, the participant variable was 1 for a participant and 0 for a non-participant. The participant awareness was measured on a seven point scale, with 1 indicating very weak interaction and 7 indicating very strong interaction. The participant awareness was equal to 0 for a non-participant. Similarly, influence on non-participant was measured on a seven point scale, with 1 indicating very weak influence and 7 indicating very strong influence. This variable was equal to 0 for a participant. Previous influence on participants was measured using the same scale as the participant awareness variable.

The energy efficiency model can be used to estimate the impact of the program on any particular sample site. This is done by calculating the difference between the expected energy efficiency predicted by the model and the energy efficiency that would be expected for the site in the absence of the program. In the absence of the program, the participant variable would be equal to 0. Moreover, in the absence of the program we can set the rated awareness to the lowest value of response, i.e., to 1. Under these assumptions, the impact of the program on expected energy efficiency can be calculated for a program participant as

Added Efficiency = 0.115 + 0.013 * (participant awareness –1)

- 0.012 * (previous influence on participant–1)

In other words, for a participant, the program increased the expected building efficiency by 0.115 plus 0.013 times the level of awareness of the rebates less the level of awareness expected in the absence of the program. In addition, the model indicates an expected decrease in building efficiency by 0.012 times the level of previous PG&E influence less the level of previous influence expected in the absence of the program

For a non-participant, the energy efficiency model implies that the program increased the expected energy efficiency by:

Added Efficiency = 0.020 * (influence on non-participant –1)

In other words, for a non-participant, the program increased the expected building efficiency by 0.020 times the level of influence of PG&E on non-participants less the level of influence expected in the absence of the program.

Table 42 shows the added efficiency due to the program for both participants and non-participants. The top portion of the table represents the impact of the program on expected energy efficiency for a program participant. Levels of participant awareness are on the horizontal axis, and the vertical axis represents the previous PG&E influence on participants. The bottom portion of the table represents the influence of the program on expected energy efficiency for non-participants. The values in the table show the increase in expected efficiency due to the program, for both participants and non-participants, evaluated using the preceding two equations.

		Participant Awareness						
		1	**2**	**3**	**4**	**5**	**6**	**7**
Previous Influence	**1**	**0.115**	0.128	0.141	0.154	0.167	0.18	0.193
	2	0.103	**0.116**	0.129	0.142	0.155	0.168	0.181
	3	0.091	0.104	**0.117**	0.13	0.143	0.156	0.169
	4	0.079	0.092	0.105	**0.118**	0.131	0.144	0.157
	5	0.067	0.08	0.093	0.106	**0.119**	0.132	0.145
	6	0.055	0.068	0.081	0.094	0.107	**0.12**	0.133
	7	0.043	0.056	0.069	0.082	0.095	0.108	**0.121**
		Non-participant Influence						
		1	**2**	**3**	**4**	**5**	**6**	**7**
		0.000	0.020	0.040	0.060	0.080	0.100	0.120

Table 42: Added Efficiency Due to Program

The top portion of table 42 shows that for a participant with very weak awareness and little previous influence by PG&E, the program increased the expected efficiency by .115. In other words, the percent efficiency of the site relative to baseline was .115 higher than in the absence of the program. If the participant had very strong interaction with PG&E and lots of previous influence, the program increased the expected efficiency by .121. The top portion of table 42 also shows that the added efficiency can be less than .115. This is due to the negative coefficient on the previous influence variable. However, for all but 3 of the 128 participants, the added effect was actually .094 or higher. For the majority of the participants, the expected savings were greater than .115.

The bottom portion of table 42 indicates that for a non-participant that was very weakly influenced by PG&E, there was no increased efficiency due to the program, but for a non-participant that was very strongly influenced by PG&E, the program increased the expected efficiency by .120.

This suggests that the program has two impacts. First the program has a direct net impact on the participants. Second, the program appears to have an indirect or spillover impact on the non-participants.

The next step was to use the energy-efficiency regression model to estimate the net direct impact of the program. For each participant we calculated the net annual kWh savings due to the program by multiplying the base annual energy use of the site by the estimated increase in efficiency due to the program, calculated from the preceding equation. Then these results were expanded to the population of program participants.

The final step was to use the energy-efficiency regression model to estimate the spillover impact of the program. In this analysis, we worked with the non-participants in the sample. For each non-participant, we calculated the net annual kWh savings due to the program by multiplying the base annual energy use of the site by the estimated increase in efficiency due to the program, calculated from the preceding non-participant equation. Then we used the Dodge database to expand the sample non-participants to the population of new construction. To ensure a conservative estimate of spillover, we made an adjustment to factor out any participant sites that may have been present in the non-participant population. To accomplish this, the sample spillover was projected to both the participant and new construction populations and the participant population estimate was subtracted from the new construction population estimate.

Table 43 shows the net savings estimate and the estimate of spillover, together with the relative precision of each estimate[13]. The difference-of-differences net savings was 39,046,000 kWh. The econometric approach yielded a direct net savings of 58,569,164 kWh. The difference between the two estimates can be thought of as the effect of self-selection bias not accounted for in the difference-of-differences approach. Spillover among non-participants would also give a downward bias to the difference of differences estimate of program net savings.

[13] Spillover is not being claimed and is not included in the net savings results reported elsewhere in this report.

	Estimate	Net-to-Gross Ratio	Relative Precision
Net Savings of Participants	58,569,164. KWh	69.8%	±10.5%
Spillover in NP population	44,112,699. KWh	--	±22.7%
Spillover in P population	23,713,117. KWh	--	±19.7%
Spillover Estimate	20,400,582. KWh	--	±54.2%
Total Net Savings	78,969,746. KWh	94.0%	±16.0%

Table 43: Net Energy Savings and Spillover Estimates

Summer Demand Regression Model

The objective of this task was to develop a regression model to estimate the summer demand efficiency choice of each sample site. The analysis followed the same steps as the efficiency choice for annual energy, reported in the preceding sections. The dependent variable is the summer peak demand savings divided by the summer peak baseline demand of each model.

Table 44 summarizes the final efficiency choice model for summer peak demand. Many of the variables and coefficients in this model are similar to the efficiency choice model for energy. Again, the efficiency choice was found to be a function of both participant interaction and influence on non-participants. The role of these variables in the model will be discussed in detail in the next section. In addition, the model included outlier indicator variables for three sites,

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients		
		B	Std. Error	Beta	t	Sig.
1	(Constant)	.074	.021		3.458	.001
	participant	.104	.040	.361	2.626	.009
	participant interaction	.016	.006	.295	2.878	.004
	previous influence on participant	-.012	.006	-.181	-2.1	.038
	influence on nonparticipant	.017	.007	.190	2.410	.017
	climate zone 1	-.198	.069	-.174	-2.9	.004
	owned by corporation	.046	.024	.102	1.915	.057
	owned by state govmt	.042	.020	.116	2.097	.037
	preexisting plans	-.039	.019	-.112	-2.1	.037
	Financial criteria missing	.080	.041	.110	1.969	.050
	previous influence missing	-.106	.050	-.123	-2.1	.033
	outlier site 315	.816	.137	.360	5.966	.000
	outlier site 349	.518	.119	.229	4.355	.000
	outlier site 204	.386	.118	.170	3.272	.001
	MILL	-.004	.012	-.031	-.307	.759

a. Dependent Variable: DEMAND

Table 44: Summer Peak Demand Regression Model

Table 45 shows several measures of the goodness of fit of the model. The adjusted R square was .343 indicating that the model explains over 34% of the total variation in efficiency choice. The F-statistic was 10.173, corresponding to a statistical significance of 0.000, indicating that the model as a whole was highly significant.

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.617	.380	.343	.1169426

Table 45: Demand Model Summary

Figure 35 shows a normal probability plot for the deviancies of the final model. This is a tool to assess the hypothesis of a normal probability distribution that is the basis of the regression analysis. If the hypothesis of a normal probability distribution is valid, then the plotted points should lie along the straight line. The figure suggests that this assumption is valid.

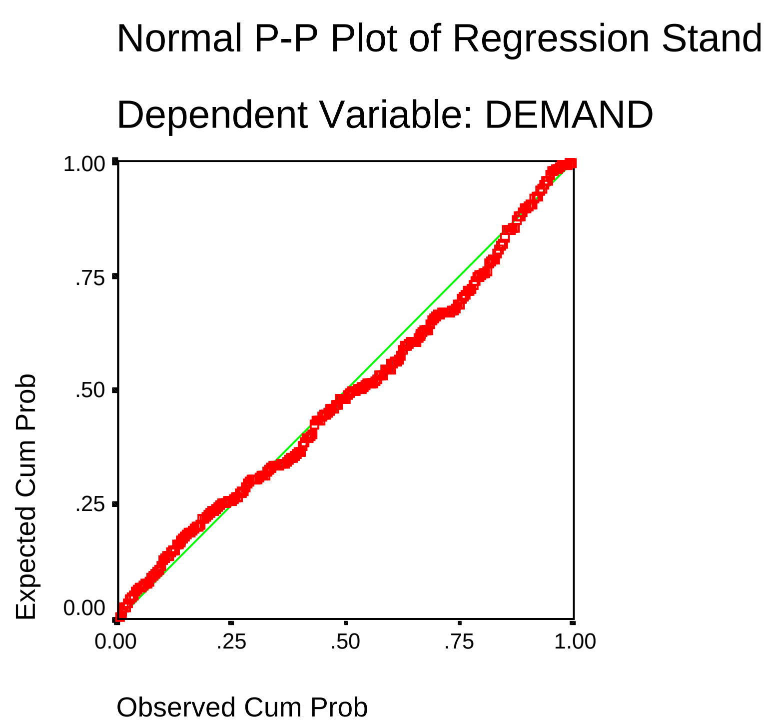

Figure 36: Normal P-P Plot of Regression Residuals

Figure 36 shows a more conventional histogram of the standardized residuals of the model. Again the assumption of a normal distribution appears to be generally satisfactory. This evidence, together with the relatively large size of the sample, indicates that standard measures of statistical significance should be valid.

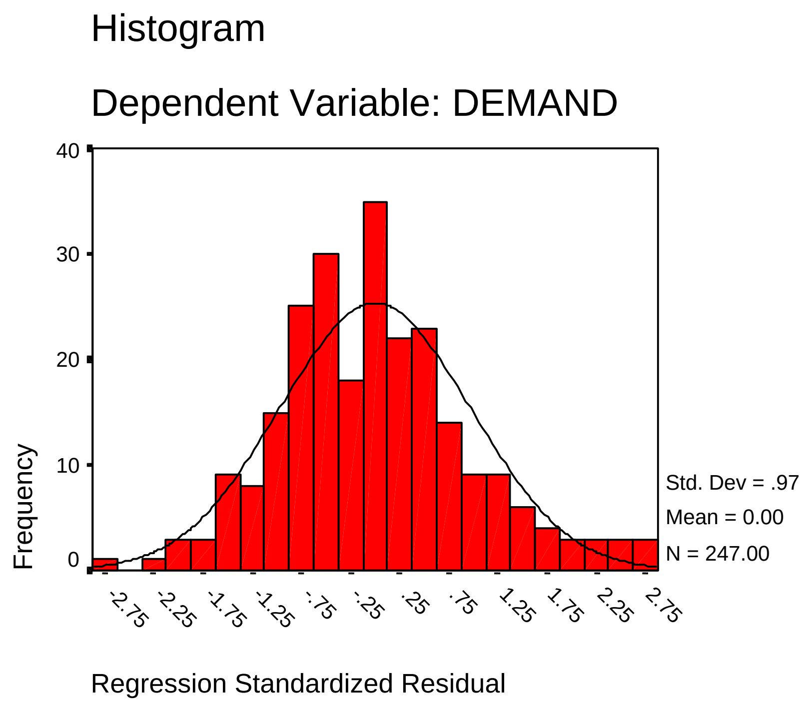

Figure 37: Demand Model Histogram

The following figure is a sequence plot of the studentized residuals of the final model. The residuals appear to be random and to indicate no remaining outliers.

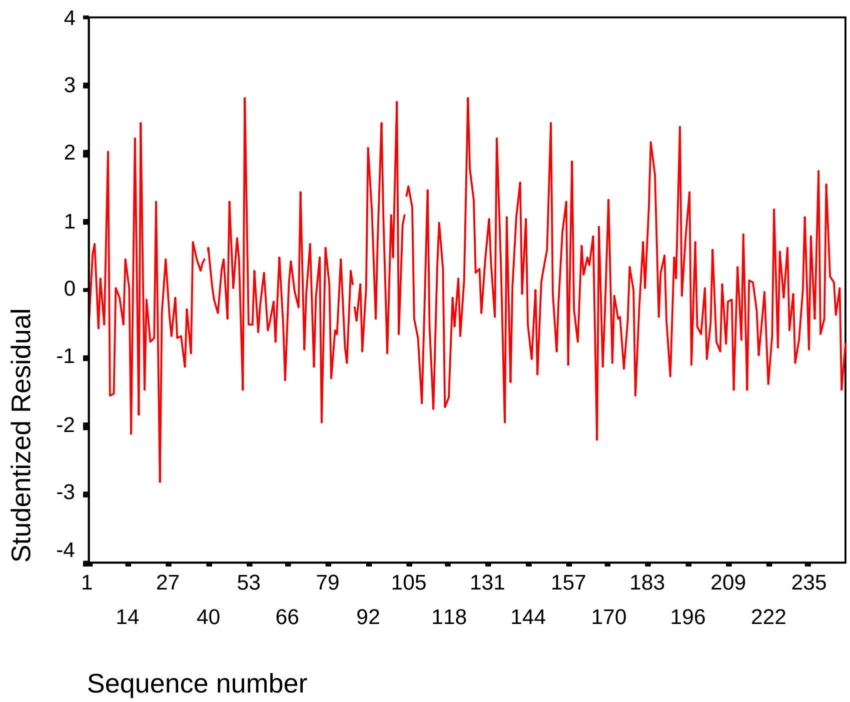

Figure 38: Demand Model Residual Plot

The remaining graph shows a scatter plot of the residuals compared to predicted values. The important issue is not the range of predicted values on the horizontal axis, but rather the range of the residuals on the vertical axis. Again this graph shows that the residuals are randomly distributed. Moreover, it shows that the residuals are homoscedastic. In other words, the variance of the residuals seems to be independent of the predicted values.

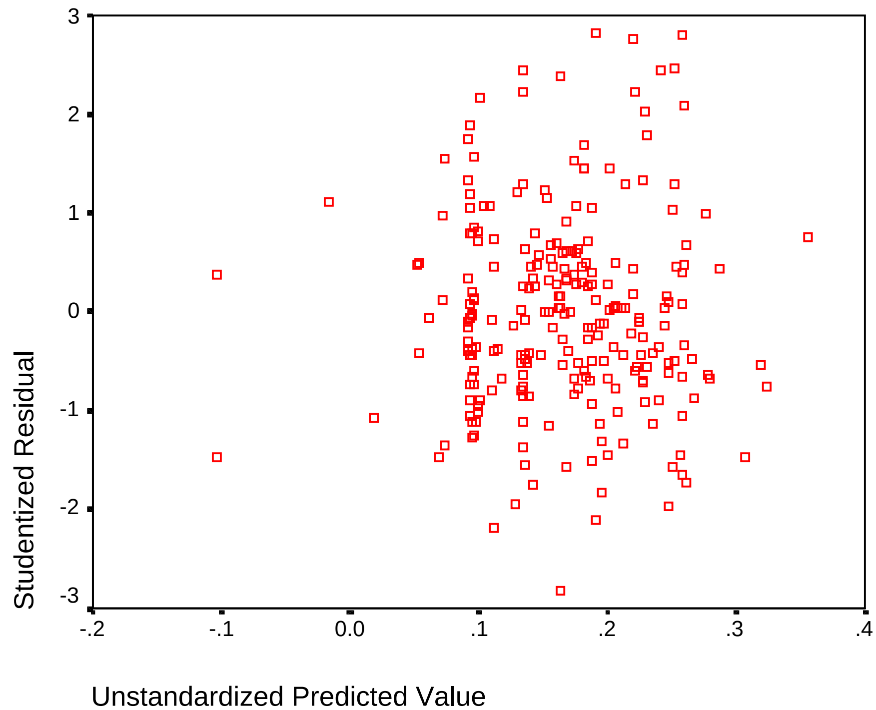

Figure 39: Demand Model Residual Scatter Plot

Analysis of Program Impact and Spillover, Peak Demand

The final energy efficiency model was described in an earlier section. Four explanatory variables in the model reflected program participation, interaction and previous influence of PG&E on the participant decision maker, and PG&E influence on non-participants on the current project. The remaining variables all reflected factors other than the program. So the analysis of the program impact and possible spillover is based on the four variables specified above. By using the multivariate regression model, these results are adjusted for non-program factors that appear to influence the efficiency choice.

For peak demand, the efficiency choice regression model can be written as follows:

Expected efficiency = .104 * Participant + .016 * participant interaction

- .012 * previous influence on participant

+ .017 * influence on non-participant + other factors

Here, the participant variable was 1 for a participant and 0 for a non-participant. The participant interaction was measured on a seven point scale, with 1 indicating very weak interaction and 7 indicating very strong interaction. The participant interaction was equal to 0 for a non-participant. Similarly, influence on non-participant was measured on a seven point scale, with 1 indicating very weak influence and 7 indicating very strong influence. This variable was equal to 0 for a participant. Previous influence on participants was measured on a seven point scale, comparable to the participant interaction scale described above.

The energy efficiency model can be used to estimate the impact of the program on any particular sample site. This is done by calculating the difference between the expected energy efficiency predicted by the model and the energy efficiency that would be expected for the site in the absence of the program. In the absence of the program, the participant variable would be equal to 0. Moreover, in the absence of the program we can set the rated influence or interaction to the lowest value of response, i.e., to 1. Under these assumptions, the impact of the program on expected energy efficiency can be calculated for a program participant as

$$\text{Added Efficiency} = 0.104 + 0.016 * (\text{participant interaction} - 1)$$
$$- 0.012 * (\text{previous PG\&E influence on participants} - 1)$$

In other words, for a participant, the program increased the expected building efficiency by 0.104 plus 0.016 times the level of interaction with PG&E less the level of interaction expected in the absence of the program. The model also indicates that for participants, the program decreased the expected building efficiency by 0.012 times the level of previous PG&E influence less the level of previous influence expected in the absence of the program.

Similarly for a non-participant, the energy efficiency model implies that the program increased the expected energy efficiency by:

$$\text{Added Efficiency} = 0.017 * (\text{influence on non-participant} - 1)$$

In other words, for a non-participant, the program increased the expected building efficiency by 0.017 times the level of influence of PG&E less the level of influence expected in the absence of the program.

Table 46 shows the added efficiency due to the program for both participants and non-participants. The top portion of the table represents the impact of the program on expected energy efficiency for a program participant. Levels of participant interaction are represented on the horizontal axis, and the vertical axis represents the previous PG&E influence on participants. The bottom portion of the table represents the impact of the program on expected energy efficiency for non-participants. The values in the table show the increase in expected efficiency due to the program, for both participants and non-participants, evaluated using the preceding two equations.

		Participant Interaction						
		1	**2**	**3**	**4**	**5**	**6**	**7**
Previous Influence on Participant	**1**	**0.104**	0.120	0.136	0.152	0.168	0.184	0.200
	2	0.092	**0.108**	0.124	0.140	0.156	0.172	0.188
	3	0.080	0.096	**0.112**	0.128	0.144	0.160	0.176
	4	0.068	0.084	0.100	**0.116**	0.132	0.148	0.164
	5	0.056	0.072	0.088	0.104	**0.120**	0.136	0.152
	6	0.044	0.060	0.076	0.092	0.108	**0.124**	0.140
	7	0.032	0.048	0.064	0.080	0.096	0.112	**0.128**
		Non-participant Influence						
		1	**2**	**3**	**4**	**5**	**6**	**7**
		0	0.017	0.034	0.051	0.068	0.085	0.102

Table 46: Added Efficiency in Peak Demand Due to Program

The top portion of table 46 shows that for a participant with very weak interaction and very little previous PG&E influence, the program increased the expected efficiency by .104. In other words, the percent efficiency of the site relative to baseline was .104 higher than in the absence of the program. If the participant had very strong interaction and strong previous influence from PG&E, the program increased the expected efficiency by .128. The top portion of table 46 also shows that the added efficiency can be less than .104. This is due to the negative coefficients on the previous influence variables. However, for all but 4 of the 128 participants, the added effect was actually .092 or higher. For the majority of the participants, the expected savings were greater than .104

The bottom portion of table 46 indicates that for a non-participant that was very weakly influenced by PG&E, there was no increased efficiency due to the program, but for a non-participant that was very strongly influenced by PG&E, the program increased the expected efficiency by .102.

This suggests that the program has two impacts. First the program has a direct net impact on the participants. Second, the program appears to have an indirect or spillover impact on the non-participants.

The next step was to use the demand-efficiency regression model to estimate the net direct impact of the program. For each participant we calculated the net annual kWh savings due to the program by multiplying the base annual energy use of the site by the estimated increase in efficiency due to the program, calculated from the preceding equation. Then these results were expanded to the population of program participants.

The final step was to use the demand -efficiency regression model to estimate the spillover impact of the program. In this analysis, we worked with the non-participants in the sample. For each non-participant, we calculated the net annual kWh savings due to the program by multiplying the base peak demand of the site by the estimated increase in efficiency due to the program, calculated from the preceding non-participant equation. Then we used the Dodge database to expand the sample non-participants to the population of new construction. To ensure a conservative estimate of spillover, we made an adjustment to factor out any participant sites that may have been present in the non-participant population. To accomplish this, the sample spillover was projected to both the participant and new construction populations and the participant population estimate was subtracted from the new construction population estimate.

Table 47 shows the net demand savings estimate and the estimate of spillover, together with the relative precision of each estimate[14]. The difference of differences net savings was 8,200 kW. The econometric approach yielded a direct net savings of 13,951 kW. The difference between the two estimates can be thought of as the effect of self-selection bias not accounted for in the difference of differences approach. Spillover among non-participants would also give a downward bias to the difference of differences estimate of program net savings.

	Estimate	**Relative Precision**
Net Savings of Participants	13,951. kW	±12.1%
Spillover in NP population	9,785. kW	±20.1%
Spillover in P population	5,215. kW	±17.4%
Spillover Estimate	4,570. kW	±47.4%
Total Net Savings	18,521. kW	±14.8%

Table 47: Net Demand Savings and Spillover

[14] Spillover is not being claimed and is not included in the net savings results reported elsewhere in this report.

Recommendations for Future Studies

The methodology used for this study has proven to be very successful. RLW Analytics and AEC were able to collect and analyze large amounts of detailed data quickly using this methodology. To be sure, this was not an inexpensive endeavor, but it has produced characteristic and energy use information that is also very valuable for studies of market transformation, building characteristics, and other market research.

The key improvements made here from the 1994 PG&E/SCE evaluation include:

- **The use of the same staff to survey buildings and build engineering models**. This approach allowed RLW Analytics and AEC to build much better models because the data was collected with a full understanding of the needs of the models. Also, because the person who developed the model was on-site, they could do a much better “reality check” on the model results.
- **Building the engineering model shortly after the site visit**. In the 1994 study, several months passed before the modeling staff could review the field data, greatly increasing the chance that errors could not be adequately corrected. In this study, the initial models were built within days or weeks of the site visit. This, combined with the point above, greatly improved the quality of the models because the building was much fresher in the mind of the modeler.
- **The use of scale variables in the econometric models**. In the 1994 study, a binary variable was used to indicate “partial participation” (a non-participant with spillover). This crude approach to a subtle issue contributed to the econometric model’s inability to identify non-participant spillover. In this study, a series of scale variables were used to isolate spillover. This more sensitive approach was successful in measuring “partial participation.”
- **Experienced construction professionals were used to recruit and survey design professionals and building owners**. The use of staff who understood the industry was the primary reason that such a high participation rate was observed. This also helped with survey completion and data quality because the respondents felt as though the surveyor understood the subject matter and could speak on their level.

Most of the cost and effort in this study involved the data collection and engineering model building tasks. To the extent that CADMAC sponsored regulatory studies like this one continue after January 1, 1998, there are several steps that could be taken in those areas to improve the cost effectiveness of the study:

- **Improvements in the model building software**. Further work to integrate the data entry, model building, and calibration modules of the software would increase the throughput and reduce the human intervention needed to turn survey data into DOE models. Because this system was developed independently of this project, work on these issues is ongoing. Future studies using the RLW Analytics / AEC team would benefit from these improvements.

- **Electronic data entry**. Related to the above point, the use of handheld computers to record survey data would streamline data entry and move quality control checks to the survey site, where the errors could most easily and accurately be corrected.
- **"Codify" engineering judgement**. A major factor in the data collection cost was the use of experienced engineers to collect the data. To the extent that some of the engineering judgement could be captured in the software, lower cost staff could be used in the data collection. This is a fine line to walk, as reductions in surveyor experience and skill could contribute to degradation in the quality of data.
- **Capture decision-maker data as the program runs.** One of the biggest challenges in this type of study is to ask a decision-maker about events that occurred as long as two years prior. The data collected for the econometric analysis could be significantly improved by collecting this data at the time the project is done. This would require a standard survey to be developed by CADMAC and administered by the utility sponsoring the program.
- **Revision of the CADMAC protocols on sampling**. A revision of the sampling protocols would benefit future studies. The CADMAC committee approved this variance from the protocols on this study as well as the 1996 Southern California Edison NRNC study and the combined 1994 PG&E/SCE NRNC study. The results of the study show that this sampling approach is effective in capturing the required information at a significantly lower cost than would be required by a sample complying with the current protocol.