Evaluation of 2020-2021 Commercial Strategic Energy Management Program

Puget Sound Energy

Date: 2/21/2023

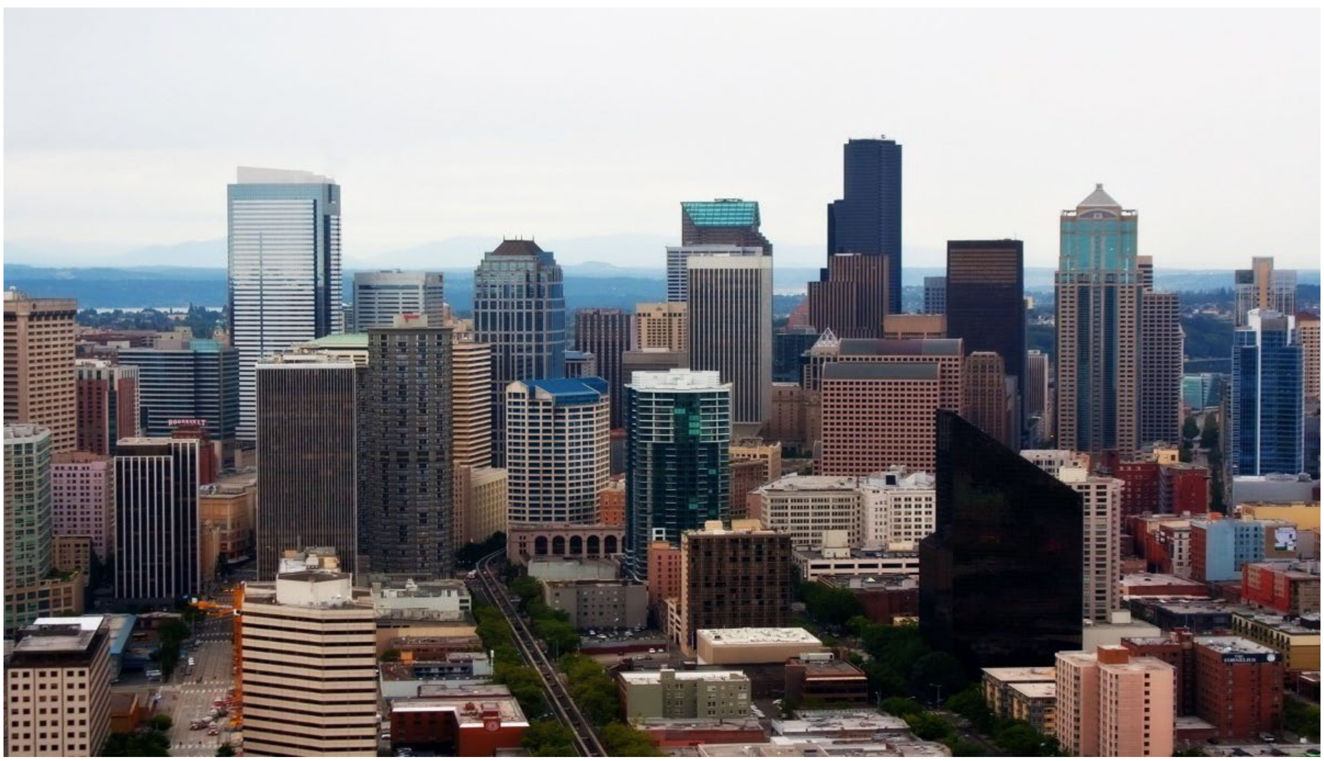


Table of Contents

List of Figures

List of Tables

Primary Contact Amit Kanungo Principal Engineer, DNV Energy Markets & Risk **amit.kanungo@dnv.com** Mobile +1-415-516-6504 Direct +1 510-891-0446 x 44123
Project Manager Andrew Wood Principal Consultant, DNV Energy Markets & Risk andrew.w.wood@dnv.com Direct+1 971-334-0035
PSE Contract Manager Kasey Curtis Sr. Market Analyst Strategic Planning, Evaluation and Research, Puget Sound Energy kasey.curtis@pse.com (425) 424-6642

1 EXECUTIVE SUMMARY

Puget Sound Energy (PSE) hired DNV to complete an independent evaluation of the program years 2020-21 commercial strategic energy management (CSEM) program. This report presents the methods, results, and findings of the evaluation of this program. The goal of the evaluation was to independently estimate program savings performance and identify opportunities to improve each of the evaluated programs.

1.1 Background and approach

DNV initiated this evaluation in 2020. Due to the impact of the COVID-19 pandemic on the program and DNV's initial findings, DNV and PSE agreed to extend the evaluation through 2022. This extension allowed DNV to observe changes to the program and further evaluate the energy savings claimed prior to any pandemic disruptions. The impact evaluation attempted an independent estimate of the ratio of energy savings being realized by the program to the energy savings tracked by PSE, referred to as the program realization rate. Impact evaluation methods were based on the program design, measures offered, and historic program performance. In general, the program was evaluated based on our review of program documentation and a representative sample of completed projects. The process evaluation relied on interviews with sampled program participants.

1.2 Evaluation results

The primary results of our evaluation are program realization rates estimated through our impact evaluation activities. These realization rates are an independent estimates of the ratio of achieved savings to tracked savings for the 2020-21 biennium.

Table 1-1. Evaluated program realization rates

Compliance Program	PSE Program	Electricity Savings (kWh)		Gas Savings (therms)	
		Unweighted Realization Rate, kWh	Weighted Realization Rate, kWh	Unweighted Realization Rate, therms	Weighted Realization Rate, therms
Commercial Strategic Energy Management	**CSEM**	84%	144%	49%	75%

1.3 Key evaluation findings and recommendations

This section provides key findings and recommendations resulting from DNV's evaluation. Additional findings are presented within each program-specific section. DNV's findings and recommendations use the Consortium for Energy Efficiency's (CEE) Strategic Energy Management Minimum Elements[1] as a framework.

1.3.1 Customer Commitment

1) **Key Finding –** While some SEM participants articulated their organization's energy goals and could describe the energy team in place to achieve them, many participants stated that no energy savings goals were currently in place and that turnover in staff (or other reasons) had reduced the energy team to a single person. While the SEM program does set a savings reduction goal with the participant, this was often not tied or integrated into the customer's goals. In addition, few participants articulated that a senior or executive manager at the organization had visibility into the SEM efforts and only a few had reporting requirements outside of annual energy spending. DNV finds that the customer commitment to SEM was lower than is expected for an SEM program.

   - **Recommendation** – PSE should consider adjustments to its program design to increase SEM visibility within a customer's executive management structure, ensure that an appropriately sized energy management team is

[1] Available to the public at: https://library.cee1.org/system/files/library/11283/SEM_Minimum_Elements.pdf

maintained during the engagement, and the agreed goals for the engagement are included in the organization's goals for energy savings. If these are not in place and customers do not demonstrate commitment, PSE should consider withholding performance payments and not attribute to savings to the program.

1.3.2 Planning and Implementation

Project documentation for the CSEM program contains all the required elements of an SEM program, an agreement between the site and the program administrator, documentation of the baseline regression model, documentation of the energy savings calculation, and documentation of actions taken during the performance period. However, DNV believes there are required changes to improve this documentation and better manage the program's savings attribution risk. Each of the recommendations below will increase the customer burden and cost to participate in the program. DNV believes that increasing the cost to participate will ensure participant organizations are committed to the goals and expectations of program. Uncommitted customers will self-select out of the program.

2) **Key Finding –** DNV did not identify any Resource Management Plans (RMP) or similar documentation stating the current energy management practices of the organization at the start of the engagement period and the goals or changes to the practice that the program would help facilitate or encourage. Therefore, there is no information for the evaluator that records the counterfactual to program participation specific to energy management practices. This reduces DNV's confidence that the reductions in consumption estimated can be attributed to the program. This is often referred as an Energy Management Assessment (EMA). In the CSEM grant agreement, the RMP is something the participant "should" update according to the agreement and is similar to a facility's standard operating procedure.

    - **Recommendation** – PSE should require the completion of the RMP or an EMA at the start of any engagement or renewal period. This assessment could be completed by the customer alone, facilitated by PSE, or facilitated by a third party. The purpose of the assessment is to understand current practice and identify improvements to energy management that the program can support. By supporting these improvements, PSE can attribute measured savings during each performance period to its program. If a customer does not demonstrate improvements to their energy management practices during the engagement period or by the next assessment, PSE should consider removing them from the program or adjusting the requirements for performance payments.

3) **Key Finding –** DNV did not identify any Facility Action Plans (FAPs), Portfolio Action Matrix (PAM), or similar documentation of customer actions that were taken or planned during the performance periods. The FAP or PAM is something the grant agreement states "should" be updated by the customer. The lack of documentation across the sampled customers is below expectations for a commercial SEM program. One key pillar of the SEM program is that customers implement changes during the performance period that result in reductions in consumption compared to the baseline (energy savings). Documentation of these changes assures stakeholders that the estimated energy savings can be attributed to the program. The documentation reviewed by DNV showed that PSE typically requested documentation of the actions taken after completing the savings analysis. Customers would document activities with an email or with an attachment to an email. In many cases this is all the documentation DNV identified. However, participants that utilize a third party to implement SEM at their facilities often had registers of the actions taken since this was a requirement under the contract between the participant and the third party. Similarly, PSE documentation did not contain the Quarterly Reports participants are required to submit per their grant agreements. DNV determined that PSE did not enforce this requirement during the evaluated period. PSE concluded Quarterly Reports were too burdensome on customers and enforcing the requirement would risk future participation. Participants interviewed by DNV confirmed that participants found little value in the Quarterly Reports.

    - **Recommendation** – PSE should review and adjust its program design to require the collection of FAPs or PAMs or similar at the start of the engagement period or at minimum by the end of Year 1. The requirement should be to document potential actions or the process through which actions will be identified and recorded. PSE should at least require program participants to submit their actions prior to the calculation of annual energy savings. The Quarterly Report is considered best practice to mitigate risk, but an Annual Report should be required at minimum. PSE should not require a report in a specific format, instead the report format should fit into the participants existing or planned reporting structure provided that they include key reporting metrics required by the program. PSE should not claim energy savings if the report is not provided by an agreed deadline. This will incentivize both PSE and its customers to identify actions once it is known that savings have occurred.

1.3.3 System for Measuring and Reporting Energy Performance

The sampled savings evaluated used either PSE created spreadsheets or My Data Manager (MDM) to estimate energy savings. DNV finds that the degree day energy models that PSE utilizes to estimate CSEM energy savings are reasonable. The models utilize independent non-weather variables to account for changes in occupancy. Non-routine events (NREs), such as other energy measures or changes to building loads are accounted for. Below are DNV's findings and opportunities for improvement specific to the estimation of energy savings.

4) **Key Finding** – The Commercial Strategic Energy Management (CSEM) program is achieving electricity savings and natural gas savings. The differences between PSE's estimates of savings and DNV's estimates are due to the baseline periods used to estimate savings and challenges that DNV experienced with the acquisition of utility meter data which limited the number of sites included in the final sample.
    - **Recommendation** – PSE should not adjust current program savings using these realization rates. The savings evaluated were the last calculated before the COVID-19 pandemic and are not applicable to the current program. PSE should review the recommendations below to improve the program evaluability and mitigate risks associated with savings attribution.

5) **Key Finding –** Baseline. DNV found opportunities for improvement in the defined baseline period as 23 of the 120 sites reviewed utilized baseline period from eight or more years prior to the performance period. The baseline period is used to develop the regression equation that estimates energy consumption in the absence of the program. DNV believes the baseline should be valid for the current loads in the building, meaningful to the participant, and changed if neither of first two are true. PSE's practice for the sites evaluated was to utilize a baseline period prior to the first year of a site's program participation, unless significant changes were identified and the customer agreed to a baseline change. In general, baselines in the evaluated sample were rarely updated. This had the benefit of preventing customer confusion and burden, but carried evaluability risks that grew over time. DNV requested monthly utility meter reads for all sampled sites from the start of their baseline period. In some cases, data from the baseline start were no longer stored by PSE. This prevented DNV from independently verifying the baseline period consumption using utility consumption data. DNV asked participants if the baseline period was meaningful to them and if the building use during the baseline period was similar to its use during evaluated performance period. While there were some cases in which older baseline periods were found to still represent evaluation period loads, most older baselines with either not meaningful or not representative of building use. In some cases, the respondent did not know whether the baseline was representative of building use as they were not at the organization during one or both periods. In many cases, participants viewed the 2018-2019 period as their baseline as it represents the prior normal consumption levels before changes due to the COVID-19 pandemic.
    - **Recommendation** – PSE should adjust its program design to ensure measurement baselines are meaningful to participating customers, representative of the existing conditioning loads in the building, and suitable for evaluation. DNV suggests updating site baseline models every six years at minimum. This is equivalent to two 3-year enrolment cycles and will ensure the utility meter data is still available within PSE's systems. PSE should consider

opportunities to more regularly update baseline to align with the participant organization's energy reduction goals and staff commitments.

6) **Key Finding –** DNV found that elapsed time between the close of the performance period and the date savings are claimed is too long for too many projects. In the sample of 35 projects evaluated, 50% of the projects had savings reported more than 6 months after the end of the performance period. Five projects contained sites with more than 1.5 years between the end of the performance period and the savings reported date. While time will always be required to complete the analysis, the program should avoid or greatly reduce significant delays.

   - **Recommendation** – PSE should review its program procedures and identify key factors that delay processing and reporting of energy savings. PSE should adjust its procedures to eliminate some, if not all, of the identified key factors. For example, DNV believes requiring an Annual Report of actions completed by an agreed deadline will reduce the time PSE waits for information on actions to support the estimated reductions in consumption.

7) **Key Finding –** DNV found that CSEM participants were insufficiently engaged in the measurement process. The measurement of energy savings in an SEM program is intended to improve the participants understanding of how and when energy is consumed. The measurement process provides each participant with information and feedback regarding their attempted improvements to energy management. Participants interviewed for this evaluation regularly stated that they did not participate in the calculation of savings, that they had not reviewed PSE's calculation spreadsheets, or that they did not trust the consumption values provided by PSE's MDM system as they rarely matched their actual billing records.

   - **Recommendation** – PSE should adjust its program design to increase participant engagement in the measurement process. The measurement process should not just be a step PSE does to calculate a participant's performance incentive. The process should be a platform through which a participant can learn about their consumption and continue to identify opportunities for improvement. PSE should consider adjustments to its calculation methods that better align the data used for the analysis with a participant's actual utility meter reads.

8) **Key Finding –** DNV found that PSE was unable to provide a sufficient amount of participant consumption data supporting the energy savings claimed. PSE's measurement process was completed within MDM or completed by PSE staff using spreadsheets. In both cases, the total consumption during the baseline and performance periods is listed. When spreadsheets were used, PSE included the total consumption by calendar month for each participant. PSE lists the account and meter numbers associated with every participating site in the grant agreement workbook. To verify the consumption and savings estimated, DNV requested monthly consumption data showing the meter read date for the sampled sites using the account and meter numbers listed.

   - **Recommendation** – PSE should review its program documentation and design to determine what changes are necessary to ensure that they can provide evaluators with the utility meter data for program participants. DNV believes the recommended change to re-baseline participants every 6 years will help mitigate the risks associated with changing account and meter numbers. One possible adjustment is to store the actual monthly utility meter consumption data for the performance period by account and meter in a separate file in the annual analysis folder instead of only having the aggregated calendarized consumption in the analysis file.

1.3.4 Other Key Findings

The findings below are based on DNV's participant interviews. These are findings are not aligned with any aspect of the minimum elements. These findings align with current initiatives in Washington state.

9) **Key Finding –** The majority of participants interviewed were aware of the Clean Buildings Performance Standards in Washington and the associated potential future penalties.

   - **Recommendation** – PSE should identify methods to support CSEM participants goals specific to the Clean Buildings Performance Standards. The goals of the program and the Clean Buildings Performance Standards are well aligned, and the existence of potential future penalties have motivated many participants to reengage with SEM. PSE could organize or support ASHRAE audit training courses for program participants as part of their annual training allowance. PSE could adjust its RMP, FAP, and PAM documentation to align with the Energy Management Plan (EMP) and Operations & Maintenance (O&M) Programs required to comply with the Clean Buildings Standard. PSE could ensure the data provided through this program can be easily integrated into the reported platforms used for Clean Buildings.

10) **Key Finding –** Many large institutional participants are moving away from energy savings goals and towards decarbonization goals. These participants are currently and expect to increase electric loads in order to reduce their use of carbon-based fuels. Calculating SEM energy savings for these customers while they are increasing their electric loads will be challenging.

   - **Recommendation** – PSE should remove participants or sites from the CSEM program and transition them to applicable clean energy programs when they plan to decarbonize their heating by switching to electric heating technologies. Participant sites could return to the program after the new system is in place for one year and a new baseline has been established. PSE should complete additional review of any participating customers transitioning off a central heating plant to decentralized electric heating. It is likely best to remove all impacted sites from the program until the new baseline has been established. Participants should also be encouraged to separately meter any vehicle charging loads added to their portfolio.

2 INTRODUCTION

This report summarizes the results of the impact and process evaluations of one Puget Sound Energy (PSE) 2020-2021 non-residential demand side management program. In this report program evaluator DNV presents results for the following program:

- Commercial Strategic Energy Management (CSEM)

This program offers performance-based and milestone-based incentives to commercial customers that commit to and achieve reductions in whole facility energy consumption compared to an agreed baseline (counterfactual). Table 2-1 shows the energy savings tracked[2] over the prior four years. These programs accounted for approximately 5% - 14%% of PSE's C&I electricity savings and 10% - 34% of C&I natural gas savings depending on the program year. DNV's impact evaluation sampled from savings reported 2019 and 2020 only.

Table 2-1. Tracked energy savings, 2020-2021

Progam Year	Unique Projects/ Participating Customers	Tracked Electricity Savings (kWh)	Tracked Natural Gas Savings (therms)
2019	32	15,349,891	596,042
2020	35	13,718,635	475,521
2021	28	6,595,843	355,433
2022	21	11,640,361	196,696

2.1 Evaluation objectives and researchable issues

The expected outcomes of this evaluation were to conduct the following research:

- **Impact evaluation**: Estimate the ratio of energy savings achieved to energy savings tracked for each program. This ratio is the program realization rate. These estimates were achieved by independently reviewing savings estimation methodologies and verifying savings achievement through file reviews and inspections.
- **Process evaluation**: Provide process findings for the programs from the perspective of the program participants. When necessary, provide information on why programs are over/underperforming and recommendations for improvements.

2.2 COVID-19 adaptations

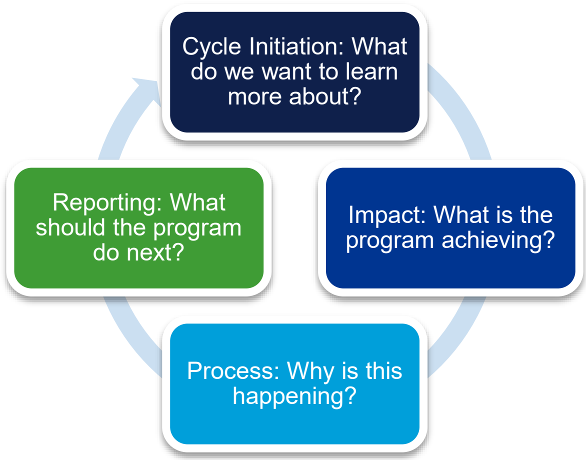
This evaluation was executed during a time of increased health and safety risk. DNV used remote methods for all data collection. Evaluated savings are based both on historic participant consumption and PSE documentation. Given the impact of COVID-19 on participant consumption and program operation, DNV does not recommend using the evaluated savings results to adjust current and future program savings. DNV does recommend reviewing and adjusting future program design and operations based on the findings and recommendations of this evaluation.

[2] Based on DSMc project tracking data provided to DNV by PSE for the purposes of program evaluation.

3 EVALUATION APPROACH AND METHODOLOGY

DNV utilized a dynamic forward-looking developmental evaluation approach. This evaluation approach provided PSE with annual program feedback structured to help improve savings reliability and program performance. DNV successfully completed two developmental cycles for this evaluation. Each cycle starts with an objective and concludes with program feedback and recommendations. Figure 3-1 shows the basic steps in each cycle. Each cycle was initiated by seeking to learn more about program savings performance.

Figure 3-1. DNV's developmental evaluation cycle



3.1 Sample design

Each impact evaluation step in the development cycle started with a review of program achievements and sample design. Each sample was designed to provide accurate independent estimates of energy savings achieved by the program and the associated program realization rates. DNV utilized a two-stage stratified random sampling approach with certainty selection to identify the sample for this impact evaluation. The sample was selected in two phases, a first phase selected in 2020 and a second phase selected in 2021. The preliminary sample design was based on 2019 program achievements with the goal of achieving 10% relative precision on site energy savings (kBtu) at the 90% confidence interval for the program. The sample targets were adjusted during the second phase of sample selection in 2021 based on actual 2020 program performance and findings from the first phase. All evaluation results present electric and gas realization rates separately. DNV intentionally oversampled in both phases for this evaluation due to an expectation that recruitment would face additional challenges during the COVID-19 pandemic.

Table 3-1 summarizes the final impact evaluation sample design implemented and the associated expected relative precision of the results. All relative precisions are shown at the 90% confidence interval for site energy savings (kBtu) which combines electricity and gas savings into one single value with consistent unit of measure for size stratification. The error ratios used in the sample design were based on DNV's experience evaluating similar programs. The full sample design is discussed in Appendix A: Sample design. The design and final achieved sample for each program is discussed in the program-specific sections of the report.

Table 3-1. Sample summary

Compliance Program	PSE Program	Phase I	Phase II	Total Sample[3]	Designed kBtu RP @ 90%
Commercial SEM	Unique Projects (Participants)	19	15	28	10%
	Unique Sites	70	50	115	

3.2 Data collection

The evaluation utilized multiple data sources to evaluate the program. All site- or project-specific data collection was completed remotely via telephone or virtual meeting interviews. DNV and PSE agreed that the additional health and safety risks associated with travel and in-person interactions due to the COVID-19 pandemic made in-person site visits inadvisable. Table 3-2 shows the data sources used to evaluate each PSE program.

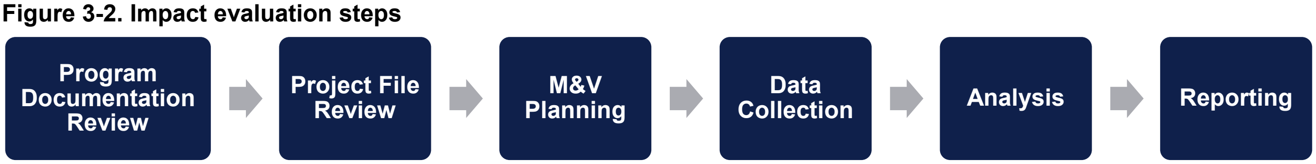
Table 3-2. Evaluation data sources

	Program Materials	Sampled Project Documentation	Project-Specific Calculations	Utility Consumption Data	EMS/BMS Trend Data	Participant Interviews	Trade Ally Interviews
Commercial SEM							
Commercial SEM	✓	✓	✓	✓		✓	

3.3 Impact evaluation methods

Program impact evaluation was initiated after the primary samples were identified. The impact evaluation steps used for this project are illustrated in Figure 3-2.

Figure 3-2. Impact evaluation steps



The steps in this process were primarily applied at the program level and are discussed in more detail in the program-specific sections. A brief description of each step is provided below:

- **Program Documentation Review:** Review program application forms, program guides, measure savings documentation, and program plans to understand the program design and theory, measures supported by the program, and the assumptions and methods used to estimate energy savings.
- **Project File Review**: A thorough review of the project files for sampled projects, focused on the energy savings calculations, assumptions, and other supporting documentation. The review identified any missing information critical to the evaluation, original calculation methodology, key uncertainty parameters to research, and any concerns with the original savings estimation methods.
  - A critical result of this step was the determination of how much sites (buildings) were included in each of the tracked energy savings sampled for evaluation.

[3] The total shows the total unique projects in the final sample. Some projects and sites were selected in both phases. DNV evaluated the savings reported in each phase.

- **M&V Planning**: Upon the completion of program document review and project file review, DNV created a program wide data collection and analysis plan. This plan documented the data to be collected through the evaluation process and the anticipated analysis method. DNV planned to complete this evaluation by:
  - Sample up to four buildings with energy savings estimates associated with the tracked energy savings claims sampled. The full sample design is discussed in Appendix A: Sample design.
  - Requesting and receiving monthly utility meter data based on account numbers and meter numbers shown in the project documentation for the sampled sites. This data would be used to verify the annual consumption recorded for each sampled site.
  - Interviewing sampled participants to verify program participation, assess program engagement, review actions completed, review changes to facility use, and collect program feedback.
  - Utilize received monthly utility meter data to independently estimate the change in energy consumption during the sampled performance period at each sampled site.
  - Estimate site specific evaluated savings for the performance period based on our independent estimate of the change in energy consumption, information collected during interviews, and tracked energy savings from reported energy conservations measures (ECMs) supported and tracked by other PSE programs.
- **Data Collection**: Data collection occurred through participant phone interviews based on an interview guide.
- **Analysis**: The evaluated savings analysis followed the M&V plan with specific modifications discussed in detail below. PSE was unable to provide monthly utility meter data for the entire sample and was unable to provide data from all baseline periods. DNV adjusted its analysis plan to maximize the number of site-specific savings estimates produced. DNV estimated site specific savings for each performance period based on the difference between actual consumption and a modeled baseline. The baseline regression model was based on consumption during the year prior to the performance period. For each sampled project, DNV produced estimates of evaluated electric and/or gas savings. A full description of DNV's baseline model development is available in Appendix B: Model selection & development.
- **Reporting.** Analysis results were recorded in a program specific spreadsheet comparing the program consumption and savings to the evaluation consumption and savings. This document is the final summary evaluation report for this program.

3.3.1 Sample extrapolation to track and program

DNV used a separate ratio estimator to obtain unbiased estimates of the total evaluated savings (either kWh or therms) for any group of interest. This estimator will yield, by design, unbiased estimates of some outcome measure, and is particularly beneficial when the outcome measure is correlated with something known for all members of the sample frame. In this case, the evaluated savings are logically correlated with tracked savings as listed in the tracking database. In general, the separate ratio estimator works as follows.

Suppose the indices:

g = Application domains which are defined by track and fuel type (kWh or therms). For some outcome measures and domains of interest, strata had to be collapsed with one another during the estimation process. This occurred with $Y_g \neq 0$ but $\sum_{i \in Sample} w_{ig} y_{ig} = 0$ (these terms are defined below).

i = Site.

And suppose:

x_{ig} = Evaluated savings for site i in group g.

y_{ig} = Tracked savings for site i in group g.

w_{ig} = Sample weight for site i in group g. This reflects the sample selection process that was used at the beginning of the study to select the original 202 sample points.

Y_g = Population total tracked savings in group g. So $Y_g = \sum_{i \in Frame} y_{ig}$

$$\hat{R}_g = \frac{\sum_{i \in Sample} w_{ig} x_{ig}}{\sum_{i \in Sample} w_{ig} y_{ig}}$$ is the Ratio estimate for group g.

Then the separate ratio estimator that will yield the total evaluated savings is:

$$\hat{T} = \sum_g \left(Y_g \cdot \hat{R}_g \right)$$

And the ratio estimate of total modeled savings to total tracked savings is:

$$\hat{R} = \frac{\hat{T}}{\sum_g Y_g}$$

The procedure used for calculating ratio estimation by domains provides the correct standard error of the estimate for each domain and overall. The procedure also takes into account defined clusters of observations (customers) and stratification.

The standard error is calculated as drawn from a finite population: the measures completed within the analysis period with associated energy impacts in the program-tracking database. This calculation uses the Finite Population Correction (FPC) factor. This factor is a reduction to the calculated variance that accounts for the fact that a relatively large fraction of the population of interest has been observed directly and is not subject to uncertainty. It is appropriate to apply precision statistics, such as confidence intervals, based on the standard error calculated in this manner when quantifying the results of the program during the study period only. The FPC factor reduces the calculated sampling error around the estimate more for smaller populations than for large.

3.4 Process evaluation methods

We conducted a process evaluation for the purpose of identifying program successes and opportunities for program improvement. DNV's process evaluation relied on interviews with program staff and program participants only. DNV's process evaluation activities focused on identifying opportunities to improve savings reliability, expand program participation, increase the savings achieved through PSE's program portfolio, and improve the PSE customer experience throughout the program. Details on the process evaluation methods and findings are presented a specific section.

DNV's process evaluation approach generates feedback that enables adaptive management of PSE's programs. The over-arching process evaluation goal is to provide the contextual information necessary to understand how programs are performing, why certain results are occurring, what is working well, and what opportunities for improvement exist. Our evaluation provides PSE with feedback focused on understanding what happened and identifying opportunities to adjust program delivery and achieve program goals.

Our team used a variety of techniques to systematically assess program processes and provided actionable recommendations that address opportunities to improve customer and stakeholder satisfaction and determine the appropriateness of program activities given current market conditions. Table 3-3 summarizes these and some other core process evaluation methods, their value to the evaluation, and the topics we address.

For each program, we present within the respective chapter the process methods applied given the availability of data and participant contacts.

Table 3-3. Process evaluation methods overview

Method	Topics	Value to the Evaluation
In-depth telephone interviews with PSE's program staff (may include program managers, Energy Advisors, outreach staff, and/or implementation contractors)	Changes to program since the last evaluation cycle; marketing/outreach activities; operations; stakeholder interaction	Ensures understanding of how specific members of PSE's team plan to use the evaluation results (helps ensure we provide results in formats that maximize their usefulness to PSE). Provides PSE staff with opportunities to contribute to evaluation's content and share perspectives on program performance. Additional basis for data collection instruments.
Telephone interviews of PSE's customers for customer satisfaction	Customer satisfaction and experience. Some programs may combine verification and satisfaction surveys.	Report salient findings on an ongoing basis to allow PSE to enhance the participant experience or adjust program design to better serve customer needs. Some programs may survey shortly following project completion to increase the probability of getting useful feedback.

4 COMMERCIAL STRATEGIC ENERGY MANAGEMENT

This section summarizes the impact and process evaluation results of PSE's 2020-21 Commercial Strategic Energy Management program (CSEM). A program overview is presented first. Sample design approach is discussed next followed by the impact and process evaluation approach. Finally, the impact and process evaluation results are presented along with findings and recommendations on how to improve the tracked savings for the CNC program and projects.

4.1 Program overview

The CSEM exists under Schedules E253 and G253. The program is offered to customers with large electric or gas sites or portfolios of sites. The program provides incentives for defined milestones and achievement of measured energy savings. Energy savings are measured at the building meter level by comparing actual consumption during a "Performance Period" to modeled baseline consumption during the same period. The baseline regression model is trained on consumption prior to program participation or the "Baseline Period". The independent variables for the regression are expected to be major energy drivers. PSE typically uses weather variables and occupancy.

Any measure or action taken that could reduce energy consumption is included in the program. However, PSE adjusts for non-routine events and other facility changes that impact consumption that are not specific to energy management. Participants typically receive incentives for traditional energy efficiency measures through other PSE programs. PSE therefore removes the energy savings tracked through other programs when calculating the CSEM program savings occurring at each participating building.

4.1.1 Program savings

Table 4-1 shows the energy savings tracked by the program from 2019-2022. Over this time the program tracking includes 56 unique customer numbers and 74 unique project numbers associated with energy savings claimed ("Performance" measures). The customer number and project number are typically the same across all measures for the customer. There are multiple project numbers for cases when the specific buildings in a portfolio are tracked separately or PSE determined the change was necessary at the time of contract renewal. DNV's impact evaluation sampled from energy savings reported in 2019 and 2020 only. All four years are shown here to demonstrate the impact the COVID-19 pandemic has on tracked program savings.

Table 4-1. Tracked energy savings, CSEM 2019-2022

Savings Reported Year	Unique[4] Customer Numbers	Unique Projects/ Transactions	Tracked Electricity Savings (kWh)	Tracked Natural Gas Savings (therms)
2019	29	32	15,349,891	596,042
2020	33	35	13,718,635	456,258
2021	26	28	6,595,843	355,433
2022	20	21	11,640,361	196,696
Grand Total	**56**	**74**	**47,304,730**	**1,604,429**

4 Customers often participate in multiple years. Therefore, the total unique counts are less than the sum of the year specific counts.

4.2 Impact evaluation

This section documents DNV's independent estimate of the program realization rate and review of the calculation methods used by the program. Each element of our evaluation process is discussed below along with relevant findings. The section concludes with our estimate of the program realization rates followed by the primary drivers of variance between PSE's tracked savings estimates and DNV's evaluated savings estimates.

DNV completed the following steps for the impact evaluation of the CSEM program:

- Sample selection: Selection of a representative sample of completed projects for evaluation
- Project file review: Review of project files provided by PSE to identify calculation methods and key parameters, and to ensure sufficient information exists to evaluate the project
- Project-specific M&V planning: Creation of a program-specific measurement and verification plan
- Data collection: Phone interviews with sampled participants to review each project, baseline assumptions, and current operating parameters
- Project-specific analysis: Estimated evaluated savings using the data received and collected to develop an independent estimate of savings.

4.2.1 Sample design

This sub-section presents an overview and summary of the sample design used to evaluate the 2020-2021 CSEM program. DNV used stratified random sampling to select an efficient representative sample of projects for evaluation. The sample was designed to provide a reliable estimate of program performance while also selecting a variety of measures to increase the breadth of the DNV's review. The sample was selected from CSEM projects with savings reported dates between July 1, 2019, and April 30, 2021. Initial sampling occurred at the project number level. Table 4-2 summarizes the planned sample design for this program. Key design elements were:

- Creation of domains based on the primary fuel saved, electricity or gas. This helped ensure sufficient results for both fuels.
- Stratification by size of savings reported and use of a certainty stratum to increase the magnitude of savings evaluated and the accuracy of the estimated savings realization rates.

DNV adjusted the sample design after reviewing Phase 1 project documentation and understanding the number of sites included within each tracked savings claim. DNV revised the Phase 2 sample down to 15 savings claims. The final primary sample included 34 unique claims associated with 28 unique participants (six participants were sampled in Phase 1 and Phase 2). One sampled claim for Phase 1 was dropped when it was determined that no savings were associated with the participant. DNV did not replace this sample point as the remaining points were expected to be sufficient for this evaluation.

Table 4-2. CSEM sample design

Savings Reported Year	Sampling Frame Population	Planned Sample 2020-2021 Evaluation	Final Evaluation Sample
Phase 1 (known at sample design)	32	20	19
Phase 2 (expected at sample design)	32	20	15
Total	**64**	**40**	**34**

4.2.2 Project file review

DNV reviewed each sampled project file for program savings methodology and accurate savings reporting. This review included the following steps:

1) Identified location of site meter numbers and/or account numbers within documentation. Extracting all meter and account numbers identified.

    a) Information was determined to be available and located within grant or renewal agreement documentation.

2) Determining if the energy savings reported in the database is supported by a single or multiple regression models. Identifying how many sites were estimated to have positive savings during the performance period and confirm that the sum of the site-specific estimates of savings equals the tracked energy savings.

    a) The energy savings tracked by PSE are the sum of savings from multiple sites for each participating customer. Each site's energy savings is typically estimated using one regression model. Table 4-3 shows the number of sites contributing to the savings claims sampled.

Table 4-3. Number of sites contributing to sampled savings claims

Sampling Phase	Final Sample, Participants	Total sites listed in savings calculators	Sites contributing to sampled savings
Phase 1	19	544	247
Phase 2	15	391	203
Total	**34**	**935**	**450**

3) Identifying the baseline period used by PSE to develop the baseline regression model for each site sampled.

    a) The baseline period is shown on the grant agreement documentation. The year the baseline starts is also shown on the site-specific calculation. Table 4-4 shows the year the baseline period started for the sites selected in the Stage 2 sample.

Table 4-4. Sampled sites by baseline year start

Year Baseline Starts	Count of Sites, Files Reviewed	Count of Sites, Final Site Sample
2010	150	25
2011	39	6
2012	43	6
2013	228	40
2014	28	2
2015	128	8
2016	181	24
2017	39	5
2018	52	4
Total	**888**	**120**

4) Identifying if actual performance period for the tracked energy savings under evaluation.

    a) Performance period dates located within site savings calculator workbooks. DNV found that the elapsed time between the end of the performance period and the date savings were reported was longer than expected for many sites. DNV did not anticipate that the sample would include performance periods that ended in 2016 or 2017. Table 4-5 shows the year the savings performance period ended for the sites in the Stage 2 sample.

Table 4-5. Sampled sites by the year the performance period ended.

Year, End of Performance Period	Count of Sites, Files Reviewed	Count of Sites, Final Site Sample
2016	13	4
2017	103	3
2018	176	24
2019	429	61
2020	214	28
Total	**935**	**120**

5) Identifying how many years the customer and site has participated in SEM and which renewal year the savings under evaluation represent.

   a) Customer participation length estimated based on start date of earliest baseline period. Another indicator is the number of years between the baseline start and the performance period start for each site. If a site has been re-baselined then this would underestimate participation history. Table 4-6 shows the number of years between the baseline period and the savings performance periods. The more years between, the more likely that the baseline regression model is no longer valid.

Table 4-6. Years between baseline and performance periods

Years	Count of Sites, Files Reviewed	Count of Sites, Final Site Sample
9	61	11
8	95	12
7	37	7
6	89	14
5	118	21
4	89	12
3	76	8
2	109	19
1	218	16
Total	**892**	**120**

6) Identifying location and existence of any information specific to the actions taken or reasons for changes in site energy consumption.

   a) This information was recorded in two places. First, communications between the participant and PSE were located within the project documentation. Second, PSE provided notes in the site savings calculator workbooks. However, in multiple cases, no notes were recorded for the site sampled by DNV.

7) Identifying any use of any non-weather independent variables, polynomial regressions, or use of more than two change points in the regression model(s).

   a) DNV did not identify the use of polynomial regressions or three-change point models in the workbooks. DNV was unable to review the models developed using My Data Manager (MDM). PSE stated that MDM regularly used an occupancy variable for sites. DNV did identify regular use of a “summer” indicator variable when modelling school energy consumption.

8) Determining if any baseline adjustments occur in the model.

   a) DNV did not identify any non-routine adjustments to the estimated baseline consumption.

9) Identifying if and how non-routine adjustments for capital projects measures (ECMs) are included in the model. Extracting the associated savings values and installation dates for all recorded ECMs.

   a) PSE does adjust for non-routine events within the site savings workbooks. The primary adjustments are for ECMs tracked by other PSE programs. PSE uses the installation data of the ECM to determine if all or part of the ECM savings should be removed. If only part of the ECM savings should be removed, then PSE uses savings load shapes to allocate annual savings to calendar months. DNV finds this to be an elegant approach to ECM NRAs. DNV believes the use of the load shapes is not necessary, but it does more accurately assign savings to months. PSE also adjusts sites for on-site solar generation not included in the baseline model.

   b) PSE often makes adjustments for portable classrooms that are moved around within participant school districts. PSE adjusts building consumption based on the square footage of portable classrooms installed.

4.2.3 Stage 2 Site Sample

During the previous file review task DNV identified which buildings supported each energy savings claim sampled. DNV sampled from these buildings to select the buildings that would be included in the impact evaluation and discussed with participants during interviews. DNV limited the number of sites per participant per phase to four buildings. Each participant's portfolio was stratified based on savings. Any site contributing more than 25% of the participants total savings was selected with certainty. Other sites were selected randomly within their stratum.

Table 4-7. Stage 2 sample result

Sampling Phase	Final Sample, Participants	Total sites listed in calculators	Sites contributing to savings	Stage 2, Sampled Sites
Phase 1	19	544	247	70
Phase 2	15	391	203	50
Total	**34**	**935**	**450**	**120**

4.2.4 Data collection, Participant Interviews

DNV successfully completed interviews with 23 of the sampled 28 participants. These interviews were completed in 2022. Given the elapsed time between the sampled performance periods and the interviews, DNV asked participants to specify if their responses were specific to the performance period under evaluation or their recent experience with the program. If a respondent had no direct knowledge of the facility during the performance period, DNV collected information on the recent engagement with the program. All interviews occurred via telephone using an interview guide. Through the in-depth interview, DNV staff captured information to:

- Verify engaged participation in the program during the performance period and in 2022.
- Verify the actions taken during the performance period to reduce energy consumption and assess if and how the participant tracks these actions.
- Identify any significant known capital improvements or non-SEM activities that impacted energy consumption during the performance period or year prior.
- Identify any operating conditions or changes to the facility that may have affected the energy savings or the validity of PSE's energy savings calculations.
- Identify known seasonal changes in facility use that might prevent modeling using weather only

- Understand basic occupancy, cooling, heating, process schedules and associated control sequences that should be reflected in consumption data, such as typical start and stop to heating and cooling seasons and use of free cooling.

4.2.5 Data collection, Utility meter data

DNV requested monthly utility meter data for all sites sampled. The initial request was provided to PSE in 2021 after Phase 2 site sampling was completed. DNV provided PSE with the customer names, addresses, listed meter numbers, and listed account numbers for each site. The meter numbers and account numbers were identified from the project documentation created at the start of each customer's engagement in the program. DNV requested data from the start of the sites baseline period through the most recent monthly meter read. PSE was unable to provide DNV will all of the data requested. Consumption data identification was attempted by PSE twice as part of this evaluation after DNV identified gaps in the first data set received. DNV intentionally chose to not use the consumption data included in the site-specific spreadsheets as this data was incomplete and processed to estimate calendar month consumption.

The difference between the data requested and the data received significantly impacted the results and applicability of this evaluation. PSE should work to remedy this data gap immediately. Potential reasons for the data gap are:

- Changes to customer and meter numbers. DNV believes the site meter and account numbers changed over the site's engagement in the program. 30 of the 120 sites in the sample had baseline start dates in 2010 or 2011. These sites had therefore participated in one version of the SEM program for close to ten years. It is highly likely that the account and meter numbers changed over this period. DNV does not know if the information provided in the agreement documentation coincided to the original or most recent numbers.
- Loss of baseline data from PSE's system. The data from a portion of the baseline period was no longer stored by PSE for the 24 sites with baseline periods that started in 2010.

PSE should consider the following adjustments to its program to ensure utility meter data is available to support the energy savings claimed by the program. PSE should expect future evaluators to request participant consumption data

- PSE should review its processes and ensure the meter and account numbers are reviewed at the time of each renewal grant agreement.
- PSE should determine if an opportunity exists to improve premise level consumption tracking even if meter numbers and account numbers change. If such an opportunity exists, PSE should consider adjustments to the program.
- PSE should implement a program change that ensures in the preservation of the baseline consumption data.

4.2.6 Project analysis

DNV adjusted its analysis plan based on the utility meter data received. Instead of developing a regression using the same baseline period as PSE, DNV chose to develop a regression from consumption during the year immediately preceding the performance period. This deviation from the program calculation method produces different estimates of savings. The methodology has both drawbacks and benefits. DNV believes the benefits outweigh the drawbacks. Key benefits that drove this decision are:

- Maximize use of consumption data provided by PSE.
- Minimize additional NRAs that much be taken for ECMs and solar production.
- Eliminate additional calculation steps required to estimate incremental savings over the engagement period.

- Test the validity of the incremental savings calculation by estimating year over year savings only using an alternative method.

Drawbacks to this methodology include:

- Mid-year changes during the baseline period that result from program participation produce less well fit models and increase modeling uncertainty.
- Mid-year ECM adjustments in the baseline are included in the baseline model and increase uncertainty. This proved to not be a factor based on the sites DNV produced models for and the timing of their ECMs.
- The baseline models for multiple schools will include the portable classrooms in place during the prior year.

DNV created 128 electric models and 97 gas models from the data provided. These models were associated with 43 electric sites and 45 gas sites. Details on DNV's baseline model development are described in Appendix B: Model selection & development.

Table 4-8. Count of evaluation models developed compared to sites sampled

Sampled Fuel	Tracked Energy Claims Sampled	Total Sites Sampled	Final Sites Evaluated	Unique Evaluation Models
Electric Savings	26	76	43	128
Gas Savings	23	64	45	97

4.2.7 Program realization rates

The site-specific results for the final evaluated sample were extrapolated back to the sampling frame to estimate the evaluated savings for the sampling frame and the program realization rate. Given the differences between the final and expected samples, there is more uncertainty in the result that planned. DNV also calculated the unweighted realization as the ratio of the sum evaluated savings over tracked savings for the final sample. DNV does not recommend adjusting future program savings based on these results due to the associated uncertainty and recent program changes. Table 4-4 provides the electric evaluation results and Table 4-5 provides the gas evaluation results.

Table 4-9. CSEM electric impact evaluation results

Evaluated Projects w/ kWh Savings	Tracked MWh Savings in Final Sample	Evaluated MWh Savings in Final Sample	Unweighted Realization Rate, kWh	Weighted Realization Rate, kWh
43 Site Specific Models	5,002	4,184	84%	144%

Table 4-10. CSEM gas impact evaluation results

Evaluated Projects w/ therm Savings	Tracked therm Savings in Final Sample	Evaluated therm Savings in Final Sample	Unweighted Realization Rate, therms	Weighted Realization Rate, therms
45 Site Specific Models	308,487	151,646	49%	75%

DNV makes the following conclusions from these analysis results.

- PSE's program achieved incremental savings during the performance periods evaluated, but the magnitude of savings is uncertain.

- The realization rates are within expectation given the challenges acquiring utility meter data and the change in methodology used.
- Site-specific results indicate that gas savings may be particularly sensitive to modeling choices as DNV estimated no gas savings at 18 of the 45 gas site evaluated. DNV has previously found small gas savings claims to be sensitive to the degree day reference temperature selected. PSE should consider further review of its gas modeling methodology to ensure savings are estimated even if the model is adjusted.
- PSE should consider if the effort to estimate and document SEM savings at sites with low consumption levels supports the efficiency potential they offer. In the files reviewed by DNV, 14% of sites with positive electric savings consumed 50,000 kWh or less during the performance period. Similarly, 26% of sites of positive natural gas savings consumed 5,000 therms or less. While it is important to offer programs to all customers, there is a limit to the savings potential these sites can achieve.
- Future evaluations should plan to sample fewer sites and develop site-specific evaluation plans based on the documentation provided for each site.

4.3 Process evaluation

DNV interviewed 23 of the sampled 28 participants in 2022 for this process evaluation. The interviews followed a guide designed around CEE's Minimum Elements and the site's sampled for impact evaluation. Process evaluation questions focused on current participant engagement in PSE's program and feedback on the program. Figure 4-1 shows participant responses to two questions asking about the frequency of engagement in SEM elements. Program improvement opportunities exist to.

- Increase the frequency of opportunity and action register updates and progress reviews. DNV that participants engaged in a SEM should review opportunities, actions, and progress at regular intervals (quarterly at minimum). This aligns with PSE's to receive quarterly updates from participants. Participants that never or annually review their opportunities, actions, and progress are less likely to be achieving savings directly attributable to PSE's program.

Figure 4-1. Participant engagement frequency

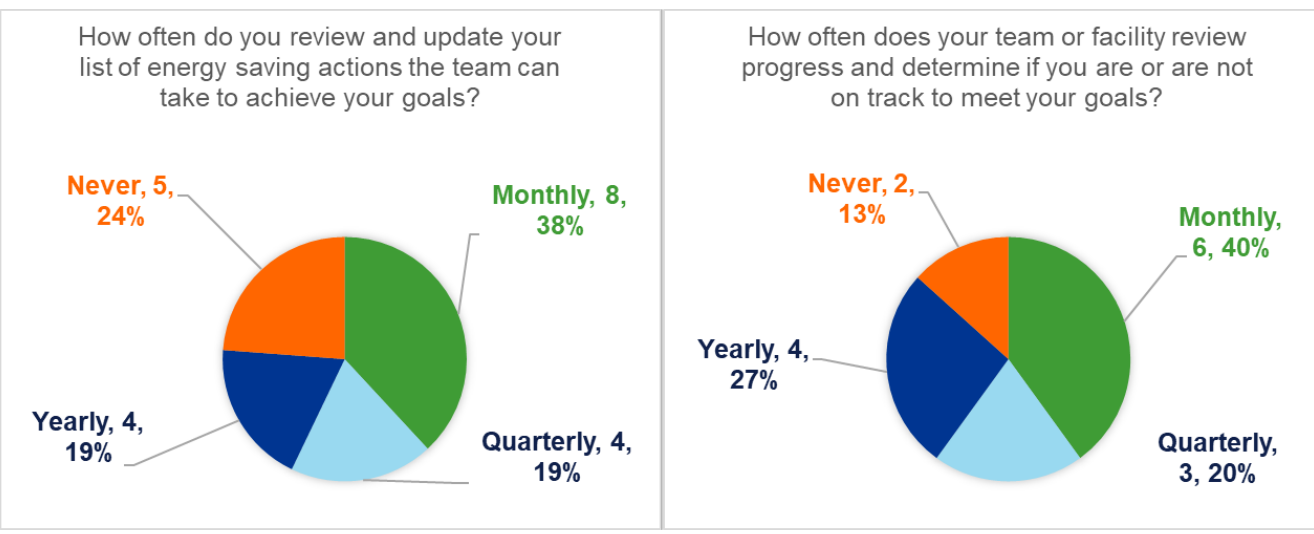
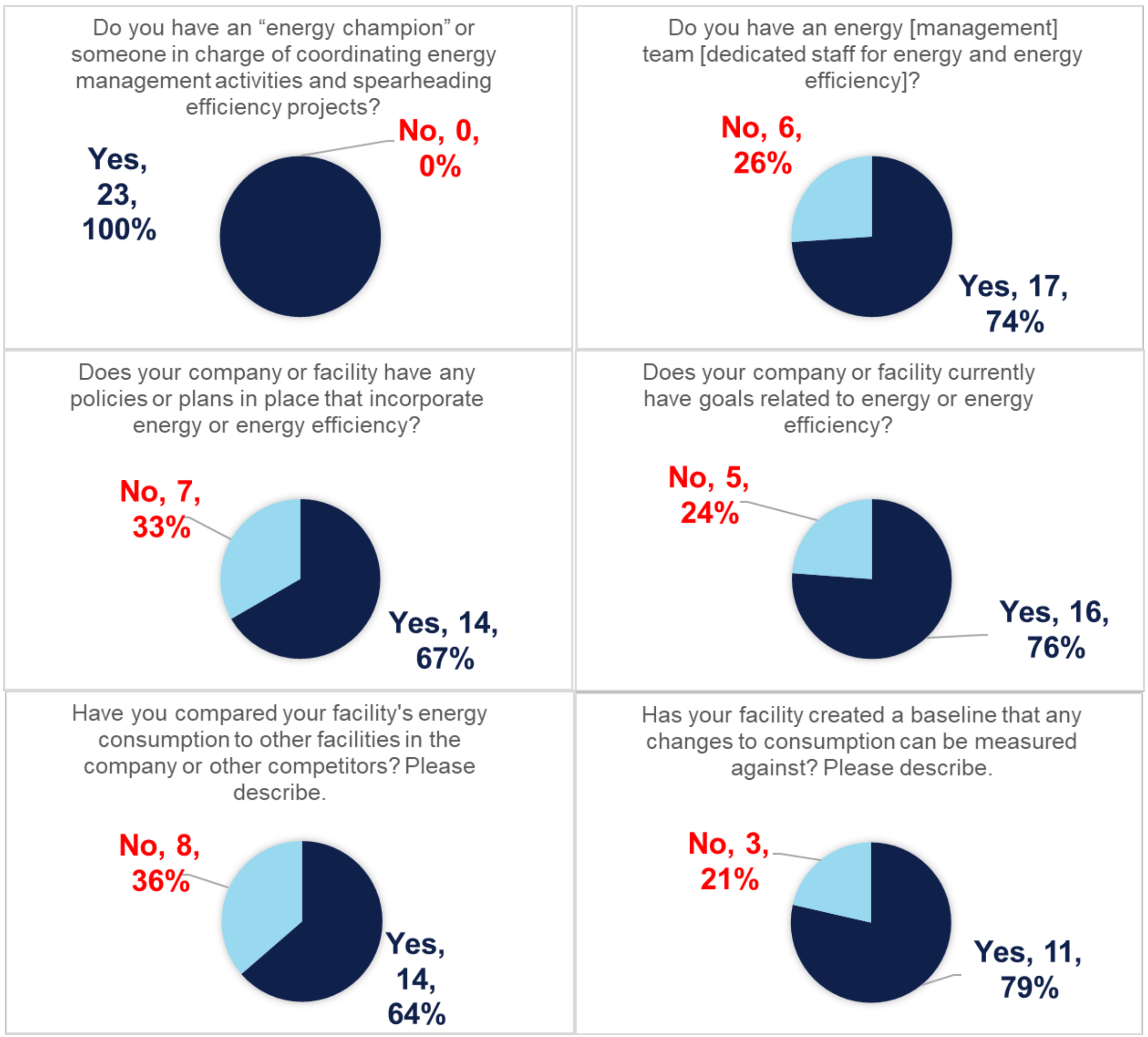


Figure 4-2 shows the responses to DNV's engagement questions. Only responses that DNV confidently classified as Yes or No are shown. DNV attempted to ask all questions during all interviews but did not ask all questions based on the responses received to earlier questions, the number of facilities sampled, and the time available for the interview. These responses demonstrate that participant engagement is lower than expected for an SEM program. Program improvement opportunities exist to:

- Increase the percent of participants with energy teams. Most program participants manage large portfolios of buildings enrolled in the program. An energy team or dedicated resources are necessary to achieve energy savings across such a large portfolio. DNV believes at least 90% of SEM participants should be expected to articulate the existence of an energy team or resources assigned to achieve energy savings reductions.
- Increase the percent of participants that can articulate their organizations goals specific to energy savings and energy efficiency. DNV believes at least 90% of SEM participants should be expected to articulate the existence of energy savings goals.

Figure 4-2. Participant engagement responses



4.3.1 Program experience and future goals

DNV's interviews included open questions on previous program experience and future goals. The questions are shown below with summarized responses. DNV reviewed responses for patterns and developed the responses below based on the patterns observed. Responses received four or more times are highlighted.

1. Thinking back on your history with PSE and this program, how have your organization's energy management practices changed?

Financial incentives have enabled strategic energy management to occur, such as funding a resource conservation manager. (4 aligned responses)

Increased energy awareness within organization and legitimacy to internal energy conversations. (4 aligned responses)

Acceptance within organization that active management of setpoints and schedules is required to achieve energy efficiency. (1 aligned response)

Limited to no change in organization's energy management practice. (4 aligned responses)

2. What are you hoping to do with the program in the future?

Increase energy savings. (1 aligned response)

More energy audits. (2 aligned responses)

Ensure compliance with Clean Buildings Performance Standard. (4 aligned responses)

Increase visibility with management. (1 aligned response)

Improve monitoring of consumption/savings. (2 aligned responses)

Take advantage of training opportunities. (1 aligned response)

Participant will focus more on capital improvements and less on SEM in near future. (1 aligned response)

Unknown/Unsure what future participation will be. Example: The organization's goals do not align with the SEM program. (1 aligned response)

3. Describe the most valuable aspects of your participation in the program?

Cash incentives. (5 aligned responses)

Funding our resource conservation manager. (1 aligned response)

Knowledge sharing. (6 aligned responses)

Legitimate measurement & verification. (1 aligned response)

Tools for consumption data review. (2 aligned responses)

Partnership in energy management. (1 aligned response)

4.3.2 Additional program feedback

DNV concluded participant interviews with an open discussion on the program. Below are the key takeaways from these discussions.

- Overall program satisfaction is high. Participants appreciate the financial and structural support PSE has provided through this program over the years.
- Most participants stated that in person trainings held prior to COVID-19 were often informative and valuable. Participants often stated that it was harder to attend and prioritize the online trainings in recent years since COVID-19. All participants understand why recent trainings were held online. Many are hopeful that increased in-person training and site audits will lead to improved efficiency in their portfolios.
- All participants are aware of the Clean Buildings Performance Standard. Many are hopeful that their SEM participation will help them comply with or at least identify buildings out of compliance with the standard. DNV believes there is an

opportunity for PSE to adjust its program documentation to align with the requirements of the Clean Buildings Performance Standard. This support would be valuable to customers and SEM participants would appreciate not having two different reporting requirements.

- Multiple institutional participants have or will be transitioning from energy savings goals to decarbonization goals. These participants stated that they are actively planning to increase their electric consumption as they switch from natural gas to electric heat pump heating and install electric vehicle charging infrastructure. Customer sites actively engaged in decarbonization efforts should not be enrolled in the CSEM program due to the challenges associated with estimating savings using utility meter data.
- Multiple participants mentioned challenges they had using the MDM system and comparing the consumption data they received to the consumption listed on their utility bills. These participants also stated they are signed up for PSE's new EnergyCap system and they are hopeful that the new system will not have the same challenges as the previous one. PSE should ensure the data and results available to participants through EnergyCap aligns with the customers utility bills or the tool risks not being valuable to participants.
- Few participants stated use of or review of PSE's SEM energy savings calculations. Most participants communicated that the calculations were done by PSE and sometimes, but not always, reviewed with the participant. DNV was surprised by the limited participant engagement in the measurement element of SEM. PSE should review its program design and identify changes that are expected improve engagement in the measurement of savings over time.

DNV

5 APPENDICES

5.1 Appendix A: Sample design

This appendix discusses our approach to sample development for the selected compliance program. First, we summarize the program participants for the chosen compliance programs at the time of sampling, then discuss our sample design approach for the programs and lay out the preliminary sample design for the evaluation (selection for Phase I). The structure of the Phase II sample is the same. However, DNV did update the number of projects selected based on Phase I findings and updated program tracking data.

5.1.1 Historic participant data

Program tracking data with customer contact information and program enrolment were provided by PSE. Table 8-1 presents the electric (kWh), gas (therm), and combined site savings (kBtu) for each compliance and PSE program. For the new construction program measures were further separated by end-use.

Table 5-1. Preliminary sample frame, 2019 annual program savings

Compliance Program	PSE Program	Accounts	kWh Total Savings	Therm Total Savings	kBtu Savings (kWh+therms)
Commercial SEM	Commercial SEM	32	15,349,891	596,042	111,978,778

5.1.2 Sample design approach

Our initial sample design approach followed the principals of model-based statistical sampling (“MBSS”) to construct the sample design and provide the framework for the subsequent analysis. MBSS techniques have been used to create a very efficient and flexible structure for collecting data on countless energy efficiency evaluations, demand response evaluations, and interval load data analyses, e.g., load research and end-use metering, projects.

The key to this project was to develop statistically reliable data that could be dynamically analyzed. In 2021, DNV compared 2020 program achievements to this initial sample design and made adjustments where necessary to ensure the final sample selected represented program achievements over the biennium. The following sections fully describe the sample design and analysis approach that we used in this project.

5.1.3 Background

Conventional methods are documented in standard texts such as Cochran’s Sampling Techniques.[5] MBSS is grounded in theory of model-assisted survey sampling developed by C.E. Sarndal and others.[6,7] MBSS methodology has been applied in load research for more than 30 years and in energy efficiency evaluation for more than 20 years. This fusion of theory and practice has led to important advances in both model-based theory and interval load data collection practice, including the use of the error ratio for preliminary sample design, the model-based methodology for efficient stratified ratio estimation, and effective methods for domains estimation.

MBSS and conventional methodologies are currently taught in the Association of Edison Illuminating Companies’ Advanced Methods in Load Research seminar. MBSS methodology is also documented in The California Evaluation Framework.[8]

5 Sampling Techniques, by W. G. Cochran, 3rd. Ed., Wiley, 1977.
6 Model Assisted Survey Sampling, by Carl Erik Sarndal, Bengt Swensson and Jan Wretman, Springer-Verlag, 1992.
7 Wright, R. L. (1983), “Finite population sampling with multivariate auxiliary information,” Journal of the American Statistical Association, 78, 879-884.
8 The report can be downloaded from the webaccount http://www.calmac.org/calmac-filings.asp

MBSS has been used successfully for decades in countless load research and program evaluation studies. It has also been examined in public utility hearings and in at least two Electric Power Research Institute, Inc. (EPRI) studies.

5.1.4 The role of the statistical model

MBSS uses a statistical model to guide the planning and the sample design. The parameters of the model, especially the error ratio, are used to represent prior information about the population to be sampled. The model describes the nature of the variation in the relationship between any target y variable of the study, in our case the normalized daily consumption of the customer, and one or more x variables that can be developed from known billing data and other supporting information. The x variable is usually a measure of the size of the customer, e.g., annual use, and assumes good information is available in the billing to support the analysis. The model is used to help choose the sample size n, to assess the expected statistical precision of any sample design, and to help formulate a sample design that is efficiently stratified for ratio estimation using case weights.

The model is used as a guide to the sample design, but the results of the study itself are not strongly dependent on the accuracy of the model.[9] Once the sample design is selected, the subsequent analysis of the data is based only on the sample design and not on the model used to develop the sample design. The resulting estimates will be essentially unbiased in repeated sampling and the confidence intervals will also be valid, provided that the sample design has been followed to select the sample customers. The results will be consistent with traditional sampling theory as found in texts such as Cochran's Sampling Techniques and consistent with standard load and market research practice.

5.1.5 Stratified ratio estimation

We assumed that the data collected and analyzed in the study is for a given population of N accounts in a given customer class. In this study, annual energy savings were the unit of measure. We let y denote any customer characteristic to be determined and we let x denote any suitable characteristic of the customer that is known from tracking data such as measure, quantity, project, or customer. We define the population ratio B by the equation

$$B = \frac{\sum_{i=1}^{N} y_i}{\sum_{i=1}^{N} x_i}.$$

Here the summations are over the entire N units (e.g., customers) in the target population. We note that the population mean or total of y is equal to B times the population mean or total of x. The latter is assumed to be known from the billing or tracking data.

We assumed that a sample of n customers is selected following a stratified sample design. For each sample customer we define the case weight w to be equal to the number of customers in the target population within the stratum containing the given customer divided by the number of customers in the sample within the given stratum. The case weight is used to avoid any bias that might otherwise arise from the different sampling fractions used from one stratum to another.

Using the case weight, we define the combined ratio estimator of B by the equation:[10]

9 Other methods, called model-dependent sampling, are much more dependent on the accuracy of the model. Such methods are not commonly used in load research applications since they would be more difficult to defend than MBSS and conventional methods.

10 This equation gives the same result as the conventional stratum-weighted equation: $b = \frac{\sum_{h=1}^{L} N_h \bar{y}_h}{\sum_{h=1}^{L} N_h \bar{x}_h}$.

$$b = \frac{\sum_{i=1}^{n} w_i \, y_i}{\sum_{i=1}^{n} w_i \, x_i}$$

Then, if desired, the population mean or total of y can be estimated as b times the population mean or total of x, known from the tracking data. Using the case weights, we calculate the relative precision at the 90% level of confidence in three steps:

1. Calculate the sample residual $e_i = y_i - b\,x_i$ for each unit in the sample.
2. Calculate $se(b) = \frac{\sqrt{\sum_{i=1}^{n} w_i (w_i - 1)\, e_i^2}}{\sum_{i=1}^{n} w_i x_i}$. [11]
3. Calculate $rp = \frac{1.645\ se(b)}{b}$.

A 90% confidence interval for B is calculated using the equation $b \pm rp\ b$. A confidence interval for the mean or total can be calculated in a similar way.

We can also use the sample data to estimate a measure of population variability called the error ratio, denoted er. The error ratio is the key determinant of the expected relative precision, along with the sample size n. We estimate the error ratio from the sample using the following equation:

$$\hat{e}r = \frac{\sqrt{\left(\sum_{i=1}^{n} w_i \, e_i^2 / x_i^{\gamma}\right)\left(\sum_{i=1}^{n} w_i \, x_i^{\gamma}\right)}}{\sum_{i=1}^{n} w_i \, y_i}$$

The parameter γ (gamma) is defined in the next section. In practice, it is usually taken to be 0.8. We will not attempt to interpret the preceding equation here, but we will define both the error ratio and gamma in the following section.

5.1.6 The ratio model

The ratio model is used to choose the appropriate sample size n, to assess the expected statistical precision of any stratified sample design, and to develop an efficiently stratified sample design. The ratio model describes the relationship between y and x for the set of all units in the population. The model consists of two equations called the primary and secondary equations respectively: [12]

$$y_i = \beta \, x_i + \varepsilon_i$$

$$\sigma_i = sd(\varepsilon_i) = \sigma_0 \, x_i^{\gamma}$$

---

[11] The conventional equation is $se(b) = \frac{1}{\sum_{h=1}^{L} N_h \bar{x}_h} \sqrt{\sum_{h=1}^{L} N_h^2 \left(1 - \frac{n_h}{N_h}\right) \frac{s_h^2(e)}{n_h}}$ where $s_h^2(e) = \frac{1}{n_h - 1} \sum_{i=1}^{n_h} (e_i - \bar{e})^2$. Our equation assumes that $\frac{1}{n_h - 1} \sum_{i=1}^{n_h} (e_i - \bar{e})^2$ is approximately equal to $\frac{1}{n_h} \sum_{i=1}^{n_h} (e_i)^2$ in each stratum.

[12] The x-variable in the primary equation is sometimes different than the x-variable in the secondary equation. In the SAS modules, we refer to the later as the stratification variable. For simplicity, we will not make this distinction in the theoretical discussion given here.

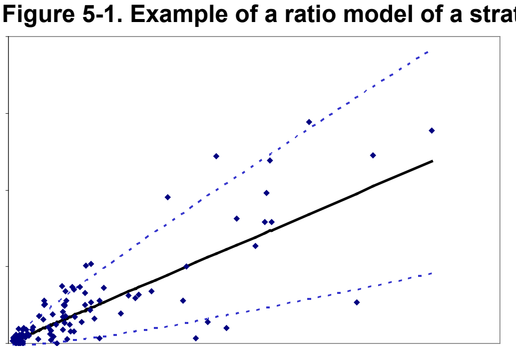
Here i denotes any customer, account, or HVAC units in the target population. $x_i > 0$ is usually known throughout the population. The primary equation describes the relationship between the y variable of interest, e.g., normalized daily use, and the x variable used in the ratio estimate, i.e., actual daily use. Since we assume that $E(\varepsilon_i) = 0$, the primary equation can also be written as $\mu_i = E(y_i) = \beta x_i$. Here μ_i denotes the expected value of y for unit i. The primary equation says that under the model, the expected value of y_i is equal to a fixed constant β times the known x_i.

The quantity, $\varepsilon_i = y_i - \mu_i$, is called the residual. The N residuals are considered to be N independent random variables. The standard deviation of ε_i is denoted as σ_i. We refer to σ_i as the residual standard deviation of each customer i. The secondary equation is used to estimate the residual standard deviation and to guide the development of an efficient sample design.

To summarize, under the ratio model, the target variable y_i is a random variable with expected value μ_i and standard deviation σ_i. The expected value μ_i is determined by the primary equation of the model. The standard deviation σ_i is determined by the secondary equation of the model. There are three parameters in the model: β (beta), σ_0 (sigma-naught), and γ (gamma).

Figure 8-1 shows an example. The points of the scatterplot represent the values eaof (x, y) for each site in the population. The solid line represents the equation $y = \beta x$, i.e., the expected value of y given x. This is a line through the origin with slope given by the parameter β. The two dashed lines represent the equation $y = \beta x \pm \sigma$, i.e., the one-standard deviation interval around the expected value. Here $\sigma = \sigma_0 x^\gamma$ so the dashed lines are determined by the two parameters σ_0 and γ.[13]

Figure 5-1. Example of a ratio model of a stratified sample



Now we are finally positioned to define the error ratio. The error ratio is defined by the equation:

[13] The role of gamma can be seen by rewriting this equation as $\log(\sigma) = \alpha + \gamma \log(x)$ where $\alpha = \log(\sigma_0)$. This shows that for each site in the population the log of sigma is a constant plus gamma times the log of the value of x for the site. Gamma is the slope in the relationship between the log of x and the log of sigma.

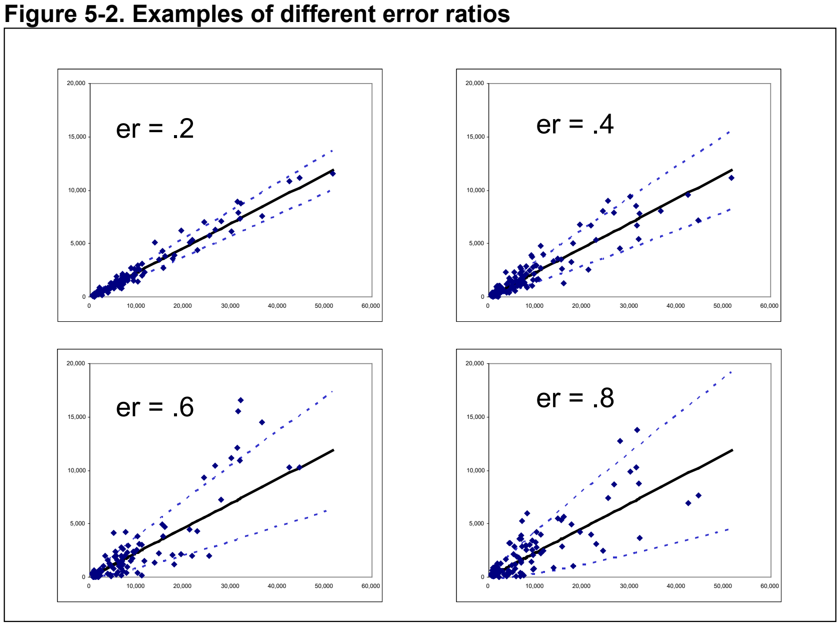
$$er = \frac{\sum_{i=1}^{N} \sigma_i}{\sum_{i=1}^{N} \mu_i}$$

The error ratio can be regarded as an alternative parameter to σ_0 since under the preceding ratio model, σ_0 can be calculated from the error ratio using the equation

$$\sigma_0 = er \frac{\sum_{i=1}^{N} \mu_i}{\sum_{i=1}^{N} x_i^{\gamma}}$$

The error ratio is the key measure of variability when stratified ratio estimation is to be used to analyze the data. Figure 8-2 shows some examples. If the error ratio is close to zero, there is a strong relationship between x and y. If the error ratio is larger, the relationship is weaker.

Figure 5-2. Examples of different error ratios



5.1.7 Choosing the sample size

We assumed that the ratio model provides a reasonably accurate description of the relationship between y and x in the target population. We also assumed that the sample design will be efficiently stratified as discussed previously and that the analysis will use stratified ratio estimation.

Under these assumptions and the added assumption that the population size N is large, then the expected relative precision is given by the equation: $rp = z\frac{er}{\sqrt{n}}$. Where z is the standard normal deviate or 1.645 for 90% confidence and 1.96 for 95% confidence, "er" is the error ratio and "rp" is the required relative precision. If the population is relatively small, the finite population correction factor can be added, giving

$$rp = z\sqrt{1-\frac{n}{N}}\frac{er}{\sqrt{n}}.$$

In Cochran, the relative precision "rp" is referred to as the desired relative precision "D." If D, is specified, then the preceding equations can be solved to determine the required n. If the population size N is large, we have

$$n = \left(\frac{z\ er}{D}\right)^2.$$

Please note, the error ratio (er) and the z-value has a modest impact on the sample size whereas the desired relative precision has a significant impact. For example, halving the desired relative precision from ±10% to ±5% effectively quadruples the sample size.

If the population is small, the sample size can be calculated in two steps.

First, calculate $n_0 = \left(\frac{1.645\ er}{D}\right)^2$. Then calculate $n = \frac{n_0}{1+n_0/N}$. These equations and some reasonable assumptions are generally enough to develop a preliminary plan.

5.1.8 Stage 1 sampling design

For this project, we examined the required sample size to achieve 90% level of confidence at ±10% precision for each compliance program across the two-year study. For planning purposes, a preliminary sample design estimate of precision for the combined 2020 and 2021 program years was developed by projecting double the 2019 measures and program total tracking savings (kBtu) to estimate 2020-2021 program enrollment.

Table 8-2 presents the sample size requirements based on the program tracking savings data available for program participants. For this exercise, we elected to choose conservative error ratios for each program based on prior experience in evaluating similar energy efficiency programs. For lighting programs, we chose an error ratio of 0.6, for commercial kitchens and small business direct install an error ratio of 0.8, and for all other programs an error ratio 1.0. Of course, the true error ratio for each end use will not be known until the data is collected and analyzed. After the initial sample is analyzed the error ratios will be reviewed, any future updates to the sample design will incorporate updated error ratios.

Table 5-2. Preliminary sample design

Compliance Program	PSE Program	Accounts (2xPY2019)	Total kBtu	Error Ratio	Planned Sample	Expected Relative Precision
Commercial SEM	Commercial SEM	64	223,957,555	0.80	40	10%

5.1.8.1 Stratification

The preceding results assume that the sample is efficiently stratified. Under the ratio model, an efficiently stratified sample design for ratio estimation can be developed in the following steps:[14]

Use the sampling frame and the assumed model to calculate σ_i for each customer in the population.

[14] This methodology is the model-based version of the Dalenius-Hodges method of constructing strata combined with optimal allocation of the sample using the within-strata population standard deviation of the e_i. However, Dalenius-Hodges stratification is approximately optimal for stratified mean per unit estimation whereas model-based stratification is approximately optimal for stratified ratio estimation. Moreover, with conventional methods it is common to calculate the required sample size from the within-stratum population standard deviation of x_i. This practice can yield very misleading results and cannot be recommended.

Choose the desired number of strata,[15]

Sort the sampling frame by increasing σ_i.

Choose stratum cut points to divide the sum of the σ_i approximately equally between the strata.

Allocate an equal number of sample customers to each stratum.

Make added adjustments if the sample size exceeds the population size in any stratum.

Under the ratio model, σ_i is determined by the x variable together with the value of γ. Methods are available for estimating γ from a sample. Indeed, we have estimated γ in numerous load research studies. We have found that the estimated values are clustered around 0.8. We have also found that the key results are not very sensitive to γ. Therefore, in interval load data collection applications, we generally recommend the use of $\gamma = 0.8$ both in constructing strata as discussed in this section and in estimating the value of the error ratio from a given sample.

Samples were stratified based on PSE program total tracking savings (kBtu). The tables that follow show the number of accounts in the population and sample, total savings, and inclusion probability for each end-use or stratum by compliance program and PSE program.

Table 5-3. Sample design stratification for CSEM

Compliance Program	Stratum	Maximum	Accounts	Total kBtu	Sample	Inclusion Probability
Commercial SEM	1	1,006,568	12	5,878,000	3	0.21
	2	1,442,366	5	6,171,767	3	0.60
	3	2,576,427	3	6,374,140	3	1.00
	4	3,993,405	2	7,476,279	2	1.00
	5	4,288,165	3	12,484,751	2	0.67
	6	16,614,511	7	73,593,841	7	1.00
	Total		**32**	**111,978,778**	**20**	**0.59**

5.1.9 Stage 2 sample design

Each of the accounts selected in stage 1 represented a PSE customer and their portfolio of the buildings enrolled in the CSEM program. After file review, one sample point was dropped as it was not associated with an actual savings claim. The 19 CSEM portfolios included in the sample included 216 unique sites with positive energy savings in the performance period. DNV's stage 2 sample selected 70 unique sites for evaluation from the 216. The goal was to select sites using a similar stratification method to the first stage. This sampling was done in a spreadsheet. The following rules were used.

1. Up to four sites per selected savings claim could selected to minimize customer burden during interviews.
2. Only sites with positive savings, gas or electric, for the claim sampled were eligible.
3. Any site with kBtu savings greater than 25% of the total customer savings was selected.

[15] With MBSS methodology we can systematically assess the gain from increased stratification. These studies indicate that five annual-use strata are usually sufficient in most load research applications. Some applications may call for added stratification by seasonal use, customer load factor, etc.

4. All other sites were allocated into strata of equal total savings with up to 4 additional strata per customer. Each of these sites was assigned a random number. DNV then reviewed the strata by customer and the number of sites within each stratum. A sample size for each stratum was assigned.

5.1.10 Evaluating the precision of any design

For any sample design, we define the inclusion probability of each site in the population, denoted π_i, to be the probability that the site is included in the sample. For a stratified sample design, the inclusion probability is the sampling fraction in each stratum, i.e., n_h / N_h.

Under the ratio model and any sample design, the expected relative precision of the stratified ratio estimator is

$$rp = z\sqrt{\sum_{i=1}^{N}\left(\pi_i^{-1} - 1\right)\sigma_i^2} \bigg/ \sum_{i=1}^{N} \mu_i$$

Here $z = 1.645$ for the 90% level of confidence.

This key result has the following mathematical implications:

For any given sample size n, a sample design is said to be efficient if the sample design minimizes the expected relative precision. For any efficient sample design, $\pi_i = \frac{n}{\sum_{i=1}^{N}\sigma_i}\sigma_i$ provided that the right-hand side is less than 1.

If the right-hand side is greater than 1, the site should be included with certainty.

If the sample design is efficient and the population is large, then the expected relative precision is $rp = \frac{z\ er}{\sqrt{n}}$.

The model-based sample design is practically efficient as long as the number of strata is large enough.

The preceding equation can also be used to calculate the expected statistical precision of any sample design in any domain of interest.

5.2 Appendix B: Model selection & development

This appendix presents the methods used in this evaluation to develop evaluated SEM savings. The impact evaluation savings analysis relied on statistical energy consumption modeling using available historic energy consumption and weather data only to influence consumption at a sampled site. DNV used the monthly energy consumption data by read date provided by PSE through a data request.

Based on data availability and program documentation, DNV utilized its “year-before baseline” methodology to develop savings estimates for comparison with the claimed program achievements. This methodology utilizes the energy consumption in the year immediately prior to the performance period to develop the baseline regression model which was generally a different period than that used by PSE to estimate savings. This change in methodology is expected to provide different but informative estimates of savings. The primary reasons for this change in methodology were:

- To simplify the calculation by eliminating the need to consider years of non-routine adjustments due to capital energy conservation measure (ECM) installations and other facility changes. Recorded capital ECMs installed in the analyzed baseline period performance periods are considered in the evaluation analysis.
- To simplify the calculation by eliminating the need to estimate incremental savings and focus only on the change observed over the performance period.
- To maximize the number of estimates DNV could create based on the data available to the evaluation.
- To provide a comparison between the two different methodologies that can be referenced when considering program design changes.

Modeling background

Modeling criteria

DNV considers statistical criteria and the appropriateness of the model when developing models for use in evaluation. In general, the strength of a model follows from its ability to tell a concise, consistent, and compelling story.

- Concise models explain the appropriate amount of variation in the dependent variable under conditions experienced most frequently. There can be a large amount of variation in factors outside of weather that drive energy consumption. The intent of the energy consumption model is to best explain energy consumption as a function of weather and other predictor variables when those values are in the most common regions of their respective ranges.
- Consistent models have coefficient values with logical relationships. For example, a model should typically yield higher estimates of energy consumption as weather conditions become extreme or building occupancy or activity levels increase.
- Compelling models have a strong statistical fit. The probability that the coefficients are different than zero should generally be greater than 90%. Further, the overall model should account for a large amount of the observed variation in energy consumption. The adjusted R-squared statistic captures how much variation in the dependent variable (energy consumption) the model explains. Values greater than 0.8 denote a very strong statistical fit. Models that have an adjusted R-squared less than 0.5 are unable to explain half the variation in energy consumption.

To assess whether the models are consistent and concise, DNV assessed the available data on the drivers of energy consumption at SEM sites. Often, we did not have sufficient visibility into the energy drivers to assess if the models were well defined. For example, hospitals likely have factors other than weather that drive energy consumption. However, we did consider if the models made sense overall, adapting appropriately to the known variables:

- Was energy consumption predicted to change appropriately in response to the weather conditions?
- Were the predicted savings reasonable for the actions and measures implemented?

Modeling vs. Fitting

One significant risk in statistical modeling is the trap of "over-fitting" to the available data when developing regression models. Curve-fitting tries to find an equation that fits well with the present data, while modeling tries to find an equation that represents the underlying data generator. Curve-fitting can be misleading and can lead to over-fitting in the sense that the fitted curve may not accurately represent periods of time outside of what was used to create the curve; the classic example is always being able to fit an (n-1)th-degree polynomial to n data points. For these regression models, the energy consumption should be directly correlated with what actually drives usage. The DNV models are independent of any curve-fitting.

For this evaluation, DNV used adjusted R-squared values to assess the statistical fit. Adjusted R-squared is reduced when the model includes too many predictor variables. Increasing the number of variables may lead to a high R-squared value, but also can lead to interpretation issues, especially when the predictor variable is seemingly unrelated to energy consumption. The evaluation therefore limited the independent variables to weather-based variables and one non-weather variable.

Site Baseline Modeling Approach

DNV utilized a standardized regression modeling approach for gas and electric usage to estimate annual energy consumption for each sampled site (or associated meter if multiple meters serve one site). DNV utilized HDD and/or CDD to capture the underlying physical heating and cooling processes. DNV did not attempt to use any non-weather independent variables in the evaluation due to inconsistent data availability. This standardized modeling approach serves to independently verify the claimed program savings. DNV developed the best model for each site based on the standard modeling criteria. In order to find the best model for each site, DNV tested several different models using various reference temperatures:

- Heating only - uses HDD term only. This model was used for all gas models.
- Cooling only – uses CDD term only.
- Single reference temperature – uses HDD and CDD calculated using the same reference temperature.
- Dual reference temperatures – uses HDD and CDD, where unique reference temperatures are calculated separately for cooling and heating.

Model selection & development

DNV developed 225 models using site-specific data from the baseline period. DNV used meter reads from the 12-months prior to the savings performance period as the baseline period. Model development for each site occurred in two stages:

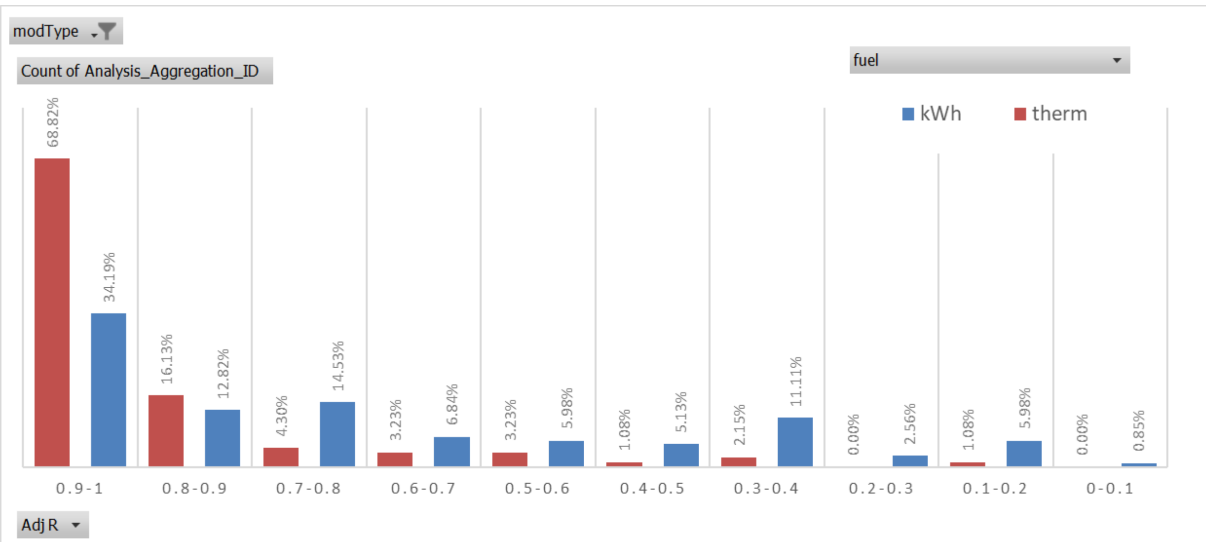
Stage 1, Determination of optimal model type reference temperatures: The first stage determines the optimal reference temperature for each potential site model type. The temperature value that produced the highest adjusted R-squared value for a type was chosen to represent that type.

Stage 2, Model type selection: The best site model type of the four types listed above was the model type with the highest adjusted R-squared value. Table 5-4 shows the model types used for the evaluation models developed. Figure 5-3 shows a distribution of the Adjusted R^2 values for the selected models (excludes constant models). This figure shows that most, but not all, model fits were above 0.70 (0.75 is often an industry standard cut off). These models are still believed to be the best models for the data provided and the distribution meets DNV's expectations for an evaluation of commercial buildings.

Table 5-4. Selected evaluation model types

Fuel	Temperature Response Model Type	Model Count
Electric	Constant	11
Electric	CDD Only	19
Electric	CDD & HDD, Single Reference Temperature	28
Electric	CDD & HDD, Dual Reference Temperature	15
Electric	HDD Only	55
Electric	**Subtotal**	**128**
Gas	Constant	4
Gas	HDD Only	93
All	**Total**	**97**

Figure 5-3. Adjusted R^2 by fuel type, selected evaluation models



Monthly Residuals

Energy savings for each month during the program are estimated as the difference between the modeled baseline energy and the actual energy consumption. This is referred to as the "monthly residual". This value is an estimate of the energy use avoided during the month due to all changes at the site. If the project installed a capital project after the baseline period, then any savings due to the capital project are included in the monthly residual.

Program Year Savings

This section discusses how incremental program year savings are determined from monthly residuals.

5.2.1.1 Performance Period Savings

Total performance period energy savings are based on the sum of monthly residuals during the performance period. DNV used the same performance periods stated by PSE in their savings calculation workbooks. Since the baseline is developed form the prior years' consumption, the sum of the residuals during the performance period is the incremental savings for that period. All performance period savings were assigned a program year. The program year is the calendar year the final read of the performance period occurred in. Any residuals after the evaluated performance period were assigned to subsequent program years. DNV's program year assignments do not directly align with the year savings were reported by PSE. These assignments are used help review consumption and residuals over time.

5.2.1.2 Capital project adjustments

Individual capital measures (ECMs) supported by other PSE programs and associated with a sampled facility and fuel combination were identified using "PSE ECM Summary" tabs in the provided site savings calculator workbooks. This tab lists all the ECMs supported by PSE for this customer since the customer joined the program. DNV did not independently very this list against program tracking data. DNV only identified two sites, each with one ECM, that required an non-routine adjustment. All others occurred outside the performance period or at sites that DNV did not receive consumption data for. The identified ECMs were accounted for by adjusting the performance period savings by the ECM savings per day times the number of days of adjustment required. This was simpler than PSE's use of monthly load shapes to allocate savings, but reasonable for this application.

5.2.1.3 Program year SEM savings

Capital measure saving values are subtracted from the program year summation of monthly model residual savings values to arrive at the total SEM program savings achieved by program year and fuel type. Following the program's guidelines, incremental savings are calculated as any SEM program savings that are greater than the SEM program savings claimed in previous years of program participation.

5.2.1.4 Savings calculation summary

The following is a summary of the steps taken to estimate evaluated program year SEM savings:

1. Monthly Meter Residuals: DNV calculated meter-level monthly energy savings as the difference between the estimated baseline consumption (using the regression model) and actual meter consumption. All calculations used monthly utility meter reads and daily weather data aggregated to each utility meter read period.
2. Program Year Assignment: DNV assigned each monthly residual to a program year based on the year the final meter read in the performance period occurred. This is done to ensure twelve months in the evaluated performance period are assigned to the same year.
3. Total Program Year Savings: DNV summed the calculated monthly residuals across all meters associated with each site over the defined performance period (first assigned program year).
4. Program Year Capital Project Savings: DNV calculated program year capital savings based on the tracked ECM savings and the number of days in the assigned program year that the measure was installed.
5. Total Program Year SEM Savings: DNV calculated the total SEM savings achieved in a program year as the difference between the Total Program Year Savings and the Program Year Capital Project Savings.
6. Incremental Program Year SEM Savings: DNV calculated Incremental Program Year SEM Savings as the Total Program Year SEM Savings since the baseline is developed from the prior year's consumption.

About DNV

DNV is a global quality assurance and risk management company. Driven by our purpose of safeguarding life, property and the environment, we enable our customers to advance the safety and sustainability of their business. We provide classification, technical assurance, software and independent expert advisory services to the maritime, oil & gas, power and renewables industries. We also provide certification, supply chain and data management services to customers across a wide range of industries. Operating in more than 100 countries, our experts are dedicated to helping customers make the world safer, smarter and greener.

Evaluation Report Response

Program: Commercial Strategic Energy Management Program

Program Manager: Max Rennie

Study Report Name: Evaluation of 2020-2021 Commercial Strategic Energy Management Program

Draft Report Date: February 21, 2023

Evaluation Analyst: Kasey Curtis

Date of Final Report Provided to Program Manager: February 21, 2023

Date of Program Manager Response: February 27, 2023

Overview

The Commercial Strategic Energy Management Program (CSEM) program offers performance-based and milestone-based incentives to commercial customers that commit to and achieve reductions in whole facility energy consumption compared to an agreed baseline. PSE offers CSEM services to any Commercial customer, school district, and public-sector government agency with a minimum portfolio baseload to meet cost-effective thresholds. The CSEM program targets larger customers with multiple facilities such that the cost of implementation can be recovered through savings achieved. Schedule 448, 449, 458, and 459 customers may utilize their Schedule 258 funding allocation for CSEM Services. While the CSEM program is specific to commercial customers, PSE has offered Strategic Energy Management options for Multi-Family customers and Industrial customers through other program offerings.

Customers qualify for the CSEM program based on their annual PSE energy purchases. A typical customer baseline for maximum program funding is 20,000,000 kWh for electric only or 2,700,000 therms for gas-only service from PSE. Funding levels are prorated based on the amount of staff a customer would need to allocate in order to achieve cost-effective savings from CSEM efforts. At a minimum, the customer needs to use 1,000,000 kWh or 135,000 therms, or the equivalent to participate in the program.

A CSEM customer employs, contracts, or designates existing staff to implement CSEM responsibilities, including accounting for resource consumption, assessing facilities, recommending actions, monitoring progress, calculating savings and communicating program information to organization stakeholders.

Monetary grants include a "start-up" grant for completion of deliverables associated with building the program foundation. The start-up deliverables include identifying an Energy Manager, setting up an energy-accounting database, writing a company resource management plan, and completing facility action plans. Once start-up deliverables are complete, the customer may qualify for "performance grants" based on achieving energy savings associated with CSEM practices and “target grants” for meeting or exceeding pre-established energy-reduction targets.

The CSEM agreement is valid for three years. Over this time, PSE anticipates a 10-12 percent reduction in overall energy use. Savings are calculated using industry standard practices and energy accounting methodologies. Reported annual savings are a variance from a fixed baseline. PSE may elect to renew a customer's CSEM agreement in three-year increments to provide continued support and additional performance incentives.

Evaluation

DNV initiated the evaluation for the CSEM program in 2020. Due to the impact of the COVID-19 pandemic on the program and DNV’s initial findings, DNV and PSE agreed to extend the evaluation through 2022. This extension allowed DNV to observe changes to the program and further evaluate the energy savings claimed prior to any pandemic disruptions. The impact evaluation attempted an independent estimate of the ratio of energy savings being realized by the program to the energy savings tracked by PSE, referred to as the program realization rate. Impact evaluation methods were based on the program design, measures offered, and historic program performance. The process evaluation relied on interviews with sampled program participants.

Key Findings, Recommendations, and PSE Response

This section provides key findings and recommendations resulting from DNV’s evaluation, along with PSE’s response to the recommendations. For the sake of organization, each finding, its associated recommendation, and PSE response are numbered below.

Finding 1: Customer Commitment

Key Finding – While some CSEM participants articulated their organization’s energy goals and could describe the energy team in place to achieve them, many participants stated that no energy savings goals were currently in place and that turnover in staff (or other reasons) had reduced the energy team to a single person. While the SEM program does set a savings reduction goal with the participant, this was often not tied or integrated into the customer’s goals. In addition, few participants articulated that a senior or executive manager at the organization had visibility into the SEM efforts and only a few had reporting requirements outside of annual energy spending. DNV finds that the customer commitment to SEM was lower than is expected for an SEM program.

Recommendation – PSE should consider adjustments to its program design to increase SEM visibility within a customer’s executive management structure, ensure that an appropriately sized energy management team is maintained during the engagement, and the agreed goals for the engagement are included in the organization’s goals for energy savings. If these are not in place and customers do not demonstrate commitment, PSE should consider withholding performance payments and not attribute to savings to the program.

PSE Response

> *We agree with this recommendation. We noticed that in recent years, having dedicated energy teams was de-prioritized by many customers as they struggled to adapt to*

shutdowns and changes resulting from COVID. Energy managers at customer locations experienced a high degree of turnover. During that period we sometimes relaxed our requirements in this area to ensure we kept customers engaged, rather than separating them from the program. But with the pandemic behind us, we are incorporating new training that emphasizes the importance of having an energy team which includes senior org. leadership, and it will once again be a requirement before a customer receives a startup incentive.

Finding 2: Project Documentation

Key Finding – DNV did not identify any Resource Management Plans (RMP) or similar documentation stating the current energy management practices of the organization at the start of the engagement period and the goals or changes to the practice that the program would help facilitate or encourage. Therefore, there is no information for the evaluator that records the counterfactual to program participation specific to energy management practices. This reduces DNV's confidence that the reductions in consumption estimated can be attributed to the program. This is often referred as an Energy Management Assessment (EMA). In the CSEM grant agreement, the RMP is something the participant "should" update according to the agreement and is similar to a facility's standard operating procedure.

Recommendation – PSE should require the completion of the RMP or an EMA at the start of any engagement or renewal period. This assessment could be completed by the customer alone, facilitated by PSE, or facilitated by a third party. The purpose of the assessment is to understand current practice and identify improvements to energy management that the program can support. By supporting these improvements, PSE can attribute measured savings during each performance period to its program. If a customer does not demonstrate improvements to their energy management practices during the engagement period or by the next assessment, PSE should consider removing them from the program or adjusting the requirements for performance payments.

PSE Response

We agree with this recommendation as well. This is another existing program requirement which, due to the difficulty recruiting and retaining program participation during the COVID pandemic, we were not enforcing as stringently as we might otherwise. Now that the worst of the pandemic is over and the CSEM program is close to being fully staffed, we will be enforcing these requirements. We are replacing the RMP with an Energy Management Plan (EMP) which will be required at the start of each engagement, and is also aligned with the Washington State Clean Buildings Law to make the customer journey easier.

Finding 3: Customer Reporting

Key Finding – DNV did not identify any Facility Action Plans (FAPs), Portfolio Action Matrix (PAM), or similar documentation of customer actions that were taken or planned during the performance periods. The FAP or PAM is something the grant agreement states "should" be updated by the customer. The lack of documentation across the sampled customers is below expectations for a commercial SEM program. One key pillar of the SEM program is that customers implement changes during the performance period that result in reductions in consumption compared to the baseline (energy savings).

Documentation of these changes assures stakeholders that the estimated energy savings can be attributed to the program. The documentation reviewed by DNV showed that PSE typically requested documentation of the actions taken after completing the savings analysis. Customers would document activities with an email or with an attachment to an email. In many cases this is all the documentation DNV identified. However, participants that utilize a third party to implement SEM at their facilities often had registers of the actions taken since this was a requirement under the contract between the participant and the third party. Similarly, PSE documentation did not contain the Quarterly Reports participants are required to submit per their grant agreements. DNV determined that PSE did not enforce this requirement during the evaluated period. PSE concluded Quarterly Reports were too burdensome on customers and enforcing the requirement would risk future participation. Participants interviewed by DNV confirmed that participants found little value in the Quarterly Reports.

Recommendation – PSE should review and adjust its program design to require the collection of FAPs or PAMs or similar at the start of the engagement period or at minimum by the end of Year 1. The requirement should be to document potential actions or the process through which actions will be identified and recorded. PSE should at least require program participants to submit their actions prior to the calculation of annual energy savings. The Quarterly Report is considered best practice to mitigate risk, but an Annual Report should be required at minimum. PSE should not require a report in a specific format, instead the report format should fit into the participants existing or planned reporting structure provided that they include key reporting metrics required by the program. PSE should not claim energy savings if the report is not provided by an agreed deadline. This will incentivize both PSE and its customers to identify actions once it is known that savings have occurred.

PSE Response

It is the program policy that FAPs are to be collected only in the first year, and quarterly reports are required quarterly. It is fair to observe that the collection of these reports was inconsistent, as the program faced a lot of turnover and experienced difficulty recruiting and retaining customers as pointed out above. In addition to the requirement for an EMP described above, we will also be changing our quarterly reports to an Opportunity Register, a living document that will contain potential projects and interim schedules, and is applicable to all resources to include water, etc. As noted, the EMP is also aligned with the Clean Buildings Law, so that in will include the customer's current EUI, target EUI, and will document the steps the customer will take to achieve their goals.

Finding 4: System for Measuring and Reporting Energy Performance

Key Finding – The CSEM program is achieving electricity savings and natural gas savings. The differences between PSE’s estimates of savings and DNV’s estimates are due to the baseline periods used to estimate savings and challenges that DNV experienced with the acquisition of utility meter data which limited the number of sites included in the final sample.

Recommendation – PSE should not adjust current program savings using these realization rates. The savings evaluated were the last calculated before the COVID-19 pandemic and are not applicable to the current program. PSE should review the recommendations below to improve the program evaluability and mitigate risks associated with savings attribution.

PSE Response

Noted and accepted.

Finding 5: Baseline

Key Finding – DNV found opportunities for improvement in the defined baseline period as 23 of the 120 sites reviewed utilized baseline period from eight or more years prior to the performance period. The baseline period is used to develop the regression equation that estimates energy consumption in the absence of the program. DNV believes the baseline should be valid for the current loads in the building, meaningful to the participant, and changed if neither of first two are true. PSE's practice for the sites evaluated was to utilize a baseline period prior to the first year of a site's program participation, unless significant changes were identified and the customer agreed to a baseline change. In general, baselines in the evaluated sample were rarely updated. This had the benefit of preventing customer confusion and burden, but carried evaluability risks that grew over time. DNV requested monthly utility meter reads for all sampled sites from the start of their baseline period. In some cases, data from the baseline start were no longer stored by PSE. This prevented DNV from independently verifying the baseline period consumption using utility consumption data. DNV asked participants if the baseline period was meaningful to them and if the building use during the baseline period was similar to its use during evaluated performance period. While there were some cases in which older baseline periods were found to still represent evaluation period loads, most older baselines with either not meaningful or not representative of building use. In some cases, the respondent did not know whether the baseline was representative of building use as they were not at the organization during one or both periods. In many cases, participants viewed the 2018-2019 period as their baseline as it represents the prior normal consumption levels before changes due to the COVID-19 pandemic.

Recommendation – PSE should adjust its program design to ensure measurement baselines are meaningful to participating customers, representative of the existing conditioning loads in the building, and suitable for evaluation. DNV suggests updating site baseline models every six years at minimum. This is equivalent to two 3-year enrolment cycles and will ensure the utility meter data is still available within PSE's systems. PSE should consider opportunities to more regularly update baseline to align with the participant organization's energy reduction goals and staff commitments.

PSE Response

While we understand this recommendation from the evaluation perspective, we have concerns about limiting a baseline period to six years. The goal of the CSEM program is to not only acquire new energy conservation, but to maintain the behaviors that led to them over time. Our CSEM measures are reported with a 3-year measure life, which means that absent intervention, we assume that the behaviors that led to energy conservation will relapse after 3 years due to inattention and/or employee or management turnover. The continued engagement with customers is critical to ensuring conservation persists past the first three years. If we were to re-adjust the baseline to take into account the behaviors adopted, we believe this would mis-represent what's actually happening at the site, given the lapse in behavioral measures which occur without our engagement. A re-baselining would assume that all behavioral measures persist indefinitely, which we have not found to be the case and is not represented in our measure life assumptions. We believe the best time to reset *baselines is upon the occurrence of major additions or reductions, major occupancy or building use changes, or fuel switching.*

Finding 6: Reporting Energy Savings

Key Finding – DNV found that elapsed time between the close of the performance period and the date savings are claimed is too long for too many projects. In the sample of 35 projects evaluated, 50% of the projects had savings reported more than 6 months after the end of the performance period. Five projects contained sites with more than 1.5 years between the end of the performance period and the savings reported date. While time will always be required to complete the analysis, the program should avoid or greatly reduce significant delays.

Recommendation – PSE should review its program procedures and identify key factors that delay processing and reporting of energy savings. PSE should adjust its procedures to eliminate some, if not all, of the identified key factors. For example, DNV believes requiring an Annual Report of actions completed by an agreed deadline will reduce the time PSE waits for information on actions to support the estimated reductions in consumption.

PSE Response

We recognize that prompt payments are important for customer satisfaction, and we agree with this recommendation. Our program is currently developing a Python tool that lets us run multiple regression models more quickly, which should result in faster payments. Some of the late responses noted in the evaluation were due to lack of documentation from the customer and lapses due to PSE staff turnover. Our goal is to pay within 12 months. We also agree on the Annual Report of Actions, and we believe the quarterly updates we are making to the Opportunity Register should help with compiling the documentation necessary to make a performance payment.

Finding 7: Customer Engagement

Key Finding – DNV found that CSEM participants were insufficiently engaged in the measurement process. The measurement of energy savings in an SEM program is intended to improve the participants understanding of how and when energy is consumed. The measurement process provides each participant with information and feedback regarding their attempted improvements to energy management. Participants interviewed for this evaluation regularly stated that they did not participate in the calculation of savings, that they had not reviewed PSE's calculation spreadsheets, or that they did not trust the consumption values provided by PSE's MDM system as they rarely matched their actual billing records.

Recommendation – PSE should adjust its program design to increase participant engagement in the measurement process. The measurement process should not just be a step PSE does to calculate a participant's performance incentive. The process should be a platform through which a participant can learn about their consumption and continue to identify opportunities for improvement. PSE should consider adjustments to its calculation methods that better align the data used for the analysis with a participant's actual utility meter reads.

PSE Response

We agree that our previous approach to measurement led to some inconsistencies in billing data that resulted in customer confusion. This is one reason why we've spent the last year transitioning to a different system, Energy Cap. Energy Cap is linked directly to our system of record for billing and consumption data, which should resolve some of the inconsistencies and lack of transparency found with MDM. Energy Cap should also provide the customer with more insight as to the basis for their energy calculations. We will consider the recommendation to involve customers more in the energy calculations, though we have not found interest among our customers in being involved in the billing analysis. We expect the transition to Energy Cap to resolve the issues identified.

Finding 8: Meter Data for Evaluations

Key Finding – DNV found that PSE was unable to provide a sufficient amount of participant consumption data supporting the energy savings claimed. PSE's measurement process was completed within MDM or completed by PSE staff using spreadsheets. In both cases, the total consumption during the baseline and performance periods is listed. When spreadsheets were used, PSE included the total consumption by calendar month for each participant. PSE lists the account and meter numbers associated with every participating site in the grant agreement workbook. To verify the consumption and savings estimated, DNV requested monthly consumption data showing the meter read date for the sampled sites using the account and meter numbers listed.

Recommendation – PSE should review its program documentation and design to determine what changes are necessary to ensure that they can provide evaluators with the utility meter data for program participants. DNV believes the recommended change to re-baseline participants every 6 years will help mitigate the risks associated with changing account and meter numbers. One possible adjustment is to store the actual monthly utility meter consumption data for the performance period by account and meter in a separate file in the annual analysis folder instead of only having the aggregated calendarized consumption in the analysis file.

PSE Response

We believe this recommendation will be addressed with the transition to Energy Cap. As noted, it will link via direct API's to SAP, our system of record for billing and consumption. To mitigate the issue associated with meter and account numbers, Energy Cap will link customers by a Business Agreement Number, which should be permanently linked to the customer even in the cases where meter or account numbers change, thereby maintaining consistent reporting for each customer.

Finding 9: Alignment with Clean Buildings Performance Standards

Key Finding – The majority of participants interviewed were aware of the Clean Buildings Performance Standards in Washington and the associated potential future penalties.

Recommendation – PSE should identify methods to support CSEM participants goals specific to the Clean Buildings Performance Standards. The goals of the program and the Clean Buildings Performance Standards are well aligned, and the existence of potential future penalties have motivated many participants to reengage with SEM. PSE could organize or support ASHRAE audit training courses for program participants as part of their annual training allowance. PSE could adjust its RMP, FAP, and PAM documentation to align with the Energy Management Plan (EMP) and Operations & Maintenance (O&M) Programs required to comply with the Clean Buildings Standard. PSE could ensure the data provided through this program can be easily integrated into the reported platforms used for Clean Buildings.

PSE Response

> *We strongly agree with this recommendation and have already taken steps to align our customer engagements with the requirements of the Clean Buildings Performance standards. We will be including training on Energy Start Portfolio Manager, and better alignment with the Clean Buildings Law was one of the reasons behind our transition to Energy Cap, which better incorporates Portfolio Manager data.*

Finding 10: Decarbonization

Key Finding – Many large institutional participants are moving away from energy savings goals and towards decarbonization goals. These participants are currently and expect to increase electric loads in order to reduce their use of carbon-based fuels. Calculating SEM energy savings for these customers while they are increasing their electric loads will be challenging.

Recommendation – PSE should remove participants or sites from the CSEM program and transition them to applicable clean energy programs when they plan to decarbonize their heating by switching to electric heating technologies. Participant sites could return to the program after the new system is in place for one year and a new baseline has been established. PSE should complete additional review of any participating customers transitioning off a central heating plant to decentralized electric heating. It is likely best to remove all impacted sites from the program until the new baseline has been established. Participants should also be encouraged to separately meter any vehicle charging loads added to their portfolio.

PSE Response

> *We agree, and we have already implemented this as program policy. When a customer transitions their gas fuel to electric, they enter a new baseline year against which any energy improvements will be measured.*