GROUP D

Non-Residential Population-Based Normalized Metered Energy Consumption (NMEC) Evaluation, Measurement, and Verification Report – Program Years 2022 - 2024

California Public Utilities Commission
CALMAC ID: CPU0406

Date: July 13, 2026

Table of contents

List of figures

List of tables

Glossary of key terms and acronyms

Avoided Cost Calculator (ACC) – ACC estimates hourly, system-level costs of providing electric or gas services for 30 years and is updated every two years, with the most recent major update in 2024. Total system benefit uses avoided costs to represent the total present value lifecycle benefits of energy efficiency programs.

Avoided energy use (AEU) – The reduction in energy consumption that occurs because of an efficiency measure, program, or action, relative to what would have happened without it. It reflects the difference between observed and counterfactual consumption over the performance period and therefore captures realized performance-period conditions.

Baseline period – The 12-month period leading up to the energy efficiency intervention or retrofit.

California Database for Energy Efficiency Resources (DEER) – This database contains information on energy efficient technologies and measures. DEER provides estimates of the energy-savings potential for these technologies in residential and non-residential applications. DEER is used by California Energy Efficiency (EE) Program Administrators (PAs), private sector implementers, and the EE industry across the country to develop and design energy efficiency programs.

California Energy Data and Reporting System (CEDARS) – This database securely manages California Energy Efficiency Program data reported to the Commission by investor-owned utilities, regional energy networks (RENs), and certain community choice aggregators (CCAs).

Coefficient of variation of the root mean square error (CV(RMSE)) – A model goodness-of-fit statistic, a measure of variability (the square root of the consumption model's squared error) relative to the average value of the variable (in this case, average energy consumption) used to determine how well the model predicting the variable (in this case, baseline consumption) fits the data. The lower the CVRMSE, the better the model fit.

Cost-effectiveness tool (CET) – An online tool designed for the CPUC to determine the cost effectiveness and examine other properties of EE programs and portfolios.

Counterfactual energy consumption – It is the energy use that would have occurred during the performance period in the absence of the implemented measure or program.

Documented savings – Meter-based savings calculated by the PAs at the end of the reporting period, which serve as the basis for final claimed savings. We refer to these as documented savings, as the PAs provide the supporting data, models, and results as part of the evaluation.

Effective useful life (EUL) – An estimate of the median number of years that installed measures are still in place and operable.

Final claimed savings – Meter-based savings calculated by the PAs at the end of a project's reporting period, covering the 12-month period after installation. Final claimed savings equal the initial savings the PAs claim in CEDARS at the time of installation, plus any true-up savings based on meter-based calculations.

Forecast realization rate (FRR) – The ratio of verified savings to initial claimed savings used to assess the reliability of planning assumptions and the extent to which programs delivered on their initial savings promise.

Forecast savings – Program-level deemed or engineering-based savings estimate calculated before installation.

Fractional savings – The percent of annual energy usage saved through program participation. For NMEC projects, the rulebook recommends that projects have a forecasted fractional savings of at least 10%.

Fractional savings uncertainty (FSU) – CV(RMSE) combined with percent savings; this statistic is similar to relative precision in that it measures the uncertainty around the expected savings. As the value of FSU decreases, confidence in the estimated savings level increases.

Granular profiles – These are mean hourly consumption profiles constructed from a sample of non-participant customers that share the characteristics of participating sites, such as industry (NAICS), climate zone, and size.

Gross realization rate (GRR) – The ratio of verified energy savings to final claimed energy savings. As a multiplier on Unit Energy Savings, the GRR considers the likelihood that not all CPUC-approved programs undertaken by PAs will come to fruition.

Gross savings – The energy savings from installed energy efficiency measures irrespective of whether those savings are from free riders, i.e., those customers who would have installed the measure(s) even without the incentives offered under the program.

Impact shape – A unitized hourly profile of annual electric or gas savings that represents the distribution of savings across the hours of the year.

Initial claimed savings – For NMEC projects, the aggregate of project-level savings or forecasted savings claimed in CEDARS following project implementation.

International Performance Measurement and Verification Protocol (IPMVP)[1] – A standardized approach to measuring and verifying energy efficiency investments. IPMVP incorporates M&V best practices in a non-prescriptive framework, allowing it to be applied flexibly based on a measure's application and the available information.

Lifecycle savings – The savings associated with the lifetime of an efficiency measure undertaken by a program participant. Equipment replaced early in its useful life might receive reduced savings to account for the untapped useful life of the outgoing equipment.

MAP-style – A pay-for-performance (P4P) population NMEC program style with an implementation structure similar to the CPUC's Market Access Programs.

Measure – A specific customer action that reduces or otherwise modifies energy end-use patterns; a product whose installation and operation at a customer's premises reduces the customer's on-site energy use.

Meter-based savings – Project-level savings measured or estimated using metered energy consumption data. These estimates are derived from metered data and energy-consumption models.

Net savings – The savings realized after accounting for free-ridership, calculated by multiplying gross savings by the net-to-gross ratio.

Net-to-gross ratio (NTGR) – A ratio or percentage of net program savings divided by gross or total impacts; used to estimate and describe the free-ridership that may be occurring within energy efficiency programs.

Non-MAP-style – A population NMEC program style recruiting and enrolling customers through direct trade ally partnerships.

Non-routine event (NRE) – A change not related to the energy efficiency intervention but affecting energy use in the baseline or the reporting period and which must be accounted for in savings estimations. Typical NREs include changes in

[1] Efficiency Valuation Organization (EVO). https://evo-world.org/en/.

facility size, changes in facility activity not affected by the energy efficiency measures (such as addition or removal of a data center), or other modifications to the facility or its operation that alter energy consumption patterns and are unrelated to the program intervention.

Normalization – A process by which consumption estimates from two different periods are put on a common basis. Baseline and performance period model predictions are observed at common values for the model's independent variables, e.g., temperature variables from typical meteorological year (TMY) data. Normalization accounts for differences in underlying drivers of consumption during the baseline and performance periods.

Normalized metered energy consumption (NMEC) – High opportunity programs or projects that provide incentives based on metered energy consumption. This initiative fulfills the directive for utilities to quickly identify high energy-efficiency savings opportunities in existing buildings using a program and project approach where incentive payment and claimed savings are based on NMEC and include only approved NMEC building programs.

NMEC rulebook – The CPUC's formal framework that defines the methods, data requirements, and reporting standards for estimating savings using meter-based approaches. It ensures that energy savings calculations are transparent, replicable, and accurately account for variables like weather or operational changes.

Normalized savings – Savings calculated as the difference between a weather normalized baseline and performance period statistical models.

On-Bill Financing (OBF) – A program style that does not offer incentives and instead finances participants' energy efficiency projects with zero-interest loans repaid through the participant's utility bill.

Parameter – Output of a regression model. For NMEC models, parameters measure how fuel consumption changes in response to a change in a given independent variable.

Peak demand – The maximum level of metered demand during a specified peak demand period for installed or implemented measures. CPUC Resolution E-4952 approved the Database for Energy-Efficient Resources (DEER) for 2020 and revised the DEER Peak Period definition to 4:00 p.m. to 9:00 p.m. effective January 1, 2020

Performance period – The 12-month period following the energy efficiency intervention or retrofit, during which savings are realized.

Projected savings – Project-level forecast savings based on deemed values or engineering calculations that serve as the basis for the PAs' population-based NMEC aggregate claims.

Program administrator (PA) – An entity tasked with the functions of portfolio management of energy efficiency programs and program choice, i.e., Pacific Gas & Electric (PG&E), Southern California Edison (SCE), Southern California Gas (SCG), and San Diego Gas & Electric (SDGE), Inland Regional Energy Network (IREN).

p value – The probability that a given parameter's true value is different from zero.

Relative precision – A ratio of the error bound divided by the value of the measurement itself. This provides the error on a relative basis, frequently used to show uncertainty as a fraction of a quantity. In this report, all relative precisions are provided at the 90% confidence interval, which means that in repeated sampling, 90 times out of 100 the true value will fall within the lower and upper bounds of the estimate.

Spline – A model that is a collection of lines with different slopes that change at defined points (nodes), allowing for more flexible response to the given independent variable than a constant linear relationship.

Synthetic control – In the context of NMEC, it is a statistical modeling approach that incorporates a granular comparison-group load profile into the regression model to control for non-program effects.

Temperature node – In a temperature spline model, a boundary temperature at which the slope changes.

Time-of-week and temperature (TOWT) model – A standard regression model approach whereby fuel consumption is modeled against temperature, included as a spline and a set of time-of-week indicator variables, generally at the daily- or hourly-level. May be split into occupied and unoccupied models. Other variables may also be included.

Total System Benefit (TSB) – TSB is an expression, in dollars, of the lifecycle energy, ancillary services, generation capacity, transmission and distribution capacity, and GHG benefits of energy efficiency activities, on an annual basis.

True-up savings – The savings claimed in CEDARS following the end of the performance period. This value is expected to be the difference between initial claimed savings and the normalized savings.

Typical meteorological year (TMY) – A data set of temperatures representing a typical year and used to normalize NMEC models to weather conditions. The CALEE CZ data sets are the standard used for NMEC.

Verified savings – Program-level savings confirmed by the third-party independent evaluation, such as the current evaluation.

1 EXECUTIVE SUMMARY

This report presents DNV's evaluation of the non-residential Population-Based Normalized Metered Energy Consumption (pop NMEC) programs for program years (PY) 2022 – 2024 on behalf of the California Public Utilities Commission (CPUC). This study evaluates the impacts of the pop NMEC programs implemented by six California program administrators (PA) and assesses how effectively the programs support energy efficiency projects.

1.1 Study background

NMEC is a program pathway that prioritizes performance-based savings estimates using premise meter data. Unlike traditional energy efficiency programs that claim final savings based on engineering or pre-determined savings estimates that are not directly empirical,[2] NMEC programs calculate and claim final savings based on actual measured changes in energy consumption at the meter (meter-based savings).

This evaluation builds on the PY2019 – PY2021 pop NMEC evaluation[3] and follows the NMEC workplan.[4] The prior evaluation included multiple residential programs and just one non-residential program. The current evaluation focuses on non-residential pop NMEC programs. Since 2021, the number of non-residential pop NMEC programs has grown from one to eight in the evaluation period, with seven still operating at the time of writing.

This evaluation estimates how much energy customers save at the meter (energy impacts), the grid impact of those savings, Total System Benefit (TSB),[5] and whether the programs motivated those savings (program influence). In addition to the actual energy saved, assessing the level of program influence is key to this evaluation as it indicates how effectively programs use ratepayer dollars to drive energy efficiency improvements by spurring new projects or expanding or accelerating planned projects.

1.2 Evaluation objectives

In this evaluation, DNV assessed the accuracy of the savings of the PY2022 – PY2024 non-residential pop NMEC programs based on claimed energy savings and analyzed how savings are distributed across participants and programs. We also assessed whether current measurement and verification (M&V) practices and the CPUC's "Rulebook for Programs and Projects Based on Normalized Metered Energy Consumption version 2.0" (NMEC rulebook)[6] provided a sufficient foundation for successful program design and implementation. To guide this work, the evaluation focused on three primary objectives:

[2] Interventions in traditional energy efficiency programs typically consist of the installation of energy efficient technologies with researched, vetted, and predictable savings that are well-established. We refer to savings from these energy efficiency technologies as "deemed" or "calculated" savings. The savings from installing a high-efficiency lightbulb, for example, are well-established, and non-NMEC energy efficiency programs could use those known values to claim savings. This contrasts with custom energy efficiency technologies and services that require unique savings calculations and do not use predefined values.

[3] DNV, "Population-Based NMEC - Program Years 2019 – 2021," calmac.org, December 22, 2023, https://www.calmac.org/%5C/publications/Final_Population-Based_NMEC_Program_Impact_Evaluation_Report_-_Program_Years_2019_to_2021_toCALMAC.pdf

[4] DNV, "Normalized Metered Energy Consumption (NMEC) Evaluation, Measurement, & Verification Work Plan," October 29, 2025, https://pda.energydataweb.com/#!/documents/4237/view.

[5] The CPUC adopted TSB as the new goal metric for energy efficiency programs starting in PY2024. TSB assigns a dollar value to the lifecycle energy, capacity, and greenhouse gas benefits of a program, encouraging "high value" load reduction and longer-duration energy savings while being fuel agnostic.

[6] CPUC, "Rulebook for Programs and Projects Based on Normalized Metered Energy Consumption," January 7, 2020, https://www.cpuc.ca.gov/-/media/cpuc-website/files/legacyfiles/n/6442463694-nmec-rulebook2-0.pdf. We evaluated PY2022 - PY2024 programs under this version of the NMEC rulebook, not version 2.1, which was released in 2025.

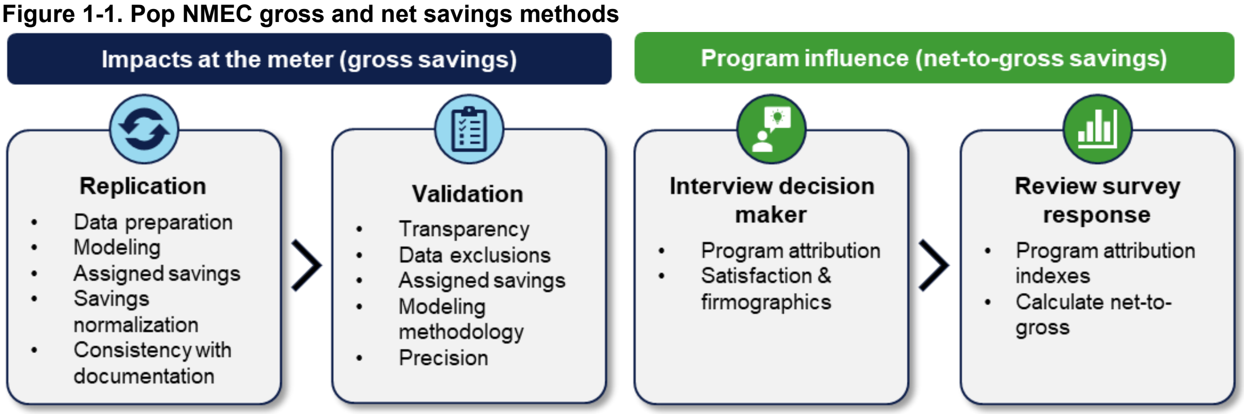
1. **Evaluate the accuracy and attribution of programs' final claimed savings**. Given the embedded evaluation structure of pop NMEC programs, DNV sought to replicate and validate the existing M&V frameworks and the program savings they produced, including TSB, and to determine program-driven savings.
2. **Assess the distribution of verified savings among participants, programs, climate zones, as well as when those savings deliver the greatest system value.** We examined participation in pop NMEC programs and evaluated how program savings are distributed across participants, program types, and sectors. We also assessed how well the programs deliver benefits during periods of high system cost and carbon intensity.
3. **Identify whether program M&V plans and practices and NMEC rulebook protocols are sufficient to ensure successful pop NMEC implementation.** DNV also assessed how programs are learning and adapting their offerings to improve savings reliability. Additionally, we addressed whether eligibility processes are being applied in a manner that is transparent, consistent with rulebook expectations, and designed to serve customers and meet policy goals.

1.3 Study approach

This study evaluated impacts at the meter (gross savings)[7] and program influence (net-to-gross or NTG savings).[8] For the gross savings evaluation, DNV followed a two-step process to validate and replicate the final claimed savings of PY2022 and PY2023 non-residential programs.[9] Because pop NMEC programs are composed of projects at individual participating sites, meter-based savings are derived from site-level energy use models and then aggregated at the program level. While validating these aggregate savings claims, DNV reviewed and assessed the underlying site-level models. These individual site models help evaluate pop NMEC eligibility, as their statistical performance indicates how well a model that is consistent with NMEC rulebook assumptions can characterize site-level consumption.

For the NTG savings evaluation, DNV investigated how much the programs influenced participant decisions to make energy efficient improvements. Figure 1-1 illustrates the gross and net savings methodology in more detail.

Figure 1-1. Pop NMEC gross and net savings methods



[7] Gross savings are the energy savings from installed energy efficiency measures irrespective of whether those savings are from free riders, those customers who would have installed the measure(s) even without the incentives offered under the program.

[8] Net-to-gross ratios (NTGR) estimate the degree to which customers would have installed the program technology or equipment without the program benefits. Gross savings multiplied by the NTGR produce net savings. Savings attributable to participants who would have purchased energy efficient technologies with or without the program influence are excluded from net savings. These participants, whom the program did not influence, are considered free riders.

[9] Final claimed savings refer to the meter-based savings calculated by the PAs at the end of a project's reporting period, covering the 12-month period after installation. Final claimed savings equal the sum of the initial savings that PAs claim in CEDARS at the time of installation and any true-up savings based on meter-based calculations.

1.4 Evaluated projects

The study evaluated eight non-residential pop NMEC programs, summarized in Table 1-1 by program delivery style and by initial claim program year. The programs adhered to three general designs.

- Five programs, similar in design to the CPUC's Market Access Program (**MAP-style**), implement projects through a marketplace platform, recruiting and managing a network of incentivized contractors (aggregators), who bundle projects across multiple sites and manage customer relationships.
- Two **non-MAP-style** programs implement projects using a traditional model, recruiting and enrolling customers through direct trade ally partnerships.
- The On-Bill Financing-Alternative Pathway (**OBF**) program is PA-administered, but, uniquely among the programs, does not offer incentives and instead finances participants' energy efficiency projects with zero-interest loans repaid through the participant's utility bill.

Table 1-1. Pop NMEC programs by initial claim year

Program style	PA*	Program name	Initial claim PY		
			2022	2023	2024
MAP-style	BayREN	Business FLEXmarket		✓	✓
	PCE	FLEXmarket Program		✓	✓
	MCE	Commercial Efficiency Market	✓	✓	✓
	SDG&E	CORE Market Access Program			✓
	SCP	SCP FLEXmarket		✓	
Non-MAP-Style	PG&E	Institutional Partnerships		✓	✓
	PG&E	NetOne Commercial Efficiency			✓
OBF	PG&E	On-Bill Financing-Alternative Pathway (OBF)	✓	✓	✓

* The six PAs with programs included in this evaluation are Bay Area Regional Energy Network (BayREN), Marin Clean Energy (MCE), Pacific Gas and Electric Company (PG&E), Peninsula Clean Energy (PCE), Sonoma Clean Power (SCP), and San Diego Gas and Electric Company (SDG&E)

1.4.1 PY2022 – PY2023 evaluated savings

In this section, we present the evaluated impacts and program influence of six PY2022 – PY2023 non-residential pop NMEC programs with a total of eight electric and two gas program-year cohorts (cohorts).[10] Figure 1-2 presents a comparison of our verified electricity, peak demand, and gas savings against the PAs' final claimed savings for these program cohorts by program style.[11] This comparison of verified and claimed savings is the gross realization rate (GRR). This figure does not include the PY2024 program cohorts as they do not yet have final claimed savings that reflect meter-based results. The final claimed savings represent the meter-based savings claims the PAs made in the CEDARs tracking database.[12]

Impacts at the meter measure the basic efficiency savings and reduced grid demand without any consideration of program influence. These benefits enhance reliability and lower system costs for all ratepayers. The figure does not include results

[10] Cohort refers to all customers with installations in a given program year within a program.

[11] Verified savings refer to program-level savings confirmed by a third-party independent evaluation, such as the current evaluation.

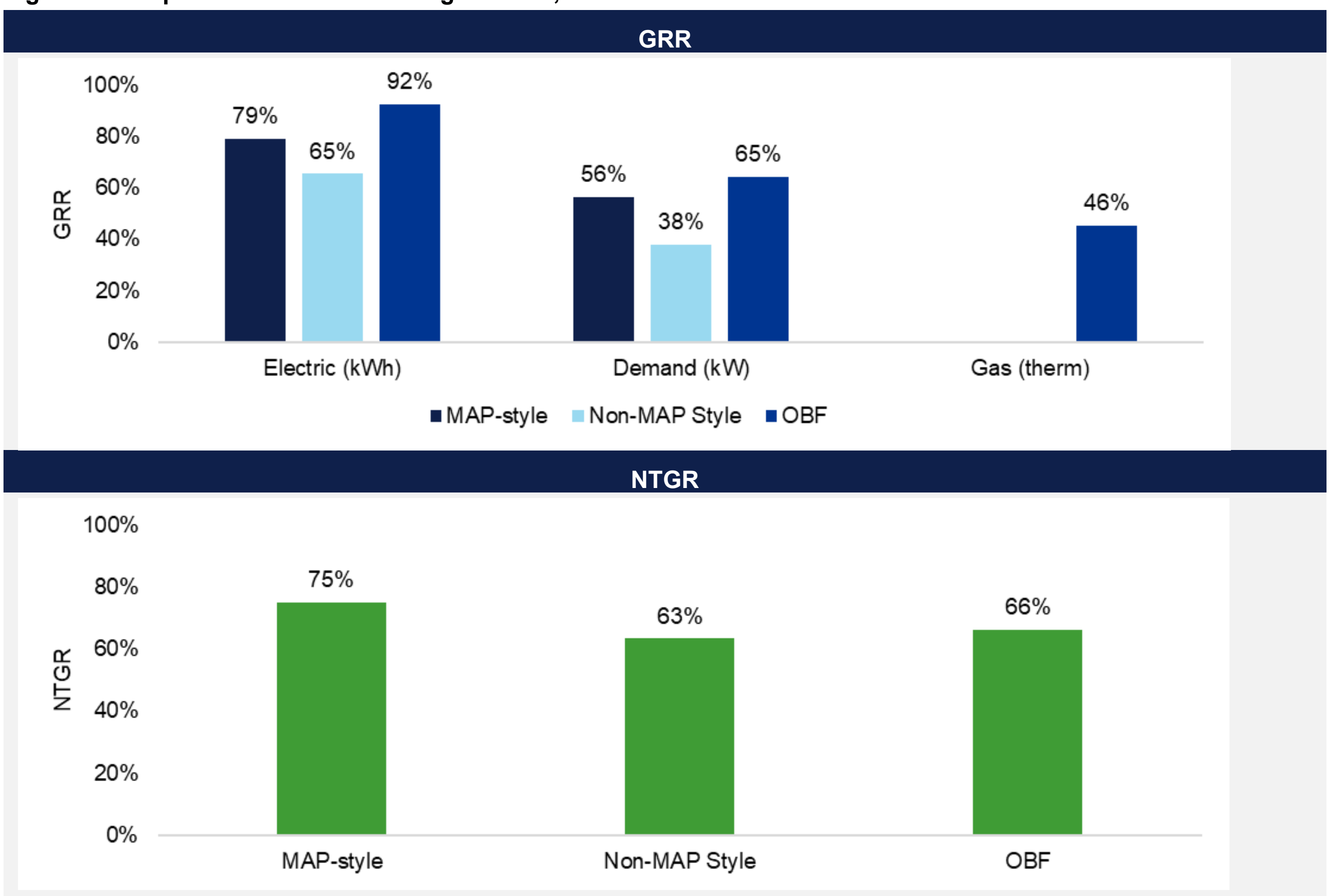
[12] The California Energy Data and Reporting System (CEDARS) is a database that hosts California Energy Efficiency Program data reported to the Commission by program administrators (PAs), including investor-owned utilities, regional energy networks (RENs), and certain community choice aggregators (CCAs).

from SCP because it could not provide the FLEXmarket models and data we needed to replicate and validate the savings claims.[13] The gas cohorts were smaller, and several models showed an increase in energy use.

The figure also presents net-to-gross ratios (NTGRs), which capture the degree of program influence on a customer's decision to install program equipment. This measure provides some confidence that programs did not incentivize work that would have occurred regardless of the program. Because this evaluation focuses on non-residential customers, we applied the same NTG methodology used for site-level NMEC, which, consistent with California non-residential NTG protocols, relied on participant interviews to assess program influence.[14]

We interviewed a representative sample of participants, covering 162 sites (12% of the population). Information from the interviews helped determine the NTGRs. We calculated NTGRs only for those sites with first-year gross electric savings. We were not able to produce different ratios for lifecycle electric savings claims or for demand or gas, as we did not have the project-level savings and measure life information needed for those calculations. The NTGRs range from 50% to 100% across PAs.

Figure 1-2. Pop NMEC evaluated savings results, PY2022 – PY2023



[13] The PA deleted all program information, including customer data and modeled results, required to claim final savings.

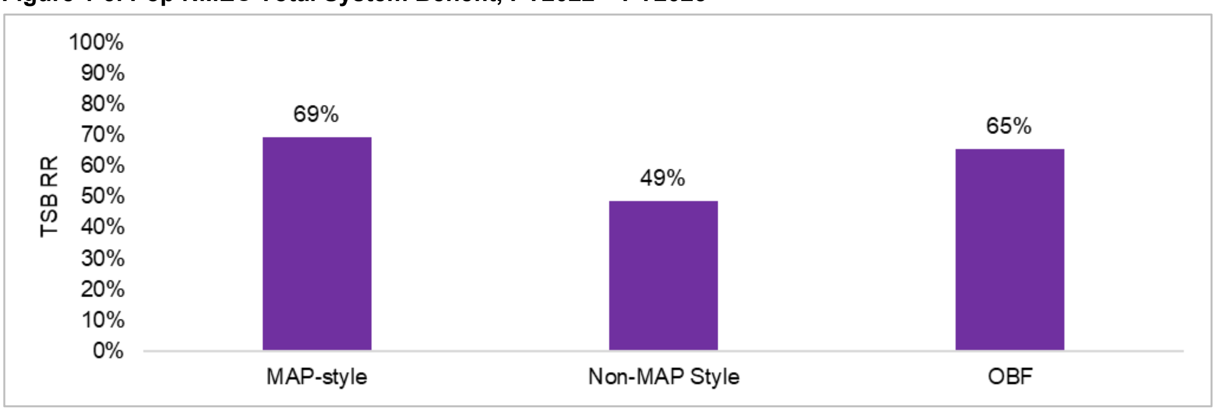
[14] California Non-residential Net-to-Gross Working Group, *Methodological Framework for Using the Self-Report Approach to Estimating Net-to-Gross Ratios for Non-residential Customers*, (CPUC), February 20, 2015, https://pda.energydataweb.com/api/view/1255/IALC_NTGR_Methods_2013-14.pdf.

1.4.2 PY2022 – PY2023 Total System Benefit

Figure 1-3 presents the TSB realization rate (RR) for these same cohorts. TSB is a newly adopted metric that incorporates the value of savings to the grid in each hour extended over the lifetime of the program savings. TSB became the primary goal by which programs are assessed starting in PY2024.[15] Because NMEC claims span two program years, the cohorts included in this evaluation submitted initial claims in PY2022 and PY2023 and were trued-up in PY2023 and PY2024, respectively. For reporting, NMEC TSB savings are associated with the installation year to remain consistent with other program tracks that also claim TSB savings in the year of installation. For this first TSB evaluation, we updated the verified energy savings, the NTGR, and the avoided cost year applied in the TSB calculation. We used the results discussed in Section 4 as the verified savings and NTGR inputs. Within the TSB calculation, the avoided cost year determines both the avoided cost version and the savings start date. All projects within a cohort should be installed in the year of the initial claim and should use the same avoided cost year for both the initial and true-up claims, aligning TSB with the period over which the cohort's impacts actually accrue.

However, when we replicated the TSB values in CEDARs, we found that the PAs used different avoided cost years for the initial and true-up claims. This inconsistency undermines savings measurability. To take just one example, if a project that used two different avoided cost years sought to zero out its initial claim with an equivalent negative true-up claim, either because the project was never actually installed or it underperformed initial expectations, the TSB metric would not, in fact, be zeroed out because the avoided costs would be different. To address this issue, we used the initial year avoided cost version and the savings start date as the basis for the verified TSB. Consistent with this transitional context, pop NMEC programs have not yet fully adopted cohort-specific, meter-based hourly profiles of energy consumption. For PY2022 – PY2024 TSB claims, PAs relied on existing generic profiles of energy savings shapes. Additionally, three cohorts used different energy savings profiles for their initial and true-up claims,[16] effectively mixing profiles in unintended ways and producing problematic results. To address these inconsistencies, the claimed TSB values presented here place all program claims on an initial-year basis, including consistent use of the initial claim impact profiles. With this consistent approach, the evaluated TSB results are broadly consistent with a combination of GRR and NTGR, both of which are reflected in the verified TSB calculations.

Figure 1-3. Pop NMEC Total System Benefit, PY2022 – PY2023



---

[15] From CPUC Decision 21-05-031: "Beginning with the energy efficiency potential and goals study being conducted in 2021, energy efficiency goals for program administrators shall be expressed in terms of a Total System Benefit (TSB) metric. Beginning in program year 2022, the TSB metric shall be reported by each energy efficiency program administrator. Beginning in program year 2024, the TSB metric will replace the energy and peak demand savings goals as the single goals metric, but portfolio outcomes will continue to be reported in terms of energy and peak demand savings, in addition to TSB." https://docs.cpuc.ca.gov/PublishedDocs/Published/G000/M385/K864/385864616.PDF

[16] BayREN used a commercial indoor lighting profile for the initial claim and an indoor compact fluorescent (CFL) profile for the true-up claim; MCE's 2022 cohort used a refrigerator and freezer recycling profile for the initial claim and a commercial indoor lighting profile for the true-up claim; MCE's 2023 cohort used a commercial indoor lighting profile for the initial claim and an indoor CFL profile for the true-up claim.

1.5 Key findings & recommendations

Pop NMEC programs spent $18.6 million and procured $17.8 million verified TSB. Despite the verified GRR and NTGR values, underlying documentation, eligibility, and other program risks remain and need to be addressed. The findings and recommendations below lay out critical lessons learned and actionable next steps to further strengthen program performance and address these remaining risks.

Pop NMEC programs aim to deliver credible, performance-based savings while expanding access to energy efficiency to customers for whom traditional program approaches may be challenging or overly cumbersome. While pop NMEC can clearly motivate meaningful savings, two primary areas need substantial attention: 1) strengthening savings measurability and 2) achieving stranded savings that are consistent with core NMEC objectives. Savings measurability can be improved through a combination of stronger adherence to program rules regarding the appropriate type of load for pop NMEC meter-based methods and using technical methods that provide transparent and consistent results. Stranded savings, the up-to-code savings for which NMEC methods are designed, can be realized by focusing on customers with demonstrated technical or financial barriers and supporting comprehensive upgrades that would otherwise be deferred or forgone.

1.5.1 Savings claims accuracy recommendations

Ensuring savings measurability is a top priority as pop NMEC programs mature. To meet savings measurability and accuracy expectations, pop NMEC programs need to recruit and serve customers whose meter data will support standardized population, meter-based approaches. Additionally, pop NMEC methods need to be streamlined so that they provide consistent final estimates of savings across PAs and implementers.

1.5.1.1 Program design improvements

Savings accuracy in pop NMEC programs depends substantially on participant eligibility criteria and screening, as well as program documentation management.

As many as 15% of participants in PY2022 – PY2023 came from sectors that cannot be modeled accurately under pop NMEC. The NMEC rulebook limits pop NMEC participation to commercial or commercial-like sites and excludes sites with industrial or process-driven loads. These limits are consistent with site-level NMEC programs and support standardized, pop NMEC modeling. Nevertheless, approximately 15% of participants in the PY2022 – PY2023 programs came from sectors that commonly exhibit process- or event-driven energy consumption (manufacturing, agriculture, infrastructure and utilities, etc.), which cannot be modeled accurately under pop NMEC. The participation of these ill-suited sites affects savings accuracy because their energy consumption patterns are often so variable they cannot provide a reliable baseline from which to estimate savings.

Program savings accuracy can be improved with more homogenous populations. The NMEC rulebook stipulates pop NMEC cohorts should include sites with similar building types, equipment, drivers of energy consumption, and levels of energy use. BayREN's program was largely composed of retail sites but also included one large manufacturing site, which was enough to dramatically increase savings uncertainty. PG&E's OBF program reported difficulty forming cohorts that were both sufficiently similar in project characteristics and large enough to support pooled population level analysis but was the only program to meet 90/25 precision for its impact estimates. By addressing highly variable levels of energy consumption among their populations, all programs can better align with the NMEC rulebook and improve their savings precision, both of which are essential program goals.

Insufficient pre-participation screening allowed participation of unsuitable sites, leading to non-meter-based savings assignments, affecting more than 60% of gas OBF sites. Pop NMEC programs require effective eligibility screening to ensure that participating sites have sufficient and appropriate data to support meter-based measurement. These eligibility requirements are detailed in program M&V plans. When PAs do not identify or screen out ineligible sites during enrollment, programs cannot develop robust, transparent, and defensible final savings claims. The OBF pre-screening, for example, did not exclude all sites with inadequate pre-installation data. As a result, approximately 7% of electric sites and more than 60% of gas OBF sites ultimately received non-meter-based savings estimates. These kinds of lapses in eligibility screening directly affect savings accuracy, compromising the analysis population and undermining the use of performance-based estimates to support program-level claims.

Nearly all programs faced challenges maintaining clear, consistent, and transparent documentation from enrollment through final M&V. Because savings claims are made at the pop NMEC program-cohort level and not at the individual project level, the PAs and program implementers alone are currently responsible for gathering, organizing, and documenting data to support the pop NMEC savings claims. However, the PAs and implementers did not maintain clear records connecting individual projects to the CEDARS claims that PAs make at the program-cohort level. Only BayREN could produce documentation that reconciled with initial CEDARS claims without requiring substantial investigation from the PA and the evaluation. One PA deleted the documentation that could have supported its performance-based results, and we could not verify the savings claimed. Another PA provided two inconsistent datasets for the same purpose, neither of which included projects whose documented savings matched the final claimed savings.[17]

Recommendations 1 & 2	Eligibility rigor and auditable records

- PAs should strengthen participant eligibility criteria by enforcing existing M&V eligibility criteria—including the requirement for sufficient pre-installation data—and tightening upfront screening rules to exclude sites with known process-, event-, or production-driven loads. Eligibility screening should also include size limits, restricting participation to sites with annual energy consumption below a pre-defined threshold (for example, 1,000,000 kWh). These criteria will ensure enrollment is limited to sites suitable for standardized pop NMEC modeling, homogenous cohort formation, and compliance with NMEC rulebook precision requirements.
- PAs should propose a solution to the reporting PCG that defines a customer project-level pop NMEC dataset that links with 100% accuracy to both initial and true-up aggregate CEDARS claims by mid-September 2026. The solution should reflect an official, date-stamped set of required project claim elements, including, but not be limited to, claim and parent claim IDs; project identifiers; participant (end-user) IDs and contact information; projected[18] and true-up savings for all fuel types (kWh, kW, therms) and TSB; installed measures; EULs; realization rates; NTG values; and implementation partners (implementer, aggregator or contractor, and M&V vendor).

[17] For pop NMEC, PAs claim aggregate program-level final savings for projects that occur in a particular program year. Documented savings refer to the meter-based savings calculated by the PAs at the end of the reporting period, which serve as the basis for final claimed savings. We refer to these as documented savings, as the PAs provide us with the supporting data, models, and results as part of the evaluation.

[18] Projected savings refer to savings forecasts made at the individual project level prior to implementation.

1.5.1.2 Technical improvements

Savings accuracy in pop NMEC programs also depends on the selection and application of consistent modeling methods and on consistent and appropriate characterization of uncertainty.

The use of different methods to calculate savings increased variability and risk in savings results. Because program implementers used different modeling and non-participant selection approaches, which are required to account for non-program effects, different assumptions shape savings estimates, and those estimates are sensitive to different sources of error. As a result, differences in final claimed and verified savings may reflect methodological choice as well as underlying program performance. This reduces the confidence of stakeholders, including participating customers and contractors, in cross-program comparisons and runs counter to pop NMEC's standardization intent.

Approaches to modeling energy use at the site level varied in their alignment with NMEC rulebook expectations. While all evaluated program implementers used modeling approaches to account for routine, recurring drivers of consumption, such as weather, and non-program-related factors, methodological choices did not always best serve savings accuracy. For example, only one program normalized savings to typical weather conditions, consistent with NMEC rulebook intent. Other programs used reporting-year weather, which increases savings errors in multiple ways. Further, aspects of the modeling approaches used extend beyond capturing routine drivers of consumption and can also incorporate within-year trends and NREs into modeled consumption. This reduces the ability to distinguish routine consumption patterns from non-routine variation, with implications for eligibility screening and the interpretation of hourly savings profiles.

Most program implementers' uncertainty methods and reporting do not provide sufficient evidence to support the credibility of claimed pop NMEC savings under the NMEC rulebook's precision expectations. Characterizing uncertainty is essential for establishing claimed pop NMEC savings are valid. Despite this, only two programs reported their calculation of uncertainty metrics. In addition, common uncertainty practices (e.g., reliance on FSU[19]) may not fully reflect sources of program-level uncertainty beyond model error, such as the uncertainty introduced by the selection of comparison sites to account for non-program effects. By this evaluation's uncertainty calculations, most programs did not meet the NMEC rulebook's 90/25 precision standard. It may be appropriate to relax this standard to a 90/50 benchmark for these non-residential cohorts with relatively small sample sizes, but this will only be reasonable if the PAs take uncertainty characterization seriously as a key indicator of program success.

Recommendations 3 & 4	Consistent methods and uncertainty treatment

- The PAs and their implementers should participate in a stakeholder process to develop more consistent modeling approaches for pop NMEC programs that prioritize clear representation of weather and schedule effects, facilitate identification of NREs that influence eligibility decisions, and support accurate TSB estimates.
- The PAs should apply program-level pop NMEC savings uncertainty approaches that reflect both modeling error and other sources of cross-site variability.

1.5.2 Stranded savings recommendations

NMEC is most effective when it targets customers with demonstrated barriers to upgrading aging, below-code equipment. Its treatment of below-code savings is justified when projects would not have occurred otherwise, whether barriers are technical or financial. By tying meter-based savings to tools such as on-bill financing, NMEC can directly address capital constraints while ensuring ratepayer support delivers genuine benefits.

[19] Fractional savings uncertainty (FSU) measures the uncertainty around program energy savings. As the value of FSU decreases, confidence in the estimated savings level increases.

The evaluated pop NMEC programs motivated a substantial share of customer savings, on average 70%, and were well received by participating customers. Overall, pop NMEC programs achieved NTGRs comparable to or higher than most non-residential program pathways, such as the Commercial, Industrial, Agricultural, and Custom pathway (CIAC) , with variation across programs suggesting that targeting hard-to-reach (HTR) customers is associated with higher influence. BayREN's Business FLEXmarket program, which focuses on HTR customers, achieved particularly high NTGRs, including 90% in one program year.

OBF survey results provide strong evidence that the program enabled or accelerated upgrades when capital barriers were central, but the program would provide more value if outcomes reflected performance-based results. OBF reaches customers who face barriers in accessing traditional financing, including municipal entities such as K-12 schools and civic facilities. Three-quarters of OBF survey respondents said they would not have been able to complete the project or do the same project at the same time without the program. That said, the program's current payback process is based on exaggerated preliminary savings estimates and undermines the program's ability to benefit more of these kinds of customers. Updating payback plans to reflect meter-based savings will fully address customer barriers while also incentivizing implementers and contractors to develop projects that deliver full value to customers.

Recommendations 5 & 6	Target customers with high stranded savings potential

- PAs should prioritize customers with demonstrated barriers (e.g., limited access to financing or constrained capital cycles), as NMEC uniquely credits "stranded savings" from to-code upgrades that are more common in these segments.
- PG&E should offer a plan that achieves program-level forecast realization rates (initial claimed savings relative to verified savings) greater than 90% by mid-September 2026. This approach will better align program incentives with realized performance, address the current misalignment between initial claimed savings and actual outcomes, and align payment schedules with actual bill-level savings.

Eighty-four percent of OBF and 77% of non-OBF interventions included only lighting upgrades, which do not fully support pop NMEC's focus on whole-building savings and delivering grid value. Despite pop NMEC's whole-building, measure-agnostic design being well suited to comprehensive retrofits, lighting upgrades dominated most programs, with multiple programs installing only lighting measures. Lighting-only projects accounted for the majority of savings (66% of OBF savings and 91% of non-OBF savings in the evaluated cohorts), while mixed-measure projects sometimes delivered higher site-level reductions (for OBF, projects with both lighting and non-lighting measures achieved higher percent reductions than lighting-only). Thus, most of the pop NMEC programs functioned more like lighting programs than comprehensive whole-building programs, which suggests that many pop NMEC programs are not fully realizing their intended role in capturing stranded savings through broader modernization.

Recommendation 7	Broader technologies for greater whole-building upgrades

PAs should look beyond lighting-only projects and increase the share of mixed-measure or comprehensive projects in pop NMEC, such as those including HVAC and refrigeration-related upgrades. Pop NMEC is designed for whole-building savings, and mixed-measure projects can deliver larger site-level reductions.

Customers deemed ineligible or with canceled projects were rarely directed to other savings pathways or programs. Eligibility processes were largely handled by implementers or aggregators, with PAs having only limited visibility into those processes. As a result, the PAs had little involvement in eligibility decisions or the disposition of customers who were not good candidates for pop NMEC programs. Generally, implementers and aggregators lacked motivation to find other opportunities for customers because they were enrolled in and incentivized by their specific pop NMEC programs.

Large customers and “non-commercial-like” sectors account for substantial savings, but they also increase pop NMEC risk and are often better suited to other pathways. Across cohorts, large sites (≥1,000,000 kWh) are a small share of participants but account for a disproportionately large share of savings. While these large sites may represent good energy efficiency program participants, they are not good fits for pop NMEC because the savings are too big and too risky for the pop NMEC environment, which uses consistent model structures across all projects. Industrial, agricultural, and utility-type customers are also better served by CIAC and strategic energy management (SEM) programs, while large commercial-like customers may be better aligned with site-level NMEC, SEM, or CIAC programs. Pop NMEC should be reserved for customers where the approach adds unique value that addresses stranded savings, rather than absorbing high-risk savings from customers that have better-fit alternatives.

Recommendation 8	Clear referral pathways for poor-fit projects
PAs should implement a consistent tracking and referral process so customers/projects screened out of pop NMEC (e.g., industrial/process loads, very large sites) are redirected to appropriate alternatives (e.g., SEM, CIAC, or site-level NMEC), and PAs should report those dispositions.	

2 INTRODUCTION

This report provides DNV's impact and process evaluation of non-residential population-based NMEC (pop NMEC) programs for program years (PYs) 2022 to 2024, performed on behalf of the California Public Utilities Commission (CPUC). Pop NMEC is a pay-for-performance (P4P) savings pathway with embedded processes for estimating performance, consistent with the CPUC's "Rulebook for Programs and Projects Based on Normalized Metered Energy Consumption" (NMEC rulebook).[20] Like site-level NMEC, the pop NMEC savings pathway uses statistical tools to estimate the impact of energy efficiency programs on energy consumption using pre- and post-intervention meter data. This contrasts with most other energy efficiency programs, which claim final savings based on the expected and predefined efficiency of an intervention.[21]

However, while both site-level NMEC and pop NMEC employ a performance-based program design and statistical models of metered energy consumption to measure energy efficiency savings, the pop pathway applies a uniform modeling approach across all participating customers and incorporates a comparison group to control for non-program effects. In addition, pop NMEC offers performance-based incentives to implementers, not customers.

2.1 Background

Pop NMEC is a relatively new and evolving approach in California's energy efficiency portfolio. The previous CPUC-commissioned evaluation of pop NMEC programs (residential PY2019 – 2021; nonresidential PY2020 – 2021) demonstrated that such programs can deliver measurable, verifiable savings, particularly in residential sectors.[22] That study also identified persistent challenges related to replicability, documentation gaps, heterogeneous customer populations, and the use of novel modeling approaches not yet widely vetted across the industry. These uncertainties underscore the need for ongoing evaluation to assess the suitability of pop NMEC methods and the effects of customer heterogeneity on estimated savings.

2.2 Evaluation objectives

We evaluated how effectively pop NMEC programs are delivering on their intended outcomes and identified opportunities to improve program and regulatory development. Specifically, we assessed the accuracy of claimed energy savings, analyzed how those savings are distributed across participating customers and programs, and evaluated whether current M&V practices and the CPUC's NMEC rulebook provide a sufficient foundation for successful program design and implementation. To guide this assessment, we focused on the following three primary objectives:

- **Evaluate the accuracy and attribution of programs' final claimed savings**. Given the embedded evaluation structure of pop NMEC programs, our evaluation sought to replicate and validate the existing evaluation frameworks and program savings, and we sought to determine the portion of those savings that the programs caused. To address this objective, we investigated the following research questions:
  - What was the energy savings impact of the programs?
  - What was the total system benefit (TSB) and bill impacts of the programs?
  - How much of the energy savings and TSB were attributable to the programs?

---

[20] CPUC, *Rulebook for Programs and Projects Based on Normalized Metered Energy Consumption, version 2.0*, January 7, 2020, https://www.cpuc.ca.gov/-/media/cpuc-website/files/legacyfiles/n/6442463694-nmec-rulebook2-0.pdf. DNV evaluated PY2022–PY2024 programs under this version of the NMEC rulebook, not version 2.1, which was released in 2025.

[21] Researched, vetted, and predictable savings for EE technologies and services with well-established properties are known as "deemed" or "calculated" savings. The savings from installing a high-efficiency lightbulb, for example, are well-established, and non-NMEC EE programs can use those known values to claim savings. This contrasts with custom EE technologies and services that require unique savings calculations and do not use predefined values.

[22] DNV, *Population-Based NMEC - Program Years 2019 – 2021*, December 22, 2023, https://www.calmac.org/%5C/publications/Final_Population-Based_NMEC_Program_Impact_Evaluation_Report_-_Program_Years_2019_to_2021_toCALMAC.pdf.

- **Assess the distribution of verified savings among participating customers, programs, and climate zones, as well as when those savings deliver the greatest system value.** We examined participation in pop NMEC programs and analyzed how program savings are distributed by program type, industry segment (organization type), customer size, and location (climate zone). Additionally, we investigated how well the programs deliver benefits during periods of high system cost and carbon intensity. To address this objective, we explored the following research questions:
  - Who participated in the pop NMEC programs?
  - How were the savings distributed among the customers and programs?
  - How well did the programs deliver system benefits during high-cost, high-carbon periods?
- **Identify whether program M&V plans and practices and NMEC rulebook protocols are sufficient to ensure successful pop NMEC implementation.** Since the first evaluation, pop NMEC programs have become more diverse in design and implementation, reflecting the NMEC rulebook's agnostic stance on methodology and the resulting range of program approaches. We examined how programs are adapting to improve savings reliability, whether eligibility processes are applied transparently and consistently with NMEC rulebook expectations, and whether implementer-developed methods adhere to best practices, appropriately characterize uncertainty, and support the integrity of the pop NMEC process. To address this objective, we investigated the following research questions:
  - How well are programs learning and adapting? Specifically, are they revising offering expectations and improving the accuracy of savings estimates based on implementation experience and available evaluation findings?
  - Are eligibility rules being applied in ways that are transparent and consistent with NMEC rulebook expectations? How and why are customers excluded or redirected from pop NMEC programs, and to what extent are these decisions justified? What guidance may be needed to ensure eligibility processes do not introduce risk or distort program outcomes?
  - How effectively do different pop NMEC methods account for non-program effects and produce reliable program savings estimates? Do implementer-developed methods reflect best practice and appropriately characterize uncertainty?

2.3 Evaluated programs

Between PY2022 and PY2024, the PAs claimed savings from eight distinct non-residential pop NMEC programs, including one public sector program and seven commercial programs.

BayREN, Marin Clean Energy (MCE), Peninsula Clean Energy (PCE), Sonoma Clean Power (SCP), and San Diego Gas & Electric (SDGE) operated pay-for-performance (P4P) programs with implementation structures similar to the CPUC's Market Access Programs (MAP-style). These programs relied on a marketplace platform to recruit and manage networks of incentivized contractors (aggregators), who bundled projects across multiple sites and managed customer relationships. BayREN's program targeted hard-to-reach (HTR) commercial customers, while MCE, PCE, SCP, and SDGE emphasized aggregator-led delivery to maximize system benefits.

Pacific Gas and Electric (PG&E) administered two non-MAP style programs: NetOne Commercial Efficiency, a commercial program, and Institutional Partnerships, a statewide program for public agencies. While PG&E took the lead, it implemented the statewide Institutional Partnerships program in collaboration with Southern California Edison (SCE), Southern California Gas (SCG), and SDGE. Both non-MAP style programs used a traditional program delivery model to recruit and enroll customers through direct trade ally partnerships. PG&E also administered the On-Bill Financing–Alternative Pathway (OBF) program. Unlike the other programs, OBF did not offer incentives; instead, it financed customers' energy efficiency projects through zero-interest loans repaid via their utility bills. Table 2-1 provides a summary of the programs.

Table 2-1. Pop NMEC programs evaluated in PY2022 – PY2024[23]

Primary sector	Program name and ID	PA	Claimed TSB	Claimed total expenditures	Program description
Commercial	Business FLEXmarket BAYREN06	BayREN	1,094,604	$3,294,133	Focuses on HTR businesses and offers pay-for-performance P4P incentives. It also offers no-cost technical assistance to participants.
	Commercial Efficiency Market MCE02d	MCE	3,907,484	$2,529,932	A price-based program that aims to be flexible in how participants can provide energy efficient solutions and TSB.
	FLEXmarket Program PCE01	PCE	1,121,352	$2,238,545	Pays aggregators for customer energy savings that support PCE's goals, enhance grid reliability, and encourage maximum system benefits.
	CORE Market Access Program SDGE4198	SDGE	445,675	$601,387	A P4P program that rewards aggregators for delivering peak demand reductions and energy savings that provide the greatest value to the grid and customers.
	NetOne Commercial Efficiency Program PGE_Com_003	PG&E	4,765,895	$2,813,107	Targets both large enterprise customers and smaller HTR customers to help them implement energy-saving technologies.
	On-Bill Financing (OBF) Alternative Pathway PGE_OBFAP, formerly PGE210911	PG&E	21,927,193	$2,556,923	Offers non-residential customers utility financing for all energy efficiency upgrades not eligible for other utility rebate and incentive programs.
	FLEXmarket Program SCP01	SCP	1,156,574	$2,331,358	Offers a flexible, performance-based path for aggregators to deliver cost-effective energy savings that meet customer needs, support SCP's goals, and improve grid reliability.
Public	Institutional Partnerships: DGS & DoC PGE_SW_IP_Gov SCE_SW_IP_Gov SCG_SW_IP_Gov SDGE_SW_IP_Gov	PG&E SCE SCG SDGE	736,271	$765,715	Offers support to California state agencies to facilitate energy efficiency improvements through technical assistance, integrated demand side management (IDSM), workforce development, and other support.

[23] Claimed TSB and expenditure values are for PY2023, except for NetOne Commercial Efficiency Program (PGE_Com_003) and CORE Market Access Program (SDGE4198), which reflect PY2024 values. For statewide public programs, the reported totals represent combined claims from PG&E, SCE, and SDGE, all of which had active programs in PY2023.

3 METHODOLOGY

This section outlines the methodology DNV used to evaluate the pop NMEC programs, including the data and methods used to evaluate final claimed savings, TSB, program attribution, participation, and savings timing and distribution.

3.1 Data collection

This section outlines the data sources used in the evaluation, describes the primary data collection methods, and provides information on the interviews conducted with PA staff.

3.1.1 Data sources

Table 3-1 summarizes the data used in the evaluation, which was drawn from both primary and secondary sources. We relied on this data to validate program savings, calculate NTGRs, characterize participating customers, assess the distribution and timing of savings, and conduct methodological assessments.

Table 3-1. Summary of pop NMEC PY2022 – PY2024 evaluation data sources

Data sources	Description	Purpose in evaluation
Program tracking data	PA claims filed with the CPUC that include program name and type, savings per claim, total claimed savings, etc.	Determine program information for evaluation
Detailed program participation data	Site-level information, including customer names, addresses, contact information, North American Industry Classification System (NAICS) code, claim IDs, installation dates, details on installed equipment, and claimed savings	Identify participating sites (including customer, meter, and premise IDs), fuel type, claimed savings per site, installation dates, contact information for customer surveys
PA M&V information	Methods, code, raw and intermediate data, site-level results from the PA M&V activity	Validate the energy savings claims of the programs
Customer Information System (CIS) data	Supplementary information on customer characteristics such as industry group, building type, zip code, and climate zone	Understand program participation patterns
Data Axle	Supplementary source of non-residential customer information, including size, expenses, location, and year established	Understand the characteristics of customers participating in the programs
Energy use data	Customer hourly electric and daily gas interval (AMI) data, billing data, and applicable bill rates	Validate claimed savings and quantify TSB and bill impacts
Granular load profile data	Average hourly electric and daily gas AMI data for non-participating customers, aggregated by program-defined cohorts such as solar status, industry, climate zone, and size	Control for the effect of exogenous factors on post-intervention energy consumption
Nonparticipant data	Energy use data (hourly electric and daily gas AMI) and site characteristics for matched and other non-participating customers	Used either as matched comparison groups to control for exogenous changes or to support methodological assessment
Weather data	Hourly temperature for the relevant analysis period and locations, and typical meteorological year (normal) weather data	Weather-normalize energy consumption
Telephone surveys	Surveys of participating customers and PA program staff	Determine program attribution and net savings; understand site characteristics

3.1.2 Primary data collection

This section describes the primary research methods used to evaluate the non-residential NMEC programs. It summarizes our approach to data collection, sample design, and survey implementation, including survey mode and disposition.

3.1.2.1 Data collection

Table 3-2 summarizes our primary data collection approach, including key details such as respondent group, sample frame source, and sample size.

Table 3-2. Summary of planned primary data collection effort for the evaluation of pop NMEC programs

Respondent group	Data collected	Frame source	Mode	Stratification approach	Sample size	Data collection timing
Utility program staff	Program design, implementation, customer acquisition and eligibility, M&V and other feedback influence, incentive payment structures	All PA program staff	In-depth interviews	Program	Census	December 2025 – January 2026
Non-residential participating customers	Program influence: NTG, customer characteristics, program awareness and experience	Detailed program participation data	In-depth interviews	Program, year	n = 152	November 2025 – February 2026

3.1.2.2 Sample design

The population, or sampling frame, for the evaluation consisted of all installations claimed by non-residential pop NMEC programs from PY2022 to PY2024, as provided by the PAs. We used these installations for primary data collection to support analyses of program attribution, customer characteristics, and customer experience.

Table 3-3 summarizes by program and year the number of customers covered by this evaluation. The customer counts shown in the table represent the populations from which we developed survey sample targets.

Table 3-3. Pop NMEC PY2022 – PY2024 survey types and number of customers

Program style	PA	Program	PY	Population
MAP-style	**BayREN**	**Business FLEXmarket**	**2023**	65
			2024	218
	PCE	**FLEXmarket Program**	**2023**	4
			2024	7
	MCE	**Commercial Efficiency Market**	**2022**	9
			2023	52
			2024	10
	SDGE	**CORE Market Access Program**	**2024**	3
Total				**368**
Non-MAP Style	**PG&E**	**Institutional Partnerships**	**2023**	7
			2024	6
		NetOne Commercial Efficiency	**2024**	37
Total				**50**
OBF	**PG&E**	**OBF**	**2022**	385
			2023	278
			2024	232
Total				**895**

For primary data collection, we used a census approach for SDGE and stratified sampling for all other programs. Under the census approach for SDGE, we attempted to collect data for all individuals in the population with contact information. Under the stratified sampling approach for the remaining programs, we selected sample units from program and program year, and we stratified by savings measured in kWh only, since PAs did not provide customer-level projected savings for the other claimed fuels. We aimed to complete 152 interviews to achieve a relative precision of ±10 at the 90% confidence level for each program.

After determining the required sample size for each program to achieve the target precision at the specified confidence level, we randomly selected primary sample points from the population in accordance with the stratification plan. We also selected a backup sample to replace any primary sample points we could not contact or reach. APPENDIX E provides detailed sampling plans. The primary objective of the surveys was to develop estimates of free-ridership. The survey instruments and interview guides used for data collection are included in APPENDIX H.

3.1.2.3 Survey mode and disposition

We delivered non-residential program customer surveys to site managers—the primary decision-makers for the installations—using contact information provided by the PAs. We conducted phone surveys over approximately 12 weeks, from November 20, 2025, through February 6, 2026. We supplemented missing contact information with online research and calls to publicly available business numbers to ask for appropriate contact information. We reached out to each contact up to six times, at different times of day, using phone and email.

Table 3-4 summarizes the disposition of the survey sample, including the number of sites we targeted, numbers of completed surveys, and the associated response rates. The high response rates reflect the proven best practices we adopted in fielding the survey, including:

- Providing respondents with a link to validate the legitimacy of the survey effort
- Co-branding the survey with the CPUC and PA logos
- Issuing the survey invitation from an email address with the PA domain, reducing the likelihood of being screened by respondents' spam filters
- Providing a letter from the CPUC study manager on the importance of this research and customer responses to energy efficiency programs

Table 3-4. Pop NMEC PY2022 – PY2024 survey efforts and sample size summary

Program style	PA	Program	Population	Sample target	Completed	Response rate
MAP-style	BayREN	Business FLEXmarket	283	30	31	103%
	PCE	FLEXmarket Program	11	9	6	67%
	MCE	Commercial Efficiency Market	71	25	20	80%
	SDGE	CORE Market Access Program	3	3	2	67%
Non-MAP Style	PG&E	Institutional Partnerships	13	9	8	89%
		NetOne Commercial Efficiency	37	16	15	94%
OBF	PG&E	On-Bill Financing (OBF)	895	60	80	133%
Total			**1,313**	**152**	**162**	**107%**

3.1.3 Program administrator interviews

To inform our evaluation of CPUC Group D PY2022 – PY2024 pop NMEC programs, we conducted structured, in-depth interviews (IDIs) with PA program staff responsible for the programs and program years covered in this evaluation. The IDIs covered six major areas: program design and implementation, customer recruitment and eligibility processes, incentive structures and payment management, savings traceability and M&V practices, the influence of program experience and available evaluation findings, and the role of the TSB framework on program operations. The IDI instrument can be found in APPENDIX H.

Interviews began with background questions on the PA's role and organizational structure, followed by detailed exploration of program theory, delivery models, the roles of implementers and aggregators, QA/QC practices, customer outreach methods, and eligibility screening procedures. Additional questions examined how incentives flow through the program (including documentation and audit practices), how savings models and customer lists are maintained and validated, and how internal M&V results shape program adjustments. We also asked interviewees about how previous evaluation findings guide program design, information-sharing across PAs, and their understanding of the TSB framework and its anticipated impacts on claims and incentive structures. Each interview lasted approximately one hour and followed a standardized protocol to ensure consistency across PAs.

DNV completed interviews with each of the PAs between December 15, 2025 and January 9, 2026.

3.2 Savings impacts

We classified and evaluated pop NMEC programs based on the availability of true-up claims. This classification is summarized in Table 3-5 and described in detail below.

- **Full gross evaluation of programs with trued-up claims.** We conducted a full gross savings evaluation for eight programs from PY2022 and PY2023 that had both initial and true-up claims. This group includes six unique programs as two had claims in both PY2022 and PY2023.
- **Early gross evaluation of programs without trued-up claims.** For programs with initial savings claims but no true-up savings claims, specifically those with PY2024 claims, we conducted an early gross evaluation. We reviewed the documented evaluation framework, including the proposed M&V approach, project eligibility criteria, and key assumptions used to estimate savings TSB.
- **Net evaluation for all programs.** We estimated net savings for all PY2022 – PY2024 pop NMEC programs using NTG surveys. Because CEDARS does not include customer-level data, including contact information, DNV requested this information from the PAs to support survey administration under a targeted sampling plan. Our goal for these programs was to achieve sufficient sample sizes to estimate program-level NTGRs with a relative precision of ±10% at the 90% confidence level.

Table 3-5. Summary of pop NMEC gross evaluation programs

Initial claim year	Program counts			PAs included
	Total	Full gross evaluation	Early gross evaluation	
PY2022	2	2	0	PG&E, MCE
PY2023	6	6	0	BAYREN, MCE, PCE, PG&E, SCP
PY2024	7	0	7	BAYREN, MCE, PCE, PG&E, SDGE
All Years	15	8	7	BAYREN, MCE, PCE, PG&E, SDGE, SCP

3.2.1 Gross savings evaluation

This report refers to savings impacts that do not account for attribution as "gross savings." Because pop NMEC programs are meter-based and include embedded M&V, our gross evaluation consisted of two phases. In the first phase, we replicated the embedded M&V results. In the second phase, we validated the PAs' M&V approaches for estimating program savings. After validating program savings, we calculated savings realization rates, measuring how closely evaluated savings align with final claimed savings. A realization rate of 100% indicates that final claimed and evaluated savings are equal.

Pop NMEC programs make initial savings claims based on engineering or deemed savings estimates, which are trued-up to meter-based savings results after a 12-month performance period. While this evaluation primarily focused on validating trued-up programs, we also performed an early gross evaluation of programs with initial claimed savings. The full gross savings evaluation of trued-up programs examined the credibility of final claimed savings, while the early gross savings evaluations of not-yet-trued-up programs examined the soundness of the PA M&V methods.

3.2.1.1 Full gross savings evaluation

For programs with true-up savings, we took a two-part validation approach.

Replication. First, we replicated the PAs' savings estimates to evaluate the:

- Transparency and replicability of the savings estimation process, as required by the NMEC rulebook
- Consistency between replicated results and those documented in the program M&V processes, which form the basis for true-up savings claims

This step required access to all relevant data, model specifications, and code. While open-source, push-button replication is ideal, the CPUC does not currently mandate it. In practice, diverse data sources and custom-developed code can make replication challenging, and we assessed whether the level of replicability met basic validation standards for each program. Any deviation from full replicability increases the evaluation burden and may introduce uncertainty into the validation process.

Validation. In the second step, we conducted a deeper assessment of the program M&V process itself, including the evaluation of savings precision. In particular, we assessed:

- Compliance with NMEC rulebook requirements, including:
  - Customer eligibility and screening practices
  - Customer attrition and exclusion approaches
  - Modeling approaches and associated documentation
  - Normalization of pre-installation (baseline) and post-installation (performance) period results
- Reasonableness of methods and results, determining whether the approaches are appropriate and the outcomes credible within the program context

As part of our check on the reasonableness of methods and results, we evaluated whether final claimed savings complied with the NMEC rulebook requirement of achieving 25% relative precision at the 90% confidence level. We assessed how PAs calculate program-level precision (savings uncertainty) and whether their methodologies reflect appropriate practices for program savings evaluation, such as established approaches used in billing analysis.

We also analyzed the distribution of evaluated precision levels across programs and program types to identify patterns and assess the implications of applying pop NMEC methods in varied contexts. Where precision consistently fell short of the 90/25 threshold, we considered alternative standards, such as a 90/50 rule, based on typical performance observed in billing analysis.

3.2.1.2 Early gross evaluation

For pop NMEC cohorts (programs in a program year) that had not yet completed the true-up process, we reviewed available documentation but did not validate claimed savings. At the start of the evaluation, these cohorts were still in their performance period and lacked the full year of post-installation data required for performance-based M&V and true-up calculations.

For these programs, our early gross evaluation assessed:

- Project eligibility criteria and plans to document customer attrition
- The types of projects or sites included in the program
- The methodological soundness of the proposed M&V modeling approach
- The approach to normalizing energy consumption
- Plans for determining savings for sites excluded from modeling
- The assumptions used to calculate TSB, including reliance on generalized or empirically derived savings shapes

3.2.2 Distribution of TSB and bill savings

TSB, adopted by the CPUC as the primary metric for valuing energy efficiency starting in PY2024, reflects the time-varying value of energy savings, assigning a higher monetary value to savings occurring during peak demand periods, particularly late afternoon and early evening hours when grid resources are more carbon-intensive and costly.

TSB values are determined using six inputs: hourly savings shapes, avoided cost values, expected useful life (EUL) of program measures, energy savings, gross realization rate (GRR), and program attribution values. PAs enter these inputs into the CPUC Cost Effectiveness Tool (CET) to calculate the TSB values they claim annually. We evaluated the claimed TSB values by examining and updating four of the six inputs: avoided cost values, GRR, energy savings, and program attribution. We entered the verified values for these inputs into an internal version of the CET and compared the resulting TSB values with the claimed values to assess how well the programs delivered the claimed benefits. While we did not update the claimed savings shapes for the TSB analysis because PAs did not base their TSB claims on empirical shapes, we compared verified savings shapes with program-specific claimed savings shapes.

In parallel, we calculated program-level bill impacts by multiplying the hourly electricity savings by applicable hourly electricity rates, including energy- and demand-related rates. Because only one program (OBF) claimed gas savings for a limited number of sites, we did not develop gas-related bill impacts.

3.2.3 Net savings methodology

We conducted NTG surveys or interviews with pop NMEC program customers to derive program-level NTG ratios, which are applied to gross savings to derive net savings. We followed the same methodology used for site-level NMEC, in alignment with the CA non-residential NTG protocols.[24] Based on this approach, we interviewed customers to collect information on program influence and used the responses to estimate NTGRs. Unlike residential NTGR methods, which typically focus only on counterfactual scenarios, the non-residential approach incorporates both counterfactual responses and indicators of influence, including program-related factors such as financial incentives, information, and program support, as well as non-program factors such as organizational goals and replacement schedules. We scored each component using standardized algorithms and averaged the scores to estimate free-ridership, calculating NTGRs as one minus free-ridership. We applied the resulting NTGRs to gross savings to calculate net savings for each program.

3.3 Participating customer profiles and system value

To address Research Objective 2, we evaluated the distribution of verified savings by program type, industry segment (organization type), customer size, and location (climate zone), and assessed how those savings contribute to system value. We used program tracking data, customer information system (CIS) files, third-party firmographic data, and survey

[24] California Non-residential Net-to-Gross Working Group, *Methodological Framework for Using the Self-Report Approach to Estimating Net-to-Gross Ratios for Non-residential Customers*, (CPUC), February 20, 2015, https://pda.energydataweb.com/api/view/1255/IALC_NTGR_Methods_2013-14.pdf.

responses to characterize program participation. We combined these data sources with verified savings to assess how savings are distributed and when they provide the greatest value to the grid.

Customer characterization. To understand the value that the programs provide to customers, we developed firmographic profiles for the commercial and public sector entities participating in the programs. To develop these profiles, we used survey data, tracking data, CIS files, and third-party data from Data Axle to identify customer attributes, including business size and sector.

We classified each site into a general industry category based on reported NAICS codes. Some uncertainty remains because reported NAICS codes may be inaccurate or outdated, and customer size can vary substantially within the same industry group (for example, commercial and retail). To support consistent and effective analysis despite these limitations, we categorized each facility into one of the three general organization type or industry segments listed below:

Commercial and retail

- **Commercial & retail:** Customer-facing facilities, including retail stores, grocery stores, restaurants and other food and beverage establishments, personal care services, and gas stations, for which energy consumption is primarily driven by operating hours and lighting use.
- **Office & information services:** Primarily office-based environments where energy consumption is typically driven by HVAC systems, lighting, and IT equipment.
- **Hospitality & recreational services:** Facilities such as hotels, parks, and businesses that provide leisure and recreational experiences for customers.
- **Distribution:** Medium storage and logistics facilities, including warehouses and wholesale centers, with substantial HVAC and lighting loads.

Institutions

- **Institutional & public services:** Facilities that provide government and educational services, as well as those that house religious and civic organizations.
- **Health care facilities:** Sites that deliver health services and operate continuously, with high HVAC and ventilation demands.

Industrial, agricultural, and utility

- **Industrial & production:** Agriculture, construction, and manufacturing facilities, typically with energy-intensive processes, heavy machinery, ventilation loads, and continuous or heavy-duty operations.
- **Infrastructure & utilities:** Essential backbone systems, such as transportation, communications, and utilities.

DNV characterized customer location by grouping California's 16 climate zones into the following five broader categories:

- **Coastal/Marine:** Climate zones 1, 2, 3, 5, and 7
- **Inland Basin:** Climate zones 4, 8, 9, and 10
- **Central Valley:** Climate zones 11, 12, and 13
- **Desert:** Climate zones 14 and 15
- **Mountain:** Climate zone 16

Savings distribution. We also analyzed how savings are distributed across participating customers and programs. This analysis involved linking verified customer-level savings to program tracking data. We quantified savings by customer characteristics, including industry segment (organization type), customer size (using energy consumption as a proxy), program type, and climate zone.

Grid value. We also evaluated how well programs deliver system benefits during periods of high cost and high carbon emissions. We based this analysis on time-differentiated savings estimates, including 365-daily and 24-hourly savings shapes derived from each program's evaluated 8760 hourly savings shapes. In particular, we assessed the extent to which savings occurred during summer months (June through September) and late-afternoon hours (4 PM – 7 PM), which overlap with DEER-defined peak periods. We examined the results by climate zone group to assess each program's contribution to system value.

Table 3-6 summarizes the research questions, data sources, and analytical methods for this evaluation.

Table 3-6. Research questions, data sources, and analytical methods

Research question	Data sources	Analytical methods
Who is participating in pop NMEC programs?	Survey data, customer rosters/tracking data, CIS, third-party data (Data Axle)	Firmographic profiling
How are the savings distributed among participating customers and programs?	Verified program savings, program tracking data, climate zone mapping	Savings disaggregation by customer and program type, and region
How well do the programs deliver system benefits during high-cost, high-carbon periods?	Hourly impact results, program tracking data, climate zone mapping, DEER-defined peak periods	Time-based savings analysis by region / climate zone

3.4 Program design and methodology review

Pop NMEC programs have expanded and diversified since their initial launch. To address Research Objective 3, we examined how programs are adapting based on lessons and available evaluation feedback. We also assessed PA customer eligibility screening and whether ineligible customers were appropriately redirected, such as to site-level NMEC. Additionally, we evaluated how well embedded M&V methods are accounting for non-program effects to produce reliable savings estimates.

3.4.1 Program experience and feedback loop

To assess program learning and adaptation, we examined whether and how PAs are using program lessons and results, including those from available evaluations, to improve program design, forecasting, and incentive structures. This effort focused on examining program performance, planning, and forecasting, with particular attention to how well pop NMEC programs are delivering their promised outcomes and evolving in response to program experience and findings.

To support this aim, we carried out the following activities:

We assessed program design and logic. Through reviews of program documents and interviews with PAs, we assessed whether program activities reflect a clear and coherent plan linking program design choices to expected outcomes that are consistent with NMEC policy intent. Specifically, we evaluated whether the theory of change reflected in program implementation and observed outcomes are consistent with those articulated in NMEC policy.

We evaluated program performance. We analyzed complementary realization rates to provide insights into program planning and delivery performance. Specifically, we examined:

- Gross realization rate (GRR), a comparison of verified to final claimed savings, to assess the effectiveness of program implementation and delivery

- Forecast realization rate (FRR), a comparison of verified to initial claimed savings, to assess the reliability of planning assumptions and the extent to which programs delivered on their initial savings promise

We examined these metrics by program type and year to identify trends and assess whether planning assumptions are reliable—and improving over time for programs with evaluation history—and whether customers are ultimately receiving the benefits initially promised.

We analyzed use of program experience and available evaluation feedback. We examined how programs use their experience, including available evaluation results, to improve program delivery. This analysis drew on reviews of program documentation, PA data, and available evaluation results, as well as interviews with PA staff to understand internal processes and decision-making. We focused on the extent to which programs are achieving their goals and are learning and adapting over time, as indicators of long-term effectiveness and ratepayer value. Specifically, we examined how program lessons and findings are incorporated into implementer compensation and program planning by focusing on:

- How discrepancies are handled and whether evaluation results are used to inform future incentive design
- Whether initial claims are improving based on available evaluated savings
- Whether program planning and design reflect program lessons and available evaluation findings

3.4.2 Customer eligibility assessment

The NMEC rulebook requires pop NMEC programs to define customer eligibility criteria in their M&V plans to ensure enrollment of customers suitable for pop NMEC participation, specifically, customers with stranded savings potential that can be addressed through the program and reliably measured using meter-based approaches. To evaluate whether the PY2022 – PY2024 non-residential pop NMEC programs included in this study employed appropriate eligibility criteria and practices, we examined program eligibility criteria, PA project screening practices, and how PAs handled potential projects deemed ineligible or otherwise not enrolled. Through PA interviews and program documentation review, we assessed program processes for tracking customer interactions, determining eligibility, and managing ineligible projects. We used this information to assess compliance with NMEC rulebook requirements for enrolling customers suitable for pop NMEC programs, and we also considered whether eligibility-handling practices indicate the potential underutilization of alternative participation pathways, such as site-level NMEC or custom programs, when pop NMEC remains available.

3.4.3 Methodological assessment

We reviewed and evaluated the pop NMEC methodologies that PAs and their M&V providers used to develop meter-based savings estimates, with a focus on how methods were structured and implemented. We aimed to document methodological variation, evaluate alignment with NMEC principles and billing analysis foundations, and identify implications for consistency and uncertainty in pop NMEC results.

We reviewed evaluation documents, model specifications, and savings calculation procedures that the pop NMEC programs used to generate final claimed savings for this assessment. We examined the methods across key stages of the pop NMEC process, including site-level modeling, construction and application of comparison groups, savings estimation, development of hourly impact shapes, and aggregation and uncertainty estimation.

- **Review of site-level modeling approaches.** We reviewed site-level consumption models supporting pop NMEC meter-based results, focusing on their role in weather normalization, baseline construction, and eligibility screening. We documented model structures (e.g., independent variables, temporal flexibility, use of granular profiles) and how programs use model outputs such as fractional savings uncertainty (FSU). We assessed consistency with NMEC steady-state assumptions, treatment of non-routine events (NREs), and implications for eligibility and downstream savings estimation.

- **Assessment of comparison group construction and use.** We reviewed how programs define and incorporate comparison groups, including matched non-participating customers and granular profiles. We focused on how programs use comparison groups to adjust for exogenous effects, consistent with NMEC rulebook guidance, and assessed how adjustment design choices affect comparability and consistency across customer populations.
- **Review of savings estimation and weather basis.** We reviewed how programs calculate site-level impacts and aggregate them to program-level savings estimates, including the use of avoided AEU approaches. We examined how programs treat weather in savings calculations (using actual performance-period weather versus typical meteorological year [TMY] normalization) and considered implications for interpretation and downstream applications.
- **Evaluation of hourly impact shapes and peak savings.** We reviewed how programs derive hourly (8760) impact shapes and peak demand savings, including baseline shape construction, application of comparison group adjustments, and incorporation of performance-period variability. We assessed alignment with NMEC steady-state assumptions and suitability for grid value and TSB analyses.
- **Aggregation and treatment of uncertainty.** We reviewed how programs aggregate site-level results and calculate and report uncertainty, including treatment of comparison group adjustments and propagation of standard errors. We assessed consistency with traditional billing analysis approaches to error accumulation and identified potential sources of inconsistency in reported precision.
- **Synthesis of methodological implications.** We synthesized findings across components to identify where methodological variation may introduce inconsistency, increase uncertainty, or complicate interpretation of pop NMEC results. We do not rank or select preferred methods in this report; instead, we describe current practices and their implications for the reliability and transparency of pop NMEC performance estimates.

4 RESULTS

This section presents evaluation results by the three program designs found among the pop NMEC non-residential programs. BayREN, MCE, PCE, SCP, and SDGE's programs are similar in design to the CPUC's Market Access Program (**MAP-style**). They implement projects through a marketplace platform, recruiting and managing a network of incentivized contractors (aggregators) that bundle projects across multiple sites and manage customer relationships. PG&E NetOne and Institutional Partnerships are **non-MAP-style** programs that rely on direct trade ally partnerships to recruit and enroll customers. The On-Bill Financing–Alternative Pathway (**OBF**) program is PA-administered and, uniquely, provides zero-interest financing instead of incentives, repaid through utility bills.

4.1 Gross savings and realization rates

We present gross savings evaluation results by fuel type (electricity, demand, and gas) for projects with initial claims in PY2022 and PY2023 in this section. Results for PY2024 cohorts are not included as final meter-based savings were unavailable. For PY2022 and PY2023 projects, the evaluation focused on verifying claimed savings using PA-provided data, models, and documentation, given that pop NMEC programs embed M&V savings calculations within program implementation.[25]

Each subsection below compares final claimed savings (claimed) with evaluated savings (verified), as reflected in the gross realization rate (GRR). A GRR of 100% indicates perfect alignment between claimed and verified savings. Each subsection also discusses discrepancies between these values and the factors driving those differences, and presents the relative precision of the verified savings estimates, which reflects the associated uncertainty.

We also used verified savings to estimate bill impacts. The results of that analysis are presented in APPENDIX C.

4.1.1 Electric savings

Table 4-1 presents the gross electricity realization rates and savings results by program and year (cohort), categorized by program style, for cohorts with initial claims in PY2022 and PY2023. SCP did not provide the FLEXmarket models or supporting data needed to replicate and validate its savings claims. As a result, SCP has a gross realization rate of 0% and contributes zero verified savings to the GRR results for its program style. For the remaining electric cohorts, gross realization rates range from 65% to 162%.

Only the PY2023 OBF electric savings met the NMEC rulebook uncertainty benchmark, with a relative precision value below the 90/25 threshold. This benchmark requires programs to be designed to achieve a margin of error no greater than 25% at the 90% confidence level.[26] The other programs did not meet this standard, indicating a higher but modest risk that the verified savings do not reflect the true savings. We address the reasonableness of relaxing this benchmark in Section 4.2.7. The remainder of this section explores the key contributors to these first-year GRRs.

Table 4-1. Gross electricity (kWh) realization rates

Style	PA	Program	PY	Claimed	Verified	GRR	RP%
MAP-style	**BayREN**	**Business FLEXmarket**	**2023**	2,024,632	1,587,591	78%	±93.0%
	PCE	**FLEXmarket Program**	**2023**	1,393,323	1,208,136	87%	±53.0%
	MCE	**Commercial Efficiency Market**	**2022**	852,728	1,382,486	162%	±40.0%
			2023	4,048,775	3,198,083	79%	±49.0%
	SCP	**SCP FLEXmarket**	**2023**	1,014,304	0	0%	N/A

[25] Based on the program materials that the PAs provided for MAP-style programs, DNV relied on a newer version of the model than the one the PAs and implementers used to claim savings. Therefore, findings in this report reflect the newer model's performance, and findings and takeaways account for this nuance.

[26] This benchmark is not a requirement for the acceptance of the final program impacts.

Style	PA	Program	PY	Claimed	Verified	GRR	RP%
Total				**9,333,763**	**7,376,297**	**79%**	
Non-MAP Style	**PG&E**	**Institutional Partnerships**	**2023**	743,319	486,840	65%	±33.0%
Total				**743,319**	**486,840**	**65%**	
OBF	**PG&E**	**On-Bill Financing (OBF)**	**2022**	11,706,659	11,282,202	96%	±42.0%
			2023	6,135,779	5,216,773	85%	±21.0%
Total				**17,842,438**	**16,498,975**	**92%**	

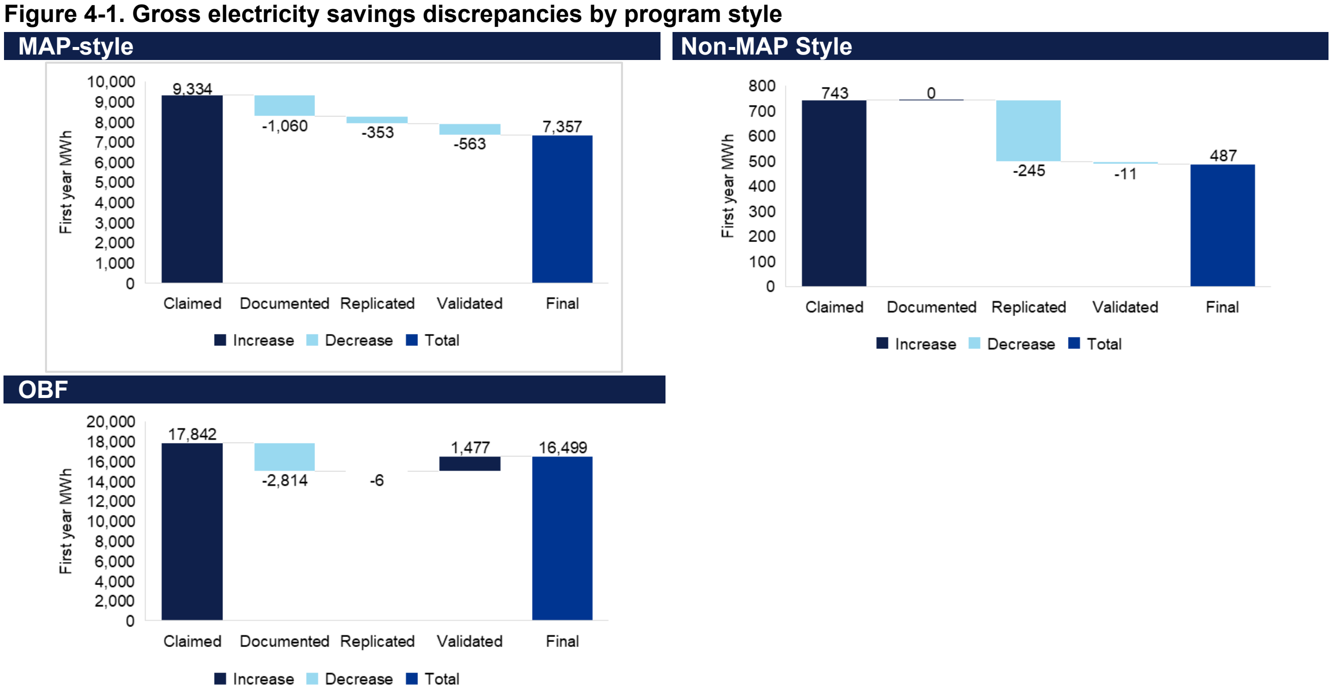
We grouped the savings discrepancies underlying the evaluation's GRR values into three high-level factors: documentation, replication, and validation. Table 4-2 summarizes these discrepancy types.

Table 4-2. Savings discrepancy factors

Discrepancy factor	Description
Documentation	Differences attributed to incorrect adjustments or unexplained changes to savings that occurred between completion of the PA analysis and entry into the tracking system.
Replication	The difference between the savings replicated using the provided code and project data and those documented in the final reports or output files.
Validation	Changes in savings due to evaluation improvements, including eligibility screening and M&V exclusion corrections, code fixes, or other methods adjustments.

Figure 4-1 shows the electricity savings impacts of each discrepancy type by program style. The first column shows total claimed savings and the last column shows total verified savings, with the middle columns reflecting the adjustments at each stage of the evaluation. Documentation adjustments were the largest source of discrepancies for OBF and the MAP-style programs, while replication adjustments were the largest for the much smaller non-MAP Institutional Partnerships program. Note that the Y-axes use different scales, reflecting substantial size differences among program styles.

Figure 4-1. Gross electricity savings discrepancies by program style



Note: Figures have different scales due to substantial differences in program size.

This section provides a high-level discussion of discrepancies. Detailed evaluation findings can be found in Section 4.2.

DNV

Documentation. The lack of models and supporting data from SCP to evaluate program savings claims was a major source of documentation discrepancies for the MAP-style programs. Even so, none of the other MAP-style programs' documentation matched the tracking claims exactly. In contrast, the non-MAP Institutional Partnerships program's documentation matched the claims in the tracking data. For OBF, the PY2022 documentation showed substantially lower savings than claimed.

Replication. The MAP-style programs, except SCP, and the non-MAP Institutional Partnerships program provided the necessary datasets and referenced open-source modeling methodologies: OpenDSM for the former and OpenEEmeter for the latter.[27] While the data and reference methodologies supported our replication efforts, we were unable to replicate the savings for these programs exactly. For MAP-style programs, differences between replicated and documented values likely reflect differences in model vintage between those used to generate claimed savings and those used in the evaluation. Despite this limitation, replication outcomes were closely aligned with documented values. Across MAP-style programs, replicated savings were within approximately 5% of documented savings (and within approximately 10% for PCE's cohort), indicating that results are not highly sensitive to the model form used in replication at the aggregate level. Conversely, we were able to replicate savings for the two OBF cohorts almost exactly. For the non-MAP Institutional Partnerships program, replication for some sites produced results that differed substantially from those documented, indicating that the provided data could not have generated the documented results.

Validation. Most validation discrepancies stemmed from inconsistent execution of M&V exclusion and eligibility screening criteria, and in some cases, coding errors that, once corrected, resulted in different savings values than those documented. A much smaller share of discrepancies resulted from a savings reasonableness check applied by DNV, which limited non-performance-based (assigned) savings to no more than 50% of baseline consumption.[28]

4.1.2 Demand savings

Table 4-3 presents the gross demand realization rates and savings results by program and year (cohort), with initial claims in PY2022 and PY2023, categorized by program style.

SCP did not provide the FLEXmarket models or supporting data needed to replicate and validate its savings claims, so evaluated savings are not available for that cohort. Further, the MCE 2022 cohort claimed zero demand savings, so, despite positive evaluated savings, a GRR is not available for that cohort. For the remaining cohorts, gross realization rates range from 0% to 92%.

Only the PY2023 OBF and PY2023 BayREN savings met the NMEC rulebook uncertainty benchmark, with relative precision values below the 90/25 threshold. The remainder of this section explores the key contributors to these GRRs.

Table 4-3. Gross peak demand (kW) realization rates

Style	PA	Program	PY	Claimed	Verified	GRR	RP%
MAP-style	BayREN	Business FLEXmarket	2023	280	221	79%	±14.0%
	PCE	FLEXmarket Program	2023	159	36	22%	±144.0%
	MCE	Commercial Efficiency Market	2022	0	304	N/A	±42.0%
			2023	465	15	3%	±363.0%
	SCP	SCP FLEXmarket	2023	116	0.0	0%	N/A
Total				**1,020**	**576**	**56%**	
Non-MAP Style	**PG&E**	**Institutional Partnerships**	**2023**	128	48	38%	±149.0%

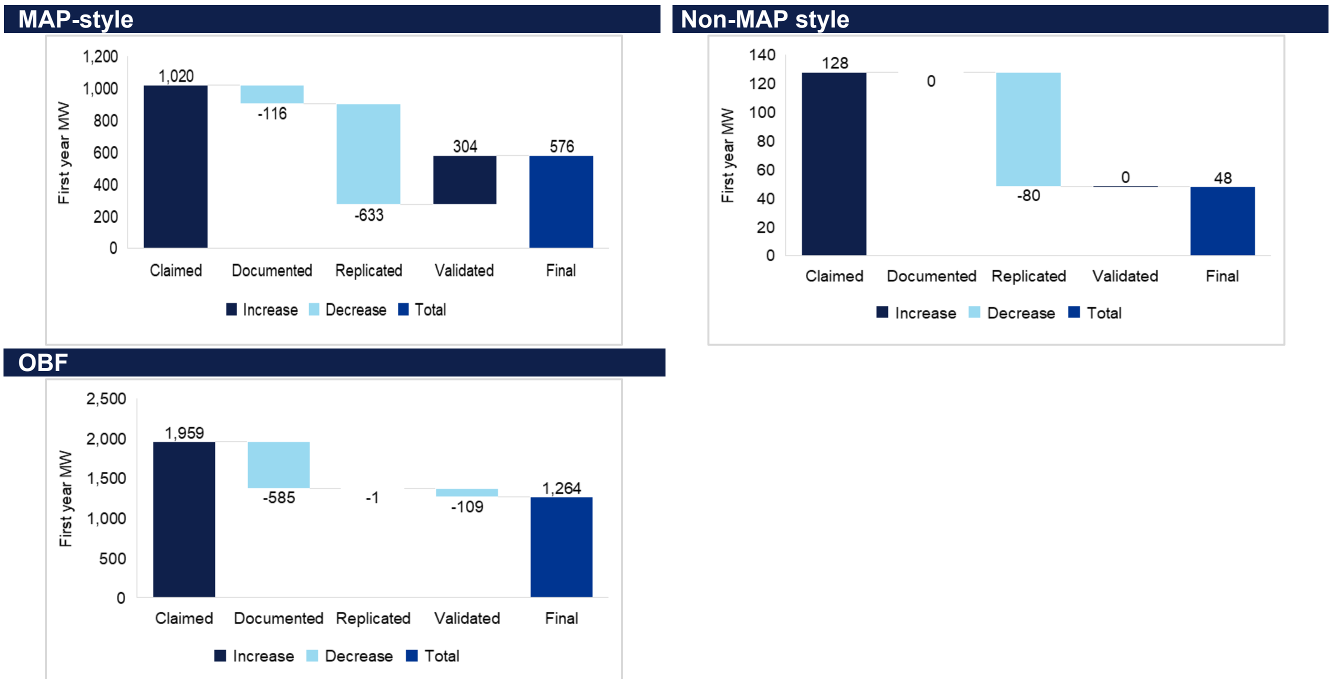
27 https://opendsm.energy/

28 Assigned savings are deemed or engineering-based estimates applied to sites for which meter-based savings cannot be developed. We discuss these in Section 4.2.5.

Style	PA	Program	PY	Claimed	Verified	GRR	RP%
Total				**128**	**48**	**38%**	
OBF	**PG&E**	**On-Bill Financing (OBF)**	**2022**	1,395	746	53%	±76.0%
			2023	564	518	92%	±22.0%
Total				**1,959**	**1,264**	**65%**	

Figure 4-2 shows the savings impacts of each type of discrepancy for each program style. Replication adjustments were the largest source of discrepancies for the MAP-style programs and non-MAP Institutional Partnerships program, while documentation adjustments were the largest source of discrepancies for the OBF program.

Figure 4-2. Gross peak demand savings discrepancies by program style



Note: Figures have different scales due to substantial differences in program size.

This section provides a high-level discussion of discrepancies. Detailed evaluation findings can be found in Section 4.2.

Documentation. The MAP-style programs did not provide any documented demand savings or methods. The non-MAP Institutional Partnerships program also did not report demand savings, but it did include limited documentation describing how peak impacts were calculated. In contrast, OBF provided both demand impact estimates and detailed code documenting the calculation method. For this program, the PY2022 documentation showed substantially lower savings than claimed.

Replication. Because there were no documented demand results to replicate for the MAP-style programs or the non-MAP Institutional Partnerships program, we relied on information from the PA M&V plans and the methodology used by the Institutional Partnerships program vendor to estimate demand savings for those programs. Using the temperature data the PAs provided, we identified DEER-defined heat-wave periods and estimated peak demand savings during those periods for these programs.[29] On the other hand, we were able to replicate the OBF demand impact estimates fully.

[29] The MAP-style program M&V plans mention DEER-defined peak periods; however, they do not clearly document whether the impacts reflect actual or typical meteorological year (TMY) weather, how heat-wave periods are defined, or the methods used to generate peak savings estimates.

Validation. For OBF, validation discrepancies came from inconsistent execution of eligibility screening and M&V exclusion criteria. For MAP-style programs, the main validation discrepancy came from the MCE PY2022 cohort, for which there were zero claimed demand savings but 304 MW of verified demand savings.

4.1.3 Gas savings

Table 4-4 presents the gross gas realization rates and savings results by program and year (cohort), categorized by program style, for cohorts with initial claims in PY2022 and PY2023.

Only the OBF program claimed gas savings. The GRR values were 0% for the PY2022 cohort and 312% for the PY2023 cohort. Quantifying gas savings with NMEC methods presents known challenges, such as difficulty fitting gas consumption data, but those challenges do not fully explain the problematic gas results in this evaluation.

Neither cohort's savings met the NMEC rulebook uncertainty benchmark, with relative precision values below the 90/25 threshold.

Table 4-4. Gross gas (therms) realization rates

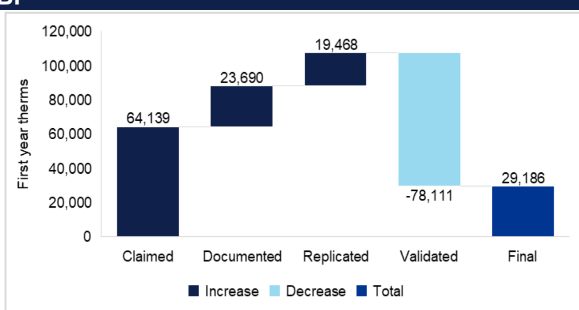
Style	PA	Program	PY	Claimed	Verified	GRR	RP%
OBF	PG&E	On-Bill Financing (OBF)	2022	54,792	0	0%	N/A
			2023	9,347	29,186	312%	±91.0%
Total				**64,139**	**29,186**	**46%**	

Figure 4-3 shows the savings impacts of each type of discrepancy for the combined OBF cohorts.

Validation adjustments were the largest source of discrepancies for the OBF gas program.

Figure 4-3. Gross gas savings discrepancies by program style

OBF



This section provides a high-level discussion of discrepancies. Detailed evaluation findings can be found in Section 4.2.

Documentation. The only documentation discrepancies came from the PY2022 cohort, which had documented savings that were 43% higher than claimed savings. PY2023 documented savings were aligned with claimed savings.

Replication. The replication discrepancies were different across the two cohorts: PY2022 replicated savings were about 10% below documented savings, while PY2023 savings were 290% above documented savings. These differences resulted from an error in the savings calculation code used to produce the documented values, which is discussed in Section 4.2.2. Because the claimed savings for PY2022 were six times greater than the PY2023 claimed savings, the two cohorts' combined discrepancies led to verified savings 30% above claimed savings.

Validation. In PY2022, all sites with meter-based results showed no savings. Documented cohort savings came only from assigned savings. As a result, verified savings for this cohort were zero. Additionally, a different coding error, discussed in Section 4.2.5, was a notable source of discrepancy between claimed and verified gas savings for both cohorts.

4.2 Gross savings validation and accuracy

The previous section presented gross realization rates and identified the primary drivers of divergence between claimed and verified savings. This section focuses on broader issues related to pop NMEC savings accuracy, supporting our top-level finding that improvements are needed to support the overall accuracy and reliability of NMEC savings. The specific findings in this section flow from the process of evaluating pop NMEC and also illuminate a wider set of opportunities to improve accuracy and reliability in a comprehensive fashion.

The findings indicate that improving the accuracy of pop NMEC savings requires both program design and ongoing technical considerations, with the former focused on eligibility criteria, screening, and documentation practices, and the latter on the consistent selection and application of M&V methods. All of the programs' documentation issues require a solution that provides visibility into the recipients of pop NMEC program interventions, given that only aggregate pop NMEC claims are recorded in CEDARS. The current system is clearly in need of substantial improvements, and building a foundational solution will not only provide visibility into customer-level activity but also ensure valuable consistency across PAs and programs. Eligibility requirements similarly play an important role in supporting savings accuracy and reliability, as NMEC is only valid when applied to an appropriate customer population. From a technical perspective, the availability of easily replicable, methodologically vetted code is critical to the consistent application of M&V methods. Additionally, because the goal of NMEC modeling is to normalize for routine drivers of consumption and isolate program impacts, NMEC may be best served by greater consistency in how modeling approaches capture routine drivers.

4.2.1 Documentation and traceability

As detailed in Sections 4.1.1 to 4.1.3 above, nearly all programs faced challenges maintaining clear, consistent, and transparent documentation, from enrollment through final M&V. For pop NMEC programs, accurate savings verification depends on well-organized, version-controlled project-level records that clearly support the cohort-level savings reported in CEDARS, the official source of claimed savings. Because savings are claimed in aggregate at the annual program level, PAs bear sole responsibility for retaining documentation sufficient to substantiate the underlying individual customer claims. In practice, documentation and data management practices often did not provide clear linkages between enrolled projects, modeled savings, and final claimed values.

Almost no programs provided documentation that clearly tied individual projects to CEDARS cohorts. We examined participating customer rosters and PA M&V impact analysis files to understand how the programs constructed the aggregated values in CEDARS. Only BayREN provided project-level records with projected savings that directly matched the CEDARS claim. For most other programs, establishing the relationship between project-level records and claimed savings required substantial follow-up. For PG&E's OBF and NetOne programs, we could not definitively link project-level projected savings to the final claimed values.[30] Table 4-5 illustrates the savings discrepancies for the OBF program.

[30] The PG&E NetOne program only claimed pop NMEC program savings in PY2024 and did not have true-up values for evaluation verification.

Table 4-5. Discrepancies between claimed and projected electric savings provided for evaluation for OBF

Program Year	CEDARS Claimed Savings	Participant Roster			PA M&V Impact-Analysis List Files		
		Number of sites	Projected Savings	Deviation from CEDARS (%)	Number of sites	Projected Savings	Deviation from CEDARS (%)
Electric (kWh)							
2022	28,141,549	575	30,523,337	-8%	549	30,411,581	-8%
2023	22,770,868	467	21,633,298	5%	424	21,943,284	4%
2024	18,992,289	348	17,361,467	9%	341	21,481,113	-13%
Gas (therm)							
2022	49,312	9	21,082	57%	25	172,787	-250%
2023	58,851	12	16,892	71%	16	57,706	2%

Documentation challenges contributed to notable discrepancies between the PA-documented savings and the final savings claimed in CEDARS. Although documented and claimed savings corresponded more closely relative to the previous evaluation,[31] differences between these values persist. For example, MCE's PY2022 electric claims reflect PA M&V or documented savings that are approximately 50% higher than the values reported in CEDARS, the primary reason GRR exceeds 100% for this cohort.

Connection to overarching findings: Clear documentation and traceability is the foundation of any ratepayer funded energy efficiency program and are arguably more critical for pop NMEC than other program types. Pop NMEC relies on a complex, multi-step process and advanced methods that must be implemented in a fully transparent manner, particularly because savings calculations are performed by implementers. As a result, documentation detailing customer eligibility, project progression, and modeled savings must be complete, accurate, and consistently maintained. Given the range of observed shortcomings and the broader need for transparency, PAs should propose to the reporting PCG a solution that defines a customer project-level pop NMEC dataset that links with 100% accuracy to both initial and true-up aggregate CEDARS claims.

4.2.2 Replicability and transparency

The NMEC rulebook requires that savings calculation methods be documented in sufficient detail to allow replication of PA M&V results, with data, tools, and algorithms available to support transparency in the M&V process. Documentation and transparency gaps in the PA M&V process therefore undermine the NMEC rulebook's intent for fully replicable claims. Where required materials were incomplete or unavailable, our replication was necessarily a partial exercise that relied on inference rather than fully traceable methods and inputs. While we could generally replicate energy savings, doing so often required substantial inference due to incomplete documentation of methods.

Table 4-6 summarizes the availability and completeness of key elements required to reproduce PA-documented savings by program type. These elements reflect the practical conditions needed to reproduce documented savings and to trace final claims back to underlying data, models, and assumptions. As shown, OBF consistently met the replication prerequisites, while MAP-style programs and the non-MAP Institutional Partnerships (Non-MAP IP) program exhibited partial or missing elements across multiple dimensions.[32]

Table 4-6. Replicability and documentation checklist by program type

Replication Dimension	Replicability Element	MAP-Style	Non-MAP (IP)	OBF
Foundational inputs	Version-controlled code and datasets	◐	✗	✓
	Documented modeling methods (open-source or equivalent)	✓	✓	✓

[31] Tables 4-3 and 4-4 of the previous evaluation show large discrepancies between documented (PA M&V results) and final claimed values for PY2019 through PY2021.
[32] OBF has operated as an NMEC program for five years and undergone prior evaluation.

Replication Dimension	Replicability Element	MAP-Style	Non-MAP (IP)	OBF
Data preparation and modeling process	Comprehensive data dictionaries	✗	✗	✓
	Fully documented data processing steps	✗	✗	✓
	Documented customer attrition and exclusion processes	✗	✗	✓
	Unique identification of final model-run configuration	✗	✗	✓
Traceability and replication	Consistent replication across sites	◐	✗	✓
	Clear steps from model outputs to reported results	✗	✗	✓
	Replication feasible at a high level	✓	◐	✓
Outcome	Ability to fully reconcile replicated and documented savings	◐	✗	✓

Key: ✓ = Provided / consistent ◐ = Partially provided / inconsistent ✗ = Not provided

The replication outcomes reflect the varying adherence to prerequisites (Table 4-7). Replication realization rates for the MAP-style programs and the non-MAP Institutional Partnerships program were not 100% because documented savings could not consistently be reproduced using the provided code, methods, and supporting materials, many of which were missing or incomplete. In these cases, after extensive replication attempts using PA-provided materials, we made reasonable and transparent modifications to code or methodological inputs to approximate documented savings; however, these adjustments did not always fully reconcile results.

As indicated in Section 4.1, for MAP-style programs, the evaluation relied on a newer version of the model than the one the PAs and implementers used to claim savings. Replication outcomes were nevertheless close to documented values, with replicated electric savings averaging approximately 96% of documented savings across cohorts with available documentation. Differences between replicated and documented savings likely reflect a combination of incomplete information needed for replication (such as detailed data preparation steps) and differences in model vintage between the claimed savings and replication approaches.[33]

For the Institutional Partnerships program, limitations in the provided code and data package prevented reproduction of the documented savings, resulting in replication outcomes that varied widely across projects and modeling steps. For example, one high-savings site had a documented counterfactual (hypothetical performance period usage absent project installation) more than twice as high as the replicated counterfactual. In addition, poor model fit, a substantial amount of missing weather data, and other data and modeling quality issues exacerbated the problem. In contrast, the OBF outcomes reflect a high degree of transparency, with the gas replication realization rate differing from 100% primarily due to a coding error, as PA-documented annual savings values in both program years were calculated by summing average daily savings for each month rather than aggregating the full monthly savings values over the performance period.

[33] MAP-style programs did not provide documented demand (kW) savings values for replication, and most cohorts did not claim gas savings. As a result, replication outcomes for MAP-style programs primarily reflect electric savings comparisons.

Table 4-7. Pop NMEC replication outcomes by program and year

Style	PA	Program	PY	Documented	Replicated
Electric (kWh)					
MAP-style	**BayREN**	**Business FLEXmarket**	**2023**	1,906,444	97%
	PCE	**FLEXmarket Program**	**2023**	1,391,415	91%
	MCE	**Commercial Efficiency Market**	**2022**	1,268,225	104%
			2023	3,707,713	96%
	SCP	**SCP FLEXmarket**	**2023**	NA	NA
Total				**8,273,796**	**96%**
Non-MAP Style	**PG&E**	**Institutional Partnerships**	**2023**	741,022	67%
Total				**741,022**	**67%**
OBF	**PG&E**	**On-Bill Financing (OBF)**	**2022**	8,892,504	100%
			2023	6,135,779	100%
Total				**15,028,283**	**100%**
Gas (therm)					
OBF	**PG&E**	**On-Bill Financing (OBF)**	**2022**	78,482	93%
			2023	9,347	398%
Total				**87,829**	**125%**

The transparency and replicability challenges in the PA M&V process were most evident in the characterization of peak demand impacts and uncertainty. Except for OBF, the PAs provided limited documentation describing how they calculated peak demand impacts. For example, beyond general references in their M&V plans, the MAP-style programs did not indicate how they generated peak demand claims, whether the claimed values reflected actual year or TMY weather, and how they defined the heat wave periods required for peak demand calculations. As noted in Section 4.1.2, they also generally did not provide documented peak demand impact values. Additionally, except for OBF, the remaining programs did not provide program-level uncertainty estimates beyond high-level references of methods used to calculate these in their M&V plans.

Connection to overarching findings: Replicability challenges complicate the evaluation process and reduce the transparency of the embedded M&V process. Replication is a necessary first step to better understand the decisions made during performance calculations. When substantial time is spent on basic replication, attention is diverted from more substantive consideration of those decisions and the implications of the methods used. These issues are further compounded by the growing variety of program methods, resulting in an increasingly complex and impractical set of approaches that creates confusion rather than transparency. In addition, gaps in documentation (such as missing information on model versions, configurations, and data processing steps) reduce transparency by making it difficult to fully trace and reproduce how savings estimates were developed.

4.2.3 Peak demand savings identification

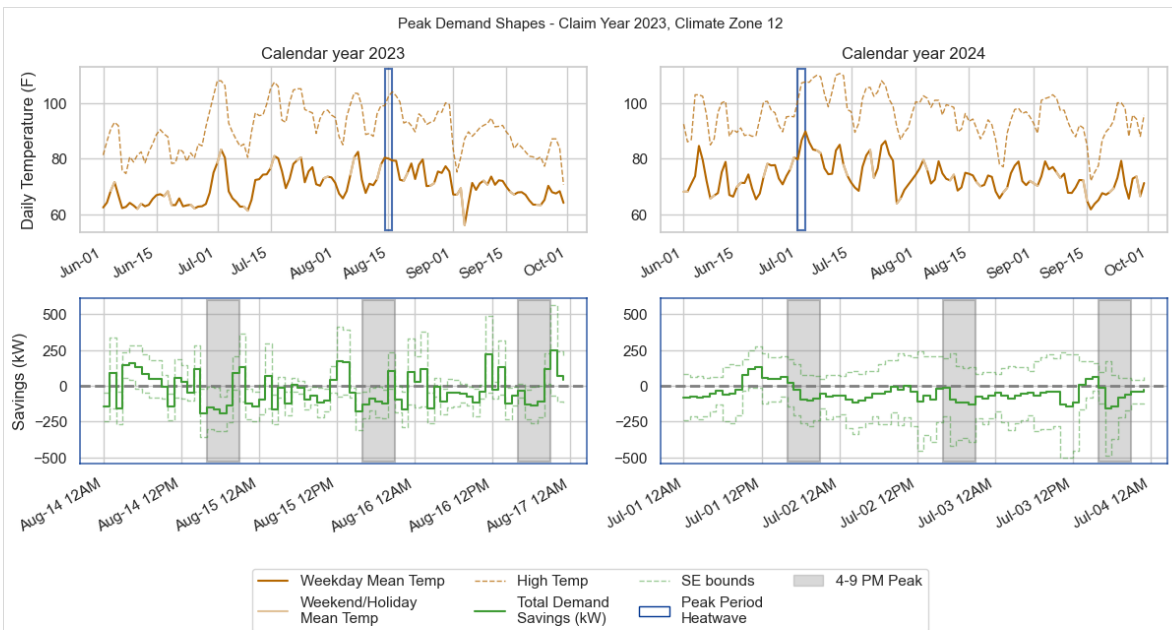
Accurate peak demand savings estimates depend on correctly identifying the peak hours and heat wave conditions that define them. Because peak impacts are concentrated in a small number of hours, differences in how heat wave days are selected (e.g., TMY-based DEER periods versus actual-year weather) affect reported peak demand savings.

Our verification of peak demand impacts indicates that most programs substantially overestimated peak savings. Peak impacts are inherently time-specific and represent load reductions during defined peak hours. For programs spanning multiple climate zones, appropriate program-level peak impacts must reflect climate-zone-specific peak conditions. Under DEER, peak demand savings are typically based on average demand reductions during a 15-hour window that covers the 4 PM – 9 PM peak hours across three consecutive hot non-holiday weekdays, identified using TMY conditions.

All programs except OBF relied on avoided-energy-use calculations that reflect reporting-period conditions, suggesting that they estimated peak demand savings using actual post-installation weather rather than DEER-defined TMY conditions. As a result, we identified heat wave periods using actual post-installation weather to estimate demand reductions by climate zone. We then aggregated the climate zone reductions to derive program-level peak-demand impacts. Thus, our verified estimates reflect climate-zone-specific heat wave conditions and the corresponding hourly demand reductions. For sites with assigned energy savings, we estimated peak demand reductions using the ratio of measured peak impacts to measured energy savings for the relevant climate zone and aggregator or project developer.

Figure 4-4 illustrates our verification approach using average hourly load impacts for MCE PY2023 customers; the PA claimed large peak demand savings that could not be verified due to a lack of supporting documentation. The top panels show weekly average and maximum temperatures and identified heat wave periods by year, while the bottom panels present estimated hourly load impacts with 90% confidence bounds over the corresponding heat wave hours. The figure shows negative aggregate hourly load changes during peak hours in both summers, consistent with evaluated negative peak impacts for this climate zone. Although most MCE customers are in climate zone 12, this cohort reflects peak demand savings driven by larger per-site impacts in the other climate zones. Additional PY2023 peak impact results, including other climate zones, are provided in APPENDIX D.

Figure 4-4. Peak period weather and hourly impacts for MCE PY2023, climate zone 12



4.2.4 Eligibility and enrollment screening

The NMEC rulebook specifies eligibility requirements for pop NMEC programs to ensure that participating sites have energy use characteristics and data that support performance-based savings estimation. In this section, we address two dimensions of eligibility and their associated enrollment screening and discuss the implications of each for program savings accuracy and evaluation outcomes.

4.2.4.1 Load suitability and customer mix

The inclusion of non-residential customers in pop NMEC programs increases the importance of enrolling customers with load patterns that are well suited to standardized, pop NMEC modeling. Because pop NMEC requires isolating program impacts from non-program factors by applying uniform models to all sites in the same manner, participating sites must have reasonably regular usage patterns driven by recurring schedules and weather conditions. When sites exhibit process- or event-driven load that is weakly related to schedule or weather, savings estimates based on these methods become less credible and programs are more likely to rely on assigned savings, which reduces the transparency and accuracy of reported impacts.

The NMEC rulebook limits pop NMEC participation to commercial or commercial-like sites and excludes sites with industrial or process-driven loads to support standardized, pop NMEC modeling. Nevertheless, we found that program participation extended beyond these permissible categories. Approximately 15% of customers across programs came from sectors or subsectors that commonly exhibit process- or event-driven loads (for example, manufacturing, agriculture, infrastructure and utilities, and fairgrounds), which are more difficult to model consistently under pop NMEC. These kinds of customers affect savings accuracy because their loads often do not support credible counterfactuals for estimating savings.

Two customer load profiles from the OBF program highlight the types of load pattern issues observed in these industries. Figure 4-5 shows the first example, an agricultural site with irregular usage pattern that does not align with recurring schedules or weather condition changes. The upper panel shows the baseline period load profile while the bottom panel shows the performance period load profile. In the baseline period, usage includes pronounced peaks and step changes, including a winter peak that does not persist into the performance period, suggesting drivers unrelated to recurring schedule- and weather-correlated consumption. These types of irregular loads lead to models that do not produce credible counterfactuals, preventing some sites from supporting modeled savings estimates. In PY2022, the OBF electric program included 36 agricultural sites, the majority of which received assigned savings estimates due to modeling limitations or data insufficiency. In PY2023, the PA significantly reduced agricultural participation, retaining only six sites, of which just two supported modeled savings.

Figure 4-5. Example of difficult-to-model agricultural load

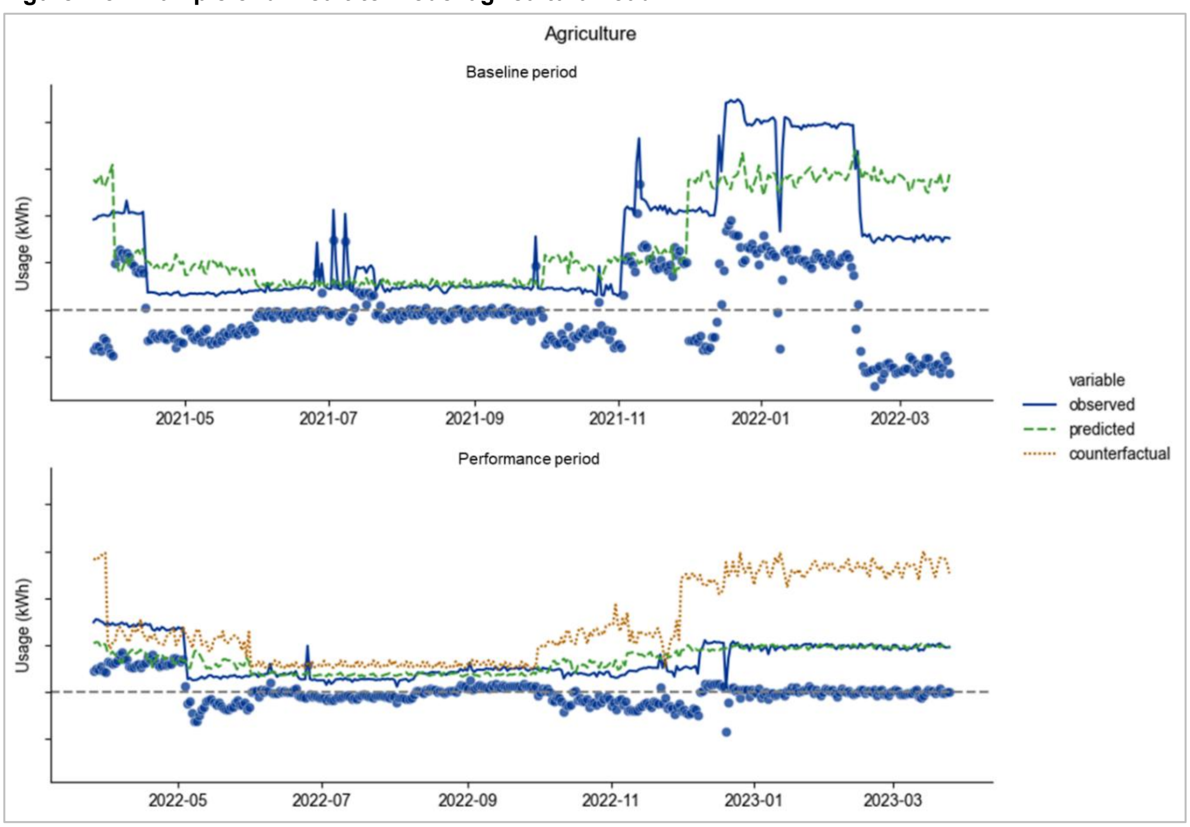
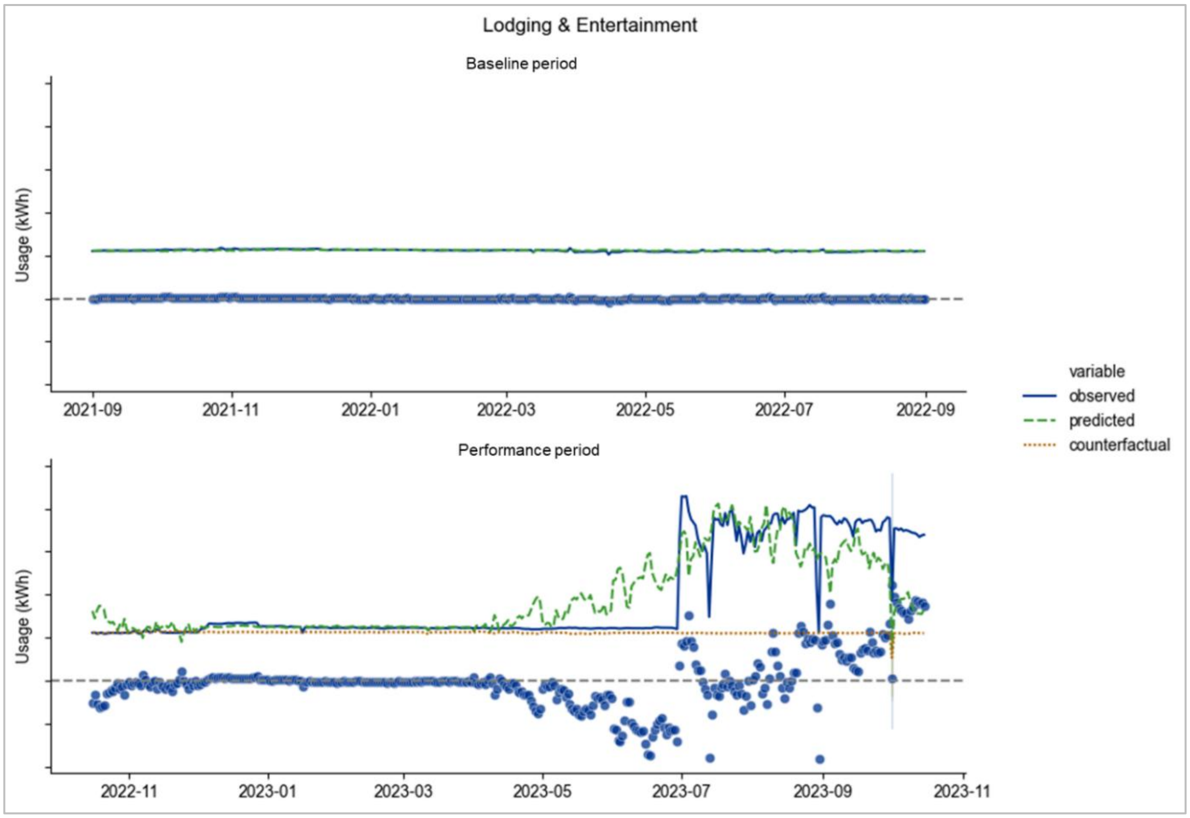


Figure 4-6 shows the second example for a lodging and entertainment site with a relatively flat baseline period load profile followed by a large structural shift during the performance period. This site's baseline period load appears relatively constant, but the pattern changes substantially in the performance period. Load irregularity of this type is typical across the lodging and entertainment sites included in the program. This particular site passed model CVRMSE checks and was only excluded from receiving modeled results due to a savings outlier check, discussed in the next section.[34] The uniform model applied across all sites does not capture the drivers of these types of energy-use changes and thus produces a poor counterfactual. As a result, such sites can yield misleading modeled savings.

Figure 4-6. Example of lodging and entertainment load



Connection to overarching findings: Non-residential pop NMEC programs need to tighten eligibility criteria and enrollment practices to better prioritize customers with load patterns suited to pop NMEC. Strengthening screening to exclude sites without recurring schedules, clear weather responsiveness, or those dominated by process- or event-driven loads would reduce reliance on exclusions and assigned savings and improve the credibility, consistency, and transparency of reported savings.

4.2.4.2 Pre-participation project screening

Pop NMEC programs require effective eligibility screening to ensure that participating sites have sufficient and appropriate data to support meter-based measurement, as specified in program M&V plans. Eligibility screening therefore directly affects savings accuracy by determining which sites enter the analysis population and whether program-level claims are supported

34 The modeled result produced a negative outlier savings for this site, which was replaced with an assigned savings value.

by performance-based estimates. When PAs do not identify or screen out ineligible sites during enrollment, programs cannot develop robust, transparent, and defensible M&V savings claims.

Table 4-8 summarizes the specific eligibility screening criteria prescribed in the current M&V plans for the MAP-style, OBF, and non-MAP-style programs. The table highlights the types of pre-participation eligibility checks that are largely shared across programs and are in place to support effective performance-based model development.

Although OBF's M&V plan includes explicit pre-participation data sufficiency criterion, in practice, pre-screening did not exclude all sites with inadequate pre-installation data. As a result, approximately 7% of electric sites and more than 60% of gas sites ultimately received non-performance-based savings estimates. Operational challenges in eligibility screening, particularly for gas sites with limited or inconsistent interval data, can significantly affect the share of savings derived from modeled versus other approaches.

Table 4-8. Eligibility screening in program M&V plans

Screening Dimension	MAP-style	OBF	Non MAP-style
Pre-participation data sufficiency	✓	✓	✓
Baseline NRE (e.g., PV, EV)	✓	✓	✓
Participation in other programs	✓	✓	✓
Comparator availability	✓	✓	✓
Baseline model fit (e.g., CVRMSE < 1)	CVRMSE < 1	CVRMSE ≤ 1	CVRMSE ≤ 0.5
No major planned load changes	—	✓	✓

Connection to overarching findings: Eligibility criteria should be strict enough to limit reliance on assigned savings estimates. Because pop NMEC uses uniform models, the key information at enrollment is whether a site's load is stable and predictable enough for credible savings estimation. PG&E's planned shift to site-level FSU screening is a positive step; applying a consistent minimum threshold (e.g., FSU 90/100, with tighter requirements for large customers) would help standardize enrollment decisions. More broadly, screening based on site-level FSU and expected savings depth can strengthen savings accuracy and improve cohort savings certainty.

4.2.5 M&V exclusions and assigned savings

While eligibility screening governs which sites enter pop NMEC programs, M&V exclusions determine whether enrolled sites ultimately contribute to performance-based savings. These exclusions allow PAs to remove enrolled sites from receiving performance-based savings when post-installation conditions prevent accurate measurement. Exclusions serve as a safeguard against invalid model-based savings estimation, but also increase uncertainty.

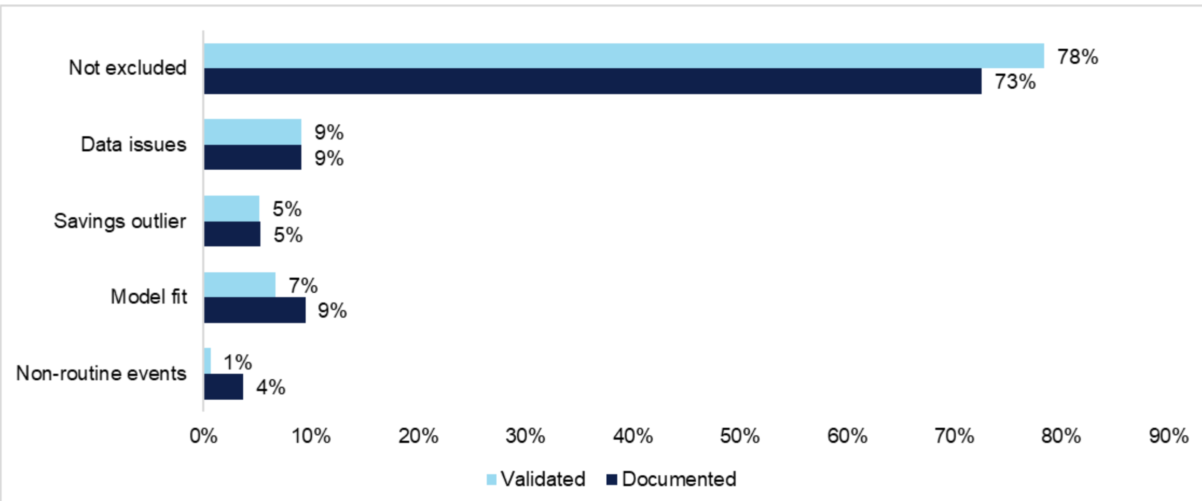
All the PAs defined the criteria for excluding sites in their M&V plans, which they applied to analysis sites during program performance assessment. We found clear PA documentation of these criteria and were able to identify where PAs applied the exclusions in the provided M&V datasets. The criteria can be grouped into four categories:

- **Model fit** exclusions applied when CVRMSE values exceed a specified threshold during savings estimation, typically defined as CVRMSE greater than one.
- **Data issue** exclusions applied when performance-period data are insufficient for valid analysis; all the PAs require 90% coverage, with some additionally requiring more than 90% of monthly meter reads.
- **Savings outlier** exclusions applied when meter-based savings fall outside defined bounds, set at ±50% relative to baseline consumption.

- **Non-routine event (NRE)** exclusions applied when post-installation site conditions, such as the installation of distributed energy resources or long-duration operational changes, affect energy consumption and invalidate performance-based modeling results; the PAs applied these exclusions based on aggregator-reported NREs or utility records of onsite generation installations.

The PAs reported a higher overall exclusion rate than the evaluation found. As part of our validation process, we reviewed each site against the PA M&V exclusion criteria to verify that the PAs followed their stated exclusion methodology and to identify additional sites that met the criteria but were not flagged. When applying the PA exclusion criteria, our evaluation excluded 22% of sites, which is fewer than the 27% of sites the PAs excluded. Figure 4-7 illustrates these patterns.

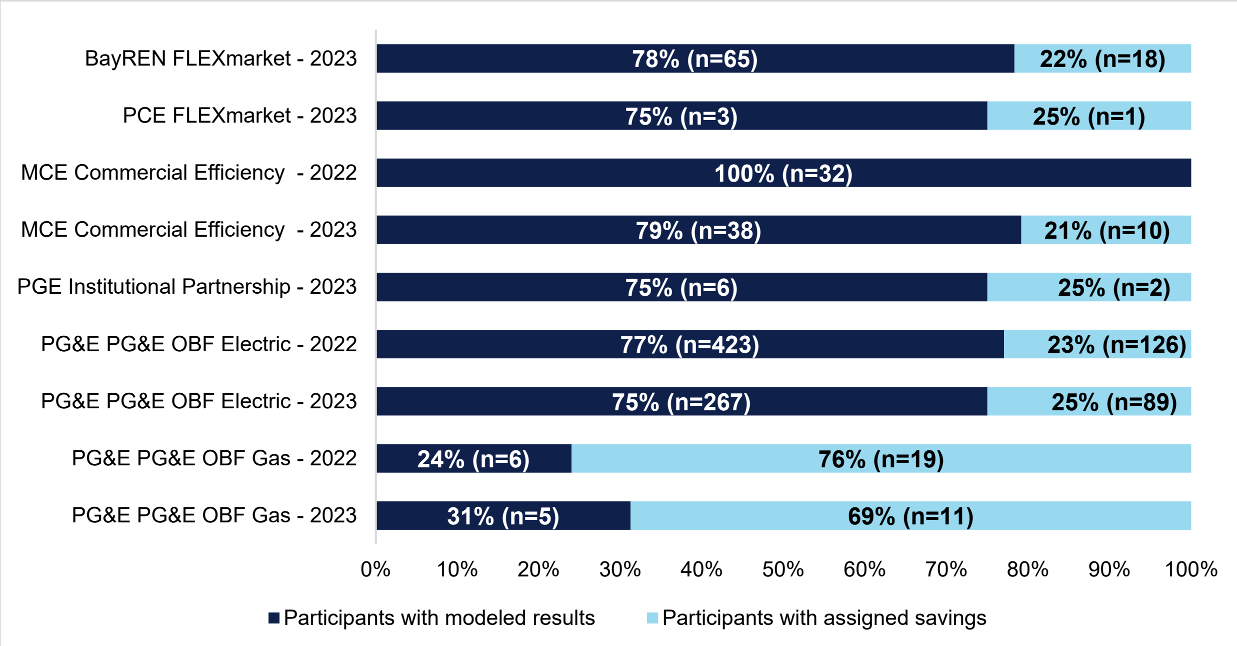
Figure 4-7. Distribution of documented and validated site exclusions by criterion



When PAs cannot develop performance-based savings for sites due to one or more M&V exclusion conditions, they rely on assigned rather than measured savings. Assigned savings are deemed or engineering estimates that PAs calculate by applying aggregator-level realization rates to each excluded site's projected savings.[35] Figure 4-8 summarizes the share of sites with modeled and assigned savings for each program. For most electric cohorts, approximately 20% to 25% of documented savings are derived from assigned savings. The share of assigned savings is substantially higher for gas cohorts than for electric cohorts, exceeding 60%. For PY2022, all claimed gas savings were assigned rather than modeled, as modeled results indicated negative savings.

[35] The PG&E M&V plan indicates that assigned savings are "based on the realization rate for the other projects … in the aggregators' portfolio," with electric and gas savings calculated separately. The provided code indicates that the program applied the prior year's realization rates, referred to as expected realization rates, to projected savings.

Figure 4-8. Share of pop NMEC customers with modeled and assigned savings



Although the PAs documented their assigned savings approaches clearly in the M&V plans, following those approaches did not always result in replication of the PA-assigned savings or required overly complex implementation. While the MAP programs and the non-MAP Institutional Partnership program assigned savings based on project developer realization rates, we could not consistently replicate the assigned savings reported by these programs using this approach. In the OBF program, the PA assigned savings to sites for partial performance periods based on site-specific ineligibility dates, which required detailed tracking of those dates and introduced substantial complexity. That complexity contributed to a coding error, under which the PA applied assigned savings beginning one year after the ineligibility date rather than immediately after the date, significantly understating assigned savings. A simpler approach, consistent with standard billing analysis practice, such as applying an average M&V percent-savings rate to excluded sites, could reduce complexity, ground the savings more directly in measured results, and improve transparency while preserving the intent of the exclusion framework.

Connection to overarching findings: Assigned savings create challenges for meter-based, P4P programs because they support savings claims that are not directly tied to measured outcomes. Although assigned savings can be appropriate when performance-period data are insufficient or post-participation conditions prevent valid estimation, reliance on assigned savings across a meaningful share of sites reduces transparency and increases uncertainty in program-level results. This dynamic increases risk for both participating customers and ratepayers and underscores the importance of monitoring the frequency of exclusions and incorporating program design features (for example, screening to exclude sites without regular usage patterns and recurring schedules) to increase the share of savings supported by measured performance.

4.2.6 Model selection and application

We assessed the performance of current pop NMEC program methods against the requirements established in the NMEC rulebook. The rulebook requires that pop NMEC methods:

- Account and normalize for routine drivers of energy consumption, such as weather, using pre- and post-installation whole building energy consumption data and account for NREs using comparison groups
- Apply the modeling approach consistently across all participating sites within a cohort

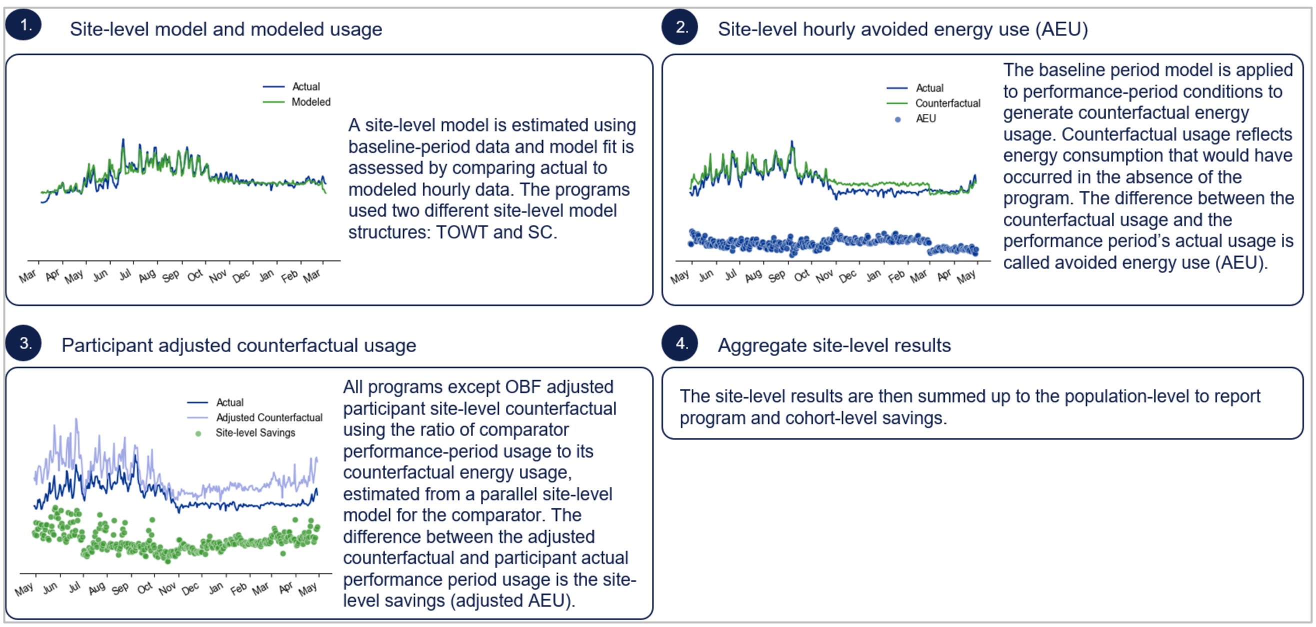
- Demonstrate acceptable performance at the program level, including meeting the aggregate savings uncertainty benchmark (90% confidence / ±25% relative precision or FSU)

The discussion below first presents the modeling approaches used by the programs and then focuses on the first two rulebook requirements and on broader methodological considerations affecting modeling performance. We address savings uncertainty separately.

4.2.6.1 Modeling approaches used in the evaluation

The evaluated programs used two modeling approaches to estimate final claimed savings for cohorts with completed true-ups. In each case, implementers developed site-level models using baseline-period data and then applied those models to performance-period conditions to generate counterfactual energy consumption. All approaches incorporated a comparison-group adjustment to account for non-program effects. By comparing adjusted counterfactual estimates to observed performance-period consumption, implementers estimated avoided energy use (AEU).[36] Figure 4-9 illustrates this general process.

Figure 4-9. General model development process



The modeling approaches used by each program type are summarized below.

1. **MAP-style and non-MAP-style Institutional Partnerships programs (TOWT).** Both the MAP-style and non-MAP-style Institutional Partnerships programs used a time-of-week temperature (TOWT) framework to model site-level hourly energy consumption as a function of piecewise linear temperature terms and time-of-week indicators for each hour of the week. The former was implemented in OpenDSM, an open-source code package for meter-based energy consumption methods, while the latter was implemented in OpenEEmeter, a predecessor to OpenDSM. Hourly load was estimated using 12 separate monthly models with three-month weighted rolling windows. This approach applies a consistent, predefined time structure within three-month periods improving transparency and comparability and emphasizing routine load drivers. However, the rolling windows allows the set of models to adjust to within-year trends or NREs which, again, get carried forward into the generation of counterfactual energy consumption.

[36] AEU differs from weather-normalized impacts in that AEU reflects performance period weather, while normalized impacts reflect TMY weather.

2. **On-Bill Financing (OBF) program (synthetic control approach).** The OBF program used a synthetic control (SC) approach to model site-level hourly energy consumption as a function of piecewise-linear temperature terms, time-of-week indicators for each hour of the week, and a granular comparison-group load profile. Unlike the MAP-style and non-MAP-style approaches, which estimate participating customer site and comparison group models in parallel, this approach directly incorporates the comparison group as an additional explanatory variable within the site-level regression. This method of controlling for non-program effects is novel, and its statistical properties have not yet been widely vetted in the evaluation context. While the SC approach does not include separate monthly models or monthly indicators, any month-to-month or other sub-annual trends present in the granular profile are reflected in modeled site consumption. This can be problematic in NMEC applications because it allows within-year trends or NREs to be carried through the modeling process. Specifically, carrying sub-annual variation through the model may obscure NREs during the eligibility phase and carry those effects into the modeled counterfactual during the performance phase, creating challenges for interpreting savings and developing hourly impact (8760) shapes.[37]

4.2.6.2 Suitability of models for capturing routine drivers of load

We assessed whether the modeling approaches used by each program type are suitable for capturing routine drivers of energy use, such as weather and recurring schedules. As described in the previous section and consistent with the NMEC rulebook, all programs used whole-building modeling frameworks to account for routine drivers of energy consumption. However, the approaches differed in how they separate routine drivers from within-year trends and NREs.

As noted above, the MAP-style and non-MAP-style programs relied on the TOWT framework to model site-level energy consumption. This approach captures routine drivers of load through weather terms and time-of-week structures that represent recurring schedules. However, because it relies on separate monthly models with three-month rolling windows, it can also adapt to within-year changes in consumption patterns. While this flexibility can improve model fit, it creates the potential to absorb within-year trends or NREs into the model rather than reflected as uncertainty, thereby carrying those effects forward into counterfactual estimates.

In contrast, the OBF program relied on a synthetic control (SC) framework that incorporates a granular comparison-group profile directly into the site-level model. The SC approach captures routine drivers through weather variables, time indicators, and information contained in the granular profile, but does not include the flexible monthly model structure used in the TOWT approaches. As a result, it does not approach the same level of temporal flexibility. However, because month-to-month and other sub-annual variations present in the granular profile are reflected in modeled site consumption, those patterns can still influence modeled consumption and counterfactual estimates. This is a fundamentally different mechanism from the rolling-window TOWT approach, as the variation originates from the comparison-group profile rather than from model flexibility itself. As discussed in the previous section, this may obscure NREs during the eligibility phase and carry those effects into the modeled counterfactual during the performance phase.

In addition to the above, while the modeling approaches adopted by the programs meet the NMEC rulebook's requirement to model whole-building consumption, only the OBF program's reported savings reflect weather-normalized values. The M&V plans for the remaining programs indicate normalization to reporting-year temperatures rather than to normal year (TMY) conditions, which does not fully meet the NMEC rulebook's normalization requirement. Relying on a single year of weather does not fully normalize for weather, since it may not reflect typical conditions.

[37] More broadly, for savings estimation and TSB analysis, the intent is typically to capture changes between baseline and performance periods rather than within-period trends. The appropriate treatment of these effects, and the role of SC in NMEC applications, remains an open question that will require broader stakeholder alignment.

4.2.6.3 Consistency of application across sites

The NMEC rulebook requires that pre-selected pop NMEC methods be applied uniformly across all sites within a cohort. **Our review found that each PA employed a defined modeling framework that was applied consistently across participating sites, rather than tailoring approaches or inputs on a site-by-site basis**.

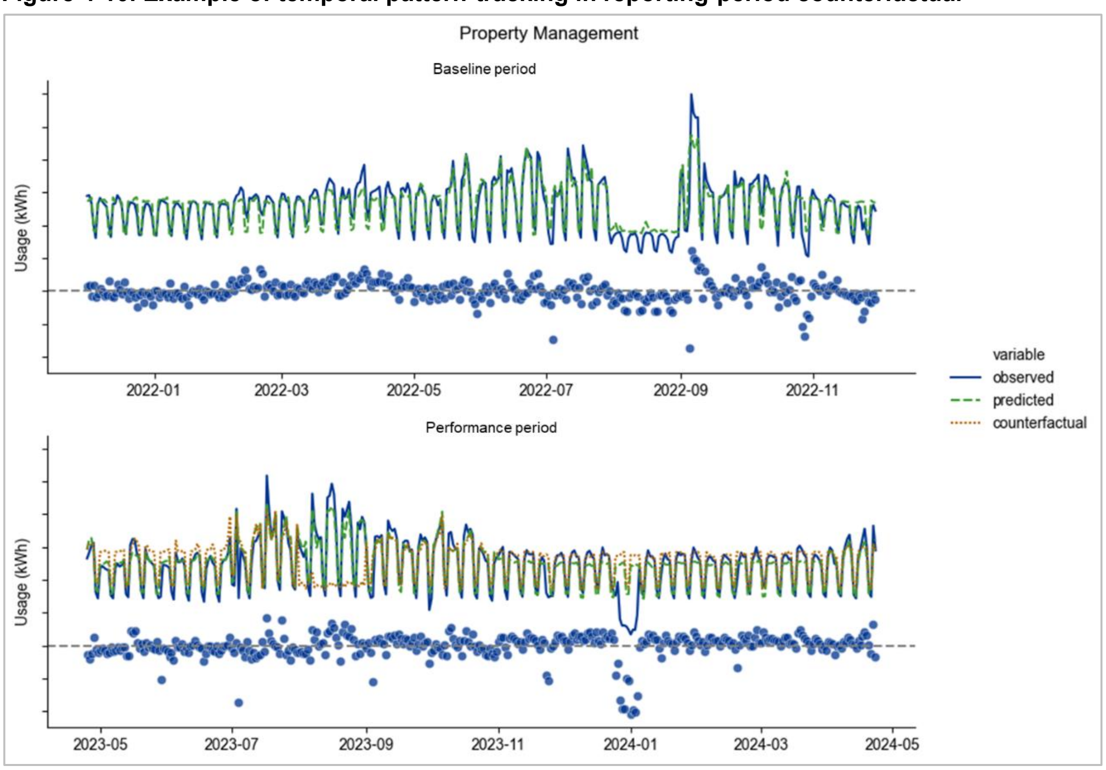
Although the programs used different modeling approaches, each employed a predefined model structure that was applied consistently across participating sites within a cohort. None of the programs varied model choice or model inputs on a site-by-site basis. As a result, all programs satisfied the NMEC rulebook's requirement for uniform application of methods.

4.2.6.4 Implications of modeling behaviors

While all programs applied the same modeling framework and used the same inputs across sites within a cohort, the modeling approaches raise methodological considerations that affect interpretability and program risk. The monthly TOWT models used by the MAP-style and non-MAP-style programs can and did capture within-year variation in consumption, including patterns that are not clearly attributable to stable, routine load drivers such as weather or recurring schedules. We observed instances where predicted counterfactual usage showed abrupt shifts tied to month-specific patterns in baseline-period data rather than routine drivers of energy use.

Figure 4-10 demonstrates an example where the model's performance-period counterfactual (orange dotted line) exhibits a step-like shift that does not correspond to clear weather- or schedule-based drivers. The apparent four-week, partial shutdown in August of the baseline period shows up in the counterfactual for the August of the performance period. That shutdown, however, appears to have been a one-time occurrence and a clear NRE. The performance period counterfactual again displays discrete shifts and pattern changes that appear to be artifacts of baseline period temporal structure rather than routine drivers of consumption. In such cases, the issue is not poor overall fit, but rather a model that tracks time-related patterns in ways that are not easily attributable to routine load drivers.

Figure 4-10. Example of temporal pattern-tracking in reporting-period counterfactual



While models using this approach may achieve strong statistical fit, their application obscures underlying load drivers for sites with irregular usage patterns. Models that too closely track time-related patterns can reduce the ability to distinguish routine load driver effects from NREs, which has implications for site eligibility screening and for understanding program-level risk. In this context, site-level consumption models that primarily capture routine drivers of energy consumption will produce residual variation that serves a useful screening function, helping assess the suitability of a customer's data for inclusion in pop NMEC. Thus, efforts to optimize model fit can reduce the visibility of this variability, which in turn affects how risk is identified and managed at the program level.

4.2.6.5 Heterogeneity in pop NMEC approaches

While the NMEC rulebook does not constrain the modeling frameworks used for pop NMEC savings estimation, the variation in methods leads programs to estimate savings and manage modeling risk in substantially different ways. As a result:

- **Savings are calculated differently across programs -** When program savings are based on different methods, the resulting estimates reflect differences in assumptions, sensitivities, and failure modes. Thus, variation in reported savings may be driven as much by method choice as by true differences in performance, making results more difficult to interpret from a stakeholder perspective. This runs counter to the standardized intent of pop NMEC, where consistency is meant to support comparability and credible performance-based claims.
- **Modeling risk is managed unevenly -** Different methods vary in how well they reveal issues such as unstable usage patterns, poor counterfactual fit, or sensitivity to non-routine behavior. This means some programs may be better at identifying high risk sites or cohorts, while others may inadvertently mask those risks, even if all are nominally compliant with the NMEC rulebook.

4.2.7 Reasonableness of gross savings

Assessing the reasonableness of gross savings helps ensure that documented and verified impacts are plausible and proportional to baseline energy use and do not reflect artifacts of data issues or implementation errors. In pop NMEC programs, these checks serve as a backstop to identify implausible values.

Consistent with this purpose, we applied conservative reasonableness checks focused on clearly implausible outcomes. We identified and adjusted assigned savings that exceeded baseline consumption as unreasonable. Where such savings were assigned to replace an outlier modeled value and the modeled value did not exceed 100% of baseline consumption, we reverted to the modeled value. In all other cases including assignments due to NREs, insufficient data, or poor model fit, or where the replaced modeled value also exceeded 100% of baseline consumption, we capped assigned savings at 50% of baseline consumption. These adjustments affected eight sites: four from MAP-style programs and four from the OBF program.

Connection to overarching findings: Savings reasonableness checks had very limited effect on overall gross savings results, indicating that implausible savings magnitudes were not a primary driver of program-level outcomes.

4.2.8 Uncertainty characterization

Program savings estimates incorporate inherent variability arising from modeling choices and cohort characteristics, including how models specify key drivers of energy use, how comparison groups are constructed, and program population size. Population size is one contributor to the precision of program savings estimates because larger cohorts provide more information for estimating savings and tend to reduce uncertainty associated with customer-level variability, model error, and comparison-group construction. In contrast, smaller cohorts, which are common in some non-residential pop NMEC programs, can limit the ability to achieve the 90/25 precision standard even when model fit and comparison group construction are otherwise reasonable.[38]

Program savings estimates have implications for performance payments, customer benefits, stakeholder expectations, and program planning. Consistent uncertainty characterization helps distinguish true performance differences from variability inherent in pop NMEC methods.

Using NMEC rulebook precision standards, only PY2023 OBF electric and demand and PY2023 BayREN demand estimates met the 90/25 savings precision requirement, while most programs aligned with the less stringent 90/50 benchmark, which is commonly used for billing analysis. As noted earlier, only two programs reported uncertainty metrics. The PG&E OBF documentation included the uncertainty approach, code, and both site- and program-level uncertainty estimates, while the PG&E Institutional Partnership documentation included site-level uncertainty estimates.

In both cases, the PA M&V process relied on FSU to characterize site-level savings uncertainty and, for OBF, to summarize program-level savings uncertainty. This approach appears to reflect an interpretation of the NMEC rulebook requirement to use FSU to characterize program savings uncertainty. FSU expresses savings uncertainty as a fraction of the savings estimate. Rolling up site-level FSU values to the program level implicitly assumes that site-level uncertainties are approximately independent and that modeling error is the only meaningful source of savings uncertainty.

In pop NMEC, uncertainty also arises from cross-site variability due to sampling (when applicable) and comparison group construction. Comparison groups, whether formed through matching or granular load profiles, represent a specific realization drawn from a larger pool of eligible non-participating sites. As a result, cohort-level savings estimates vary depending on which non-participating sites enter the comparison group and how closely those comparison-group customers match participating customer usage patterns and trends. Although comparison group methods are designed to control for average NREs and support unbiased savings estimation, the inherent variability in comparison group composition still contributes to uncertainty in program savings results.

Ignoring cross-site variability is only appropriate if site-level estimates are treated as if modeling error is the only meaningful source of savings uncertainty and comparison group randomness is negligible, conditions that generally do not hold in practice. Given this, we characterized program-level uncertainty in a way that reflects these shared sources of uncertainty by using the standard error of the mean savings across sites ($s/\sqrt{n}$) to report relative precision.[39]

Under the NMEC rulebook's pop NMEC precision standard (90/25), only the PY2023 OBF electric and demand and PY2023 BayREN demand results meet the requirement (Sections 4.1.1 and 4.1.2). However, the alternative 90/50 standard may be justified for billing-analysis-based savings estimates, particularly for non-residential programs with smaller sample sizes.[40] All but two electric savings estimates meet this less-stringent standard. BayREN's electric savings estimate does not meet the threshold due to an outlier: a single site with energy consumption and savings approximately 30 and 45 times the respective cohort mean values. This outlier significantly skews the results. Excluding that outlier site, BayREN's relative precision improves to 0.31. PCE's electric savings estimates, which have a relative precision of 0.53, fall short because the cohort had only four customers.

---

[38] The relationship between population size and uncertainty depends on other cohort and modeling conditions, including savings variability, data quality, model fit, and comparison group performance. Larger populations generally improve precision, but they do not guarantee that a cohort will meet a specific precision threshold.

[39] For standard quasi-experimental approaches with a comparison group, the standard errors are effectively calculated separately for participants and comparison group and then combined. Aggregated impacts that are comparator- or granular profile-adjusted at the site-level may under-estimate aggregate standard errors.

[40] The required level of statistical accuracy may vary across evaluations. Statistically significant difference from zero represents a substantially lower standard than the frequently proposed 90/10 relative precision. Relative precision depends on the magnitude of the savings estimate, the number of participants, and, depending on the approach, the granularity of the data. Billing analysis results with 90/50 relative precision are common and may be acceptable for program evaluation purposes. Source: DOE Uniform Methods Project, Chapter 8: Whole-Building Retrofit with Consumption Data Analysis (https://docs.nlr.gov/docs/fy17osti/68564.pdf).

Connection to overarching findings. While the NMEC rulebook establishes a 90/25 standard, program evaluation experience shows that many non-residential cohorts align more closely with the 90/50 benchmark commonly applied in billing analysis, particularly when sample sizes are small. Stakeholders currently rely on different methods to estimate and report uncertainty, which limits comparability. Addressing this gap will require the establishment of a consistent, broadly accepted approach for characterizing and reporting program level savings uncertainty.

4.3 Early gross evaluation

Section 4.1 presented the evaluation results for the PY2022 – PY2023 pop NMEC cohorts with completed true-ups. This section provides an early gross-savings assessment for cohorts that have not yet completed the true-up process, including the PY2024 versions of the programs covered in Section 4.1, as well as two new programs with only PY2024 initial claims. These cohorts were still within their performance periods at the start of the evaluation and did not have the full year of post-installation data required for performance-based M&V and true-up calculations. Accordingly, this early assessment is based on available program documentation, primarily M&V and implementation plans, rather than final gross savings results.

Specifically, the assessment covers PY2024 cohorts from the MAP-style, non-MAP, and OBF programs for which we evaluated PY2022 – PY2023 claims, as well as PY2024 cohorts from two additional programs: SDGE's Commercial Core Market Access Program (SDGE Core Market Access), a MAP-style program, and PG&E's NetOne Commercial Efficiency Program (PG&E NetOne), a non-MAP-style program. Table 4-9 summarizes our early gross savings assessments for the first group, based on a review of how program documents address permissible projects, eligibility screening, M&V and normalization approaches, M&V exclusions and assigned savings, and TSB. The table also identifies the elements that the full gross evaluation will assess once these projects submit true-up claims.

Table 4-9. Early gross savings assessment summary of PY2024 MAP-style, Non-MAP, and OBF programs

Assessment area	Early gross assessment
Project and site mix	Implementation plans indicate that MAP-style cohorts focus on small- to medium-commercial sites, non-MAP Institutional Partnerships on state agency facilities, and OBF on a broad range of commercial, industrial, agricultural, and public sector customers; the full evaluation will review inclusion of permissible load and project types
Eligibility screening	M&V plans provide clearly defined pre-participation eligibility screening (including data sufficiency, baseline model fit, baseline NRE requirements); the full evaluation will assess whether PAs applied these criteria as specified
M&V modeling approach	M&V plans describe modeling approaches for estimating avoided energy use that account for schedule, weather, and exogenous changes using comparison groups; the full evaluation will assess handling of sites with irregular load patterns
Normalization of energy use	MAP-style and non-MAP M&V plans call for weather normalization to reporting year temperatures and not TMY as required by the NMEC rulebook, while the OBF M&V plan calls for weather normalization of savings based on TMY (CZ2022 files)
M&V exclusions	M&V plans clearly specify M&V exclusion criteria based on events that trigger performance period NREs; the full evaluation will review the exclusions and the proportion of excluded sites
Assigned savings	M&V plans describe approaches for developing assigned savings estimates for sites that cannot be modeled due to reporting-period NREs; the full evaluation will assess how programs apply these approaches
TSB assumptions	All cohorts use savings shapes from the DEER and custom load shapes library as the CET tool could not, at the time programs made claims, accommodate the use of empirically derived shapes; the full evaluation will assess the implications of the selected shapes for TSB

As indicated earlier, two new programs will have true-up savings for PY2024 claims: SDGE Core Market Access and PG&E' NetOne. SDGE's program is a P4P program for commercial customers in SDGE's service territory that rewards aggregators for delivering peak demand reductions and energy savings that provide the greatest value to the grid and customers, while the pop NMEC portion of PG&E's NetOne program targets commercial customers whose building characteristics and usage profiles support meter-based assessment.

SDGE's M&V and implementation plans, the PA-provided customer roster, and CEDARS tracking data illustrate a program targeting commercial customers within SDGE's service territory, with documented eligibility screening criteria, a SC modeling approach, and a well-defined treatment of exclusions and assigned savings. However, the PY2024 cohort includes only three participating sites, which will likely limit the development of stable, program-level savings estimates. The full evaluation will assess whether the cohort included additional sites and how cohort size and any exclusions affect program-level estimates.

Similarly, a review of PG&E NetOne's M&V and implementation plans, the PA-provided roster, and CEDARS tracking data portray a program that targets PG&E commercial customers with similar building types and leverages a pop NMEC modeling framework consistent with PG&E's broader NMEC approaches. However, the key limitation to gross evaluation is documentation traceability: customer-level projected savings in the PA-provided roster do not align with initial savings claims reported in CEDARS. The full evaluation will need to reconcile customer-level records with CEDARS claims and verify final savings, accounting for all existing discrepancies.

4.4 Net savings results and ratios

Net-to-gross ratios (NTGRs) are designed to capture the degree of program influence on a customer's decision to install program equipment. They indicate the extent to which these actions would not have occurred otherwise. This evaluation followed the same NTG methodology as site-level NMEC, using interviews with participating customers to collect information on program influence, in alignment with California non-residential NTG protocols.[41] We conducted interviews with a representative sample of participating customers, covering 162 sites (12% of the population).

Table 4-10 presents the NTGRs for each program and PY. It shows the counts of projects and customers, followed by the NTGR and relative precisions (RPs). Unlike our gross savings results, this table includes PY2024 because we completed the NTG interviews with PY2024 participants following project installation but before the end of the performance period. The NTGR reflects only those sites with first-year gross electric savings. We could not produce different ratios for lifecycle electric savings claims or for demand or gas, as programs did not provide the project-level savings and measure life information needed for those calculations. The NTGRs range from 50% to 100% across PAs. Looking at individual programs, NTGRs vary from year to year. For example, the MCE Commercial Market Efficiency program had an NTGR of 100% in 2022, 67% in 2023, and 78% in 2024. Note that small sample sizes and variance increase the uncertainty of the specific-year estimates. The OBF program had the most consistent NTGRs with 65% in 2022, 68% in 2023, and 74% in 2024.

Table 4-10. Pop NMEC first-year NTGR by program and year, PY2022 – PY2024

Style	PA	Program	PY	Projects	Customers	NTGR*	RP%**
MAP-style	**BayREN**	**Business FLEXmarket**	**2023**	15	14	90%	±4.0%
			2024	16	14	84%	±14.0%
	PCE	**FLEXmarket Program**	**2023**	3	1	50%	±0.0%
			2024	3	3	80%	±0.0%
	MCE	**Commercial Efficiency Market**	**2022**	1	1	100%	±0.0%
			2023	14	8	67%	±14.0%

[41] California Non-residential Net-to-Gross Working Group, *Methodological Framework for Using the Self-Report Approach to Estimating Net-to-Gross Ratios for Non-residential Customers*, (CPUC), February 20, 2015, https://pda.energydataweb.com/api/view/1255/IALC_NTGR_Methods_2013-14.pdf.

Style	PA	Program	PY	Projects	Customers	NTGR*	RP%**
			2024	5	4	78%	±6.0%
	SDGE	**CORE Market Access Program**	**2024**	2	1	67%	±0.0%
			Total	**59**	**46**	**70%**	**±9.0%**
Non-MAP-Style	**PG&E**	**Institutional Partnerships**	**2023**	6	3	63%	±6.0%
			2024	2	2	97%	±0.0%
		NetOne Commercial Efficiency	**2024**	15	4	71%	±21.0%
			Total	**23**	**9**	**72%**	**±20.0%**
OBF	**PG&E**	**On-Bill Financing (OBF)**	**2022**	22	12	65%	±15.0%
			2023	21	8	68%	±9.0%
			2024	37	10	74%	±17.0%
Total				**80**	**30**	**70%**	**±11.0%**

* We used site-level forecasted first-year electric savings in the ratio estimation.
** Relative precision is at the 90% confidence level.

Two MAP-style programs had the highest and lowest NTGRs. BayREN's Business FLEXmarket program had the highest NTGR (90%) and PCE's FLEXmarket program the lowest NTGR (50%). The BayREN program focused on hard-to-reach (HTR) customers, while the PCE FLEXmarket included any commercial customer that could meet NMEC modeling requirements. The HTR customers credited program factors with more influence on their decision-making than the PCE Flexmarket customers did. Additionally, based on our interviews, without the program, the HTR customers would have been less likely than the PCE FLEXmarket customers to implement the same project scope at the same time.

> "The program makes the project more affordable, we can do it sooner and it is good for the environment and expanded the scope to not just change the lights but to add sensors as well."
>
> "It helps small businesses who are looking to be more efficient and cost of energy is ridiculous."
>
> - BayREN customers

> "I wish the program was more lucrative, bigger rebates, I wish the program would incentivize other types of equipment as much as they do lighting."
>
> - PCE customer

4.4.1 Program attribution index results

As noted above, NTGRs are calculated using the same methodology as site-level NMEC. The approach factors in project scope, timing, and program and non-program influences to calculate NTGRs. More information on the individual components and calculations can be found in APPENDIX F.

Table 4-11 describes the three program attribution indicators (PAI_1, PAI_2, PAI_3) and shows the average score for each indicator, broken out by program design. The program attribution indices collectively measure the extent to which program versus non-program factors influenced project decision making (PAI1), whether the project would have occurred absent the program (PAI2), and the program's effect on the project's scope and timing (PAI3). MAP-style programs had the highest PAI scores, and all three program styles had high scores for PAI_2, indicating the importance of program influence on customers' decision to move forward with the project.

Table 4-11. Program attribution index (PAI) results

Program attribution index	Basis	MAP-style	Non-MAP	OBF
PAI_1	Respondents' ratings on the importance of individual program and non-program influences in their decision to implement a project.	8.6	6.6	6.0
PAI_2	Respondents' rating on the importance of program influences collectively relative to the importance of non-program influences collectively on their decision to implement a project.	10.0	8.8	9.6
PAI_3	Respondents' ratings on the likelihood they would have implemented a similar project scope on a similar timeline in the absence of the program.	7.2	6.5	7.0

PAI values in the table are expressed on a scale of 0 to 10.

4.4.2 Participant satisfaction and program feedback

In addition to NTG questions, our survey asked about program satisfaction, program strengths, and suggested areas for improvement.

Interviewers asked, "Who first brought this project to your organization for consideration?" Table 4-12 summarizes respondents' answer to this question. For each type of program, the majority of respondents indicated that they first learned about the opportunity for energy efficiency upgrades included in their projects from program staff. Participants in MAP-style programs were the most likely to indicate that program staff first brought the project to their organization's attention.

Table 4-12. Who first brought project to respondents' organizations' consideration?

Project Initiator	MAP-style	Non-MAP	OBF
Program staff	71%	52%	53%
Non-program vendor	15%	4%	35%
Internal staff	12%	39%	10%
Don't know	0%	0%	2%

During the interviews, we asked participants, "On a scale of 0 – 10, where 0 is 'completely dissatisfied' and 10 is 'completely satisfied', how would you rate your overall satisfaction with the [program]?" Table 4-13 summarizes the response to this question.

MAP-style respondents reported an average satisfaction rating of 8.7. Respondents representing 73% of MAP-style sites were promoters (providing a rating of 8 or above) and respondents representing 14% of MAP-style sites were detractors (providing a rating of 6 or below). Promoters' satisfaction was driven by the incentives that allowed them to make upgrades they would not have been able to make otherwise and the increased energy efficiency of their buildings. One promoter said, "I thought it was too good to be true. Our bill went down 50%." Detractors cited communication challenges with their program vendor. One said, "What was promised [by the contractor] wasn't exactly what we got. Communication wasn't great. They couldn't help us figure out why we weren't saving more."

Non-MAP respondents gave an average rating of 9.5. Respondents representing 91% of non-MAP sites were promoters (providing a rating of 8 or above) and no non-MAP respondents were detractors. As with MAP-style participants, non-MAP promoters' satisfaction was driven by the incentives that allowed them to make upgrades that they would not have been able to make otherwise and the increased energy efficiency of their buildings. One promoter said, "[The program is] extremely important because the incentives make the ROI work."

OBF respondents gave an average rating of 8.3. OBF participants indicated high satisfaction (providing an 8 or above) with all aspects of the program, except for communications with contractors (7.6) and the time it took to review and approve the application (7.9). Detractors cited dissatisfaction with contractors' work quality and challenges communicating with their program vendor. One said, "The contractor installed equipment incorrectly and tried to blame the city."

Table 4-13. Program satisfaction by site

Metric	MAP-style	Non-MAP	OBF
Average satisfaction	8.7	9.5	8.3
% promoters (≥8)	73%	91%	61%
% detractors (≤6)	14%	0%	14%

Following the program satisfaction question, interviewers asked respondents about the strengths of the program. Table 4-14 presents their responses as a percentage of responding sites.

For MAP-style programs the most frequently cited strengths related to the programs' ability to improve financial viability of the project, provide energy and utility bill savings, and improve the facility.

- "[The program] makes the project more affordable, we can do it sooner and it is good for the environment. It expanded the scope to not just change the lights but to add sensors as well."
- "The project was at no cost to me and I'm saving about $300 a month."
- "We had stuff that needed to be done but not as quickly. Lighting is important for the safety of the staff—the program helped expedite the project. Safety was a big factor for us."

For non-MAP programs the most frequently cited strengths were again the improved financial viability of the project and facility improvements but also the ability to conduct a project sooner than would have been possible without the program.

- "[The program had] influence on project approvals because it helps with return-on-investment decision. [The program] is the difference often."
- "The program really influenced [our company] to thoroughly understand what we were getting into. The strength is really understanding the needs on the ground and not pigeonholing with solutions. It allows for a wide array of solutions to be wrapped into the projects."
- "The positive part is that someone came and did the project all at once. What would have taken years was weeks."

Table 4-14. Program strengths, percentage of sites – MAP-style and Non-MAP

Strength	MAP-style	Non-MAP
Improved financial viability for project	56%	61%
Energy and bill savings	44%	0%
Improved facility	19%	48%
Environmental benefits	17%	0%
Quick participation process	12%	26%
Easy application	10%	0%
Allowed for project to be initiated sooner	8%	30%
Good contractors	3%	13%

Interviewers asked OBF participants a different set of satisfaction questions to maintain consistency with previous evaluations, posing a multiple-choice question about outcomes achieved from their project. More than three-fourths (78%)

said the program resulted in lower maintenance costs, 63% said the program resulted in improved user comfort, and 37% said the program resulted in better regulatory or corporate compliance (Table 4-15).

Table 4-15. Program strengths, percentage of sites – OBF

Outcome	Percent of sites
Lower maintenance costs	78%
Improved user comfort	63%
Regulatory or corporate compliance	37%
Greater competitiveness	33%
Increased productivity	22%
Other	4%

The MAP-style and non-MAP interviews included questions asking respondents for suggested improvements, but the OBF instrument did not include a similar question due to limitations with survey length. While 49% of MAP and 30% non-MAP respondents had no suggestions for program improvements, those who did have suggestions most frequently requested increased incentives, more program or contractor feedback and support after installation, and more communication from installers during implementation. Notably, respondents representing 35% of non-MAP sites suggested streamlining program participation processes while no MAP-style respondents provided a similar recommendation (Table 4-16).

MAP-style respondents told us:

- "Provide long term commitment from a vetted installer that can follow up on issues and warranty after project completion."
- "Better communication on how the incentives are paid back to the program and what our liability or exposures look like."
- "The third parties need to improve communication, attention to detail, and quality checks."

Non-MAP respondents told us:

- "We've had stellar third parties and third parties that have no idea what they're doing. My suggestion would be it would be good to have one program serve all industries as opposed to one serving agricultural [and] one serving industrial…because lighting is not industry specific."
- "Make more money available—we would love to do more and faster—we are now looking at a solar program."

Table 4-16. Suggestions for program improvements, percentage of sites

Improvement	MAP-style	Non-MAP
Higher incentive amount	17%	30%
Feedback and support after installation	15%	0%
More installer communication	8%	0%
Better communication of payout and risk	5%	0%
Consumer advocacy	2%	0%
Quicker approval process	2%	4%
More measures	2%	0%
Streamline the application process	0%	35%

4.5 Total savings

Table 4-17 presents the total evaluated savings for non-residential pop NMEC programs with initial claims made in PY2022 and PY2023 and trued up in PY2024. The table pulls together the results from Sections 4.1 and 4.4 for final savings claims and presents the final evaluated net savings.

Table 4-17. Total savings, PY2022 – PY2023

Style	PA	Program	PY	Claimed*	GRR††	Verified†	NTGR**	Net
Electric (kWh)								
MAP-style	BayREN	Business FLEXmarket	2023	2,024,632	78%	1,587,591	90%	1,428,673
	PCE	FLEXmarket Program	2023	1,393,323	87%	1,208,136	50%	604,068
	MCE	Commercial Efficiency Market	2022	852,728	162%	1,382,486	100%	1,382,486
			2023	4,048,775	79%	3,198,083	67%	2,127,365
	SCP	SCP FLEXmarket	2023	1,014,304	0%	0	N/A	0
Total				**9,333,763**	**79%**	**7,376,297**	**75%**	**5,542,593**
Non-MAP Style	PG&E	Institutional Partnerships	2023	743,319	65%	486,840	63%	308,218
Total				**743,319**	**65%**	**486,840**	**63%**	**308,218**
OBF	PG&E	On-Bill Financing (OBF)	2022	11,706,659	96%	11,282,202	65%	7,376,304
			2023	6,135,779	85%	5,216,773	68%	3,558,361
Total				**17,842,438**	**92%**	**16,498,975**	**66%**	**10,934,665**
Demand (kW)								
MAP-style	BayREN	Business FLEXmarket	2023	280.0	79%	220.6	90%	198.5
	PCE	FLEXmarket Program	2023	159.1	22%	35.6	50%	17.8
	MCE	Commercial Efficiency Market	2022	0.0	N/A	304.4	100%	0.0
			2023	465.5	3%	15.5	67%	10.3
	SCP	SCP FLEXmarket	2023	115.8	0%	0.0	N/A	0.0
Total				**1,020.3**	**56%**	**576.1**	**39%**	**226.6**
Non-MAP Style	PG&E	Institutional Partnerships	2023	127.9	38%	48.2	63%	30.5
Total				**127.9**	**38%**	**48.2**	**63%**	**30.5**
OBF	PG&E	On-Bill Financing (OBF)	2022	1,394.8	53%	746.0	65%	487.7
			2023	564.0	92%	518.0	68%	353.3
Total				**1,958.8**	**65%**	**1,264.0**	**67%**	**841.1**
Gas (therm)								
OBF	PG&E	On-Bill Financing (OBF)	2022	54,792	0%	0	65%	0
			2023	9,347	312%	29,186	68%	19,908
Total				**64,139**	**46%**	**29,186**	**68%**	**19,908**

*Claimed savings are the sum of initial and true-up claims made by the programs in the tracking data.
†Verified savings are the evaluated program savings.
††Gross realization rate (GRR) compares verified savings with the claimed savings.
**We used site-level projected first-year electric savings in the ratio estimation

4.6 Total System Benefit

Total System Benefit (TSB) is a dollar-based metric that now serves as the primary objective for energy efficiency portfolios, reflecting an explicit policy focus on delivering savings during high-value hours. Although PAs have reported TSB in the CEDARS database for several years, TSB did not become the official primary program goal until PY2024. Because NMEC claims span two program years, the cohorts included in this evaluation submitted initial claims in PY2022 and PY2023 and were trued-up in PY2023 and PY2024. As a result, this first evaluation of TSB reflects a transition period. Looking ahead, the meter-based nature of NMEC allows future project years to apply empirical cohort-specific, meter-derived hourly impact shapes through the custom impact shapes tool, enabling TSB calculations to better reflect actual savings by hour across the year.

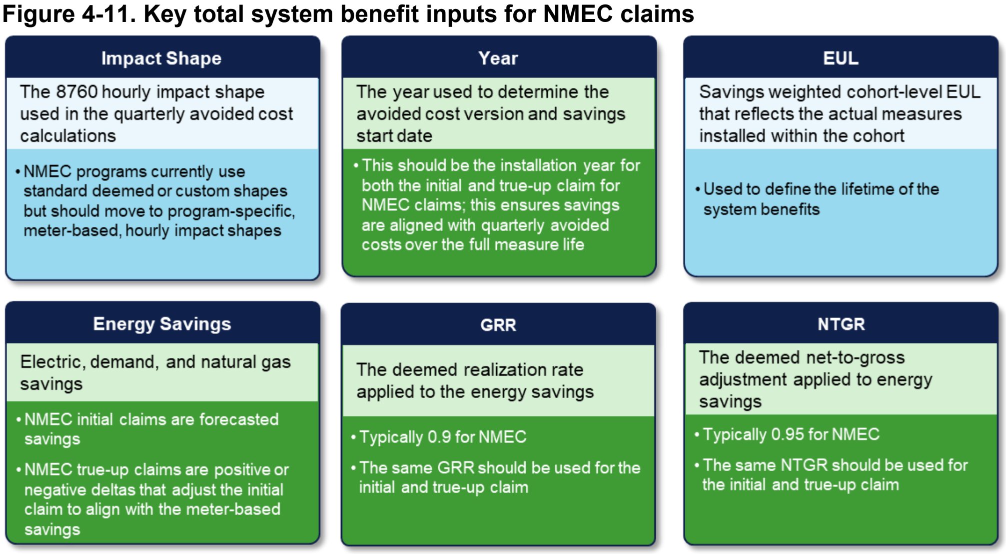
Table 4-18 presents the TSB results for PY2022 and PY2023. The TSB values in the table reflect the combined effects of multiple inputs. Our verified TSB results account for a number of these, including verified gross savings and NTGRs. Because our evaluation occurred during the transition to TSB, we did not adjust the verified value using meter-based savings shapes, which serve as another input to the TSB calculation.

Table 4-18. Total system benefits, PY2022 - PY2023

Style	PA	Program	PY	Claimed	TSB RR	Verified
MAP-style	BayREN	Business FLEXmarket	2023	1,651,356	96%	1,586,707
	PCE	FLEXmarket Program	2023	1,360,538	51%	699,962
	MCE	Commercial Efficiency Market	2022	852,310	183%	1,560,545
			2023	4,004,183	60%	2,391,104
	SCP	SCP FLEXmarket	2023	1,156,574	0%	0
Total				**9,024,961**	**69%**	**6,238,317**
Non-MAP Style	PG&E	Institutional Partnerships	2023	729,692	49%	355,475
Total				**729,692**	**49%**	**355,475**
OBF	PG&E	On-Bill Financing (OBF)	2022	7,326,222	99%	7,265,364
			2023	9,904,225	40%	3,989,012
Total				**17,230,448**	**65%**	**11,254,376**

Figure 4-11 shows the six inputs used in TSB calculations. PAs enter these inputs into the CPUC's Cost Effectiveness Tool (CET) when making their annual claims. For this evaluation, we updated the four inputs shown in green. Future evaluations should update all six inputs, including the impact shape and EUL. With the introduction of the custom impact shape tool in PY2025, PAs can use empirical savings shapes for future pop NMEC claims, enabling future evaluations to update the shapes. Future evaluations will also need more information about how PAs select EULs in order to evaluate those choices.

Figure 4-11. Key total system benefit inputs for NMEC claims



We found that under the current claims approach, PAs used different assumptions for initial and true-up claims, which affected TSB results in ways that were likely not intentional. This section addresses those impacts for each TSB input.

Impact shapes. Pop NMEC programs will transition to empirical cohort-specific, meter-based hourly impact shapes. However, for PY2022 – PY2024 TSB claims, PAs relied on existing DEER and Custom load shapes. Table 4-19 identifies the load shapes used for both initial and true-up claims. Three cohorts applied different impact shapes across the two claims, effectively mixing shapes in ways not intended by the underlying methodologies and producing problematic results. PG&E applied multiple impact shapes within a single program, likely to approximate the savings shapes of the mix of measures installed through the program.

Table 4-19. Load shape and EUL changes between initial and true-up claims

PA	Program	PY	Electric end use shape	Initial claim	True-up claim	Same shapes used for initial and true-up?
BAY	Business FLEXmarket	2023	1 = Commercial Indoor Lighting	1,142,277	0	No
			DEER:Indoor_CFL_Ltg	0	882,354	
MCE	Commercial Efficiency Market	2022	DEER:RefgFrzr_Recycling	1,082,699	0	No
			1 = Commercial Indoor Lighting	0	-229,971	
MCE	Commercial Efficiency Market	2023	1 = Commercial Indoor Lighting	4,077,687	0	No
			DEER:Indoor_CFL_Ltg	0	-14,456	
			DEER:HVAC_Eff_HP	0	-14,456	
PCE	FLEXmarket Program	2023	DEER:Indoor_CFL_Ltg	1,052,161	341,162	Yes
PGE	On-Bill Financing (OBF)	2022	DEER:Indoor_Non-CFL_Ltg	22,140,968	-15,705,345	Yes
			DEER:HVAC_Split-Package_AC	6,000,581	-729,545	
PGE	On-Bill Financing (OBF)	2023	DEER:Indoor_Non-CFL_Ltg	18,672,042	-14,481,065	Yes
			DEER:HVAC_Split-Package_AC	4,098,825	-2,154,023	
SCP	SCP FLEXmarket	2023	DEER:Indoor_CFL_Ltg	335,253	0	N/A**
			DEER:HVAC_Chillers	598,151	0	
			DEER:HVAC_Split-Package_HP	80,900	0	
SW	Institutional Partnerships	2023	2 = Commercial Outdoor Lighting	752,164	-8,845	Yes

* MCE had multiple true-up rows with a range of EULs.
** SCP did not true-up the program claims.

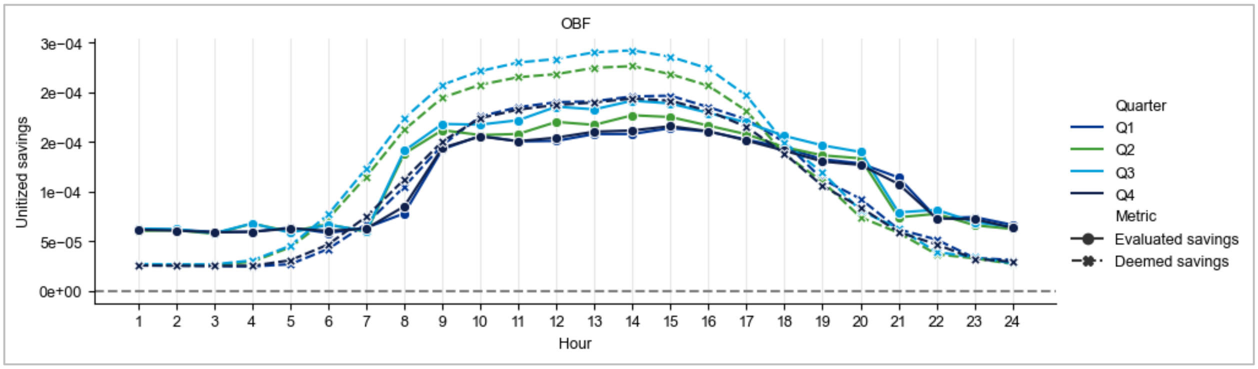
Year. Each cohort had at least two claims: initial claimed savings one year and the true-up claim the following year. This distinct feature of NMEC pathways poses challenges to TSB calculations. When replicating the TSB values in CEDARs, we found that the PAs used different avoided cost years for the initial and the true-up claims. Effectively, this means that if a true-up energy savings claim sought to zero out the initial claim, the TSB metric would not be zeroed out because the avoided costs would be different. To address this issue, we used the initial year as the basis of the verified TSB.

EUL. Given that pop NMEC programs claim savings at the cohort rather than the project level, and less project documentation is required, we were unable to verify the claimed EULs. Additionally, the PAs claimed a range of EULs for the same cohorts, sometimes claiming different EULs for the initial and true-up claims. This had the effect of changing the lifetime of the system benefits only for the true-up delta and not for the entire program claim. We used the initial claim EULs when calculating verified TSB in order to produce internally consistent results.

Other inputs. Because we used the verified energy savings produced in this evaluation when producing the verified TSB results, we set the GRR to 1. Additionally, we also used the NTGR from this study when updating the results.

Similarity to deemed shapes. To compare the DEER (deemed) savings shapes used by PAs in their TSB claims with empirical (verified) savings shapes, Figure 4-12 presents a case study of the PY2023 PG&E OBF program. The figure compares empirical hourly savings with the deemed savings shape. We generated the deemed shape by calculating a savings-weighted average of the end-use shapes listed in Table 4-19 used by the OBF program in its initial claim. While there is some alignment between the deemed and empirical shapes for this population, the deemed shape overstates midday savings and understates early evening and overnight savings relative to observed performance. In particular, the empirical shapes show savings persisting into peak afternoon hours, which is not reflected in the deemed shapes.

Figure 4-12. Comparison of PY2023 OBF empirical and DEER savings shapes



4.7 Participant profiles and system value

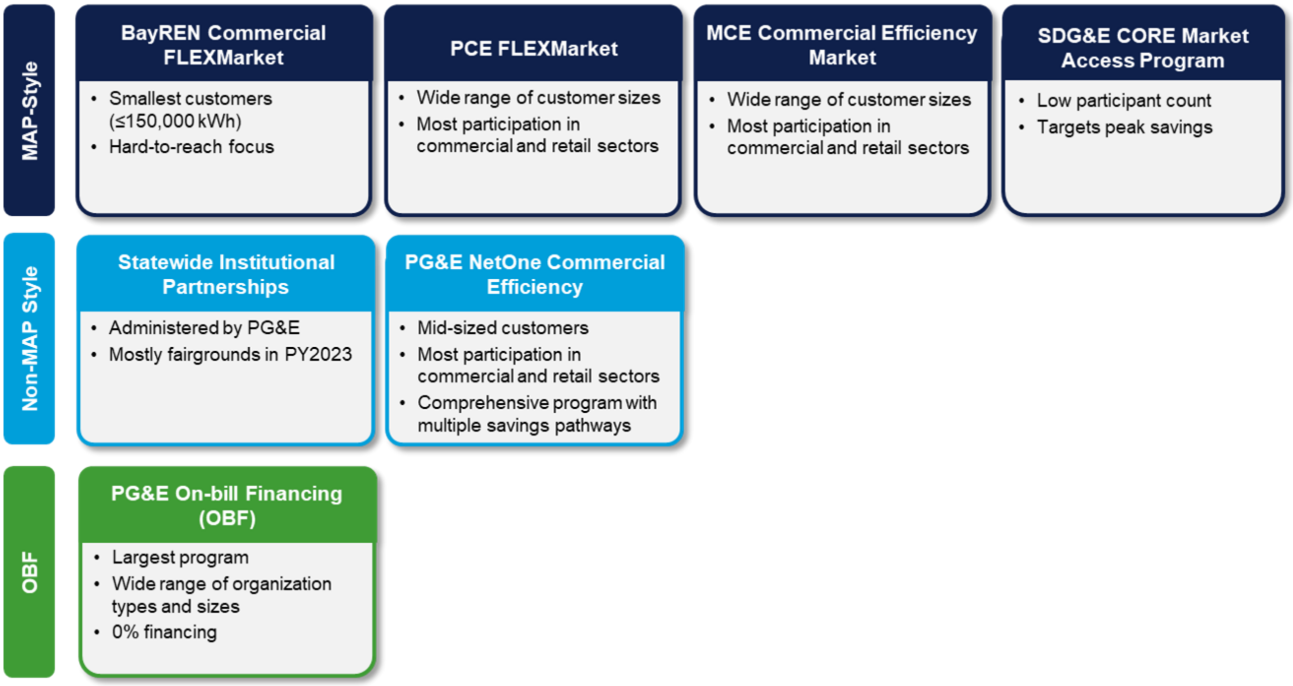
Pop NMEC, like site-level NMEC, is designed to unlock the potential savings stranded when customers maintain equipment beyond its EUL. Traditional energy efficiency programs that claim savings based on calculated results generally assume a hypothetical, up-to-code baseline. Incentives are calculated based on the above code savings, so customers operating below that baseline are not incentivized to upgrade to that baseline and may not be able to afford to upgrade beyond it. Because NMEC measures savings against a pre-intervention, existing conditions metered baseline, customers can receive credit for any improvement in efficiency, including upgrades that simply bring equipment up to code. This can be especially valuable for customers who are unable to raise capital or lack technical support for site improvements. Helping customers make these kinds of upgrades is how NMEC programs motivate energy savings that otherwise would not have happened.

We examined which customers NMEC programs target and how savings are distributed among them to assess implications for stranded savings and the application of program eligibility screening. This section presents findings on the distribution of savings across programs and customers, as well as the timing of those savings, to assess implications for system value. Results show which customers pop NMEC programs serve, how benefits accrue across participants, and when savings contribute most to the grid, shedding light on how program design and participant characteristics drive differences in savings outcomes.

Throughout this section, we include data from as many program years as possible for each analysis. For firmographic analyses, we draw on data from PY2022 through PY2024, as firmographic information is available for all cohorts. Analyses that rely on verified savings results are limited to PY2022 and PY2023.

As discussed throughout this report, the evaluated programs adhere to three primary designs: MAP-style, non-MAP style, and OBF. Figure 4-13 provides an overview of the unique features of the different programs included in each program type.

Figure 4-13. Program design overview



Note: We excluded SCP's program from this figure as we were unable to verify any savings for the program.

In the sections that follow, we investigate the distribution of key firmographics and summarize any differences in savings performance. Additionally, APPENDIX G presents a summary of customer profiles, including years in operation, number of employees, and square footage, by program.

4.7.1 Industry segments

Analyzing the primary activity at each site differentiates the customer types served across programs and supports our assessment of the application of eligibility criteria. Given that general NMEC methods assume that energy efficiency improvements can be isolated using whole-facility metered data and that pop NMEC assumes that no site-level model tailoring is needed, we were particularly interested in any performance patterns across industry segments. The NMEC rulebook requires that non-residential program customers be commercial or commercial like.[42] It additionally requires that program populations consist of sites that "can reasonably be expected to have similar types of equipment holdings, as well as drivers and levels of energy consumption."

Table 4-20. breaks down each program's participating customers by organization type and industry classification (see Section 3.3 for details on how we classified sites). While commercial, retail, and institutional NAICS classifications suggest alignment with the NMEC rulebook's "commercial-like" requirement, the NAICS classifications are designed for broader classification purposes and may not fully capture the specific intent of NMEC eligibility. As a result, the NAICS code can include sites that do not align with NMEC requirements (for example, fairgrounds and hotels are in the same NAICS grouping, with the former having irregular activities and load patterns while the latter do not). Additionally, the reported NAICS codes may be inaccurate or outdated, which could affect the accuracy of these classifications, and participant size can vary substantially within the same industry group (e.g., commercial and retail), introducing further uncertainty. Eligibility

[42] CPUC, *Rulebook for NMEC, 2.0*, p. 12, https://www.cpuc.ca.gov/-/media/cpuc-website/files/legacyfiles/n/6442463694-nmec-rulebook2-0.pdf.

should therefore be validated using more detailed site-level information available through screening and site visits, where possible.

Table 4-20. Industry classifications of pop NMEC program participants, PY2022–PY2024

Organization type/ industry	MAP-style				Non-MAP style		OBF
	BayREN Commercial FlexMarket (n=327)	MCE Commercial Efficiency Market (n=76)	PCE Commercial Efficiency Market (n=11)	SDGE Core Market Access (n=3)	PG&E NetOne (n=71)	PG&E IP (n=21)	PG&E OBF (n=982)
Commercial and retail	**86%**	**87%**	**100%**	**67%**	**92%**	**86%**	**48%**
Commercial & retail	80%	49%	82%	67%	48%	0%	22%
Distribution	1%	7%	0%	0%	13%	0%	4%
Office & information services	5%	25%	18%	0%	28%	0%	11%
Hospitality & recreational services	0%	7%	0%	0%	3%	86%	12%
Institutions	**2%**	**11%**	**0%**	**33%**	**1%**	**14%**	**35%**
Institutional & public services	2%	9%	0%	0%	1%	14%	29%
Health care facilities	0%	1%	0%	33%	0%	0%	6%
Industrial, agricultural, and utility	**12%**	**3%**	**0%**	**0%**	**7%**	**0%**	**17%**
Industrial & production	11%	1%	0%	0%	4%	0%	12%
Infrastructure & utilities	1%	1%	0%	0%	3%	0%	5%

MAP-style programs. All MAP-style programs had more participating customers in the overarching commercial and retail sector than the institutional sector. Within the overarching commercial and retail category, large shares of customers in the BayREN Commercial FlexMarket (80%, n=327) and PCE Commercial Efficiency Market (82%, n=11) programs were commercial & retail. MCE's Commercial Efficiency Market program also had a large share of office & information services (25%) customers in addition to commercial & retail (49%). MCE's program enrolls the broadest range of customers among the MAP-style programs, reaching all sectors. Programs with small participant counts, namely PCE's Commercial Efficiency Market (n=11) and SDGE's Core Market Access (n=3) programs, reflect case-specific enrollment rather than broad market coverage.

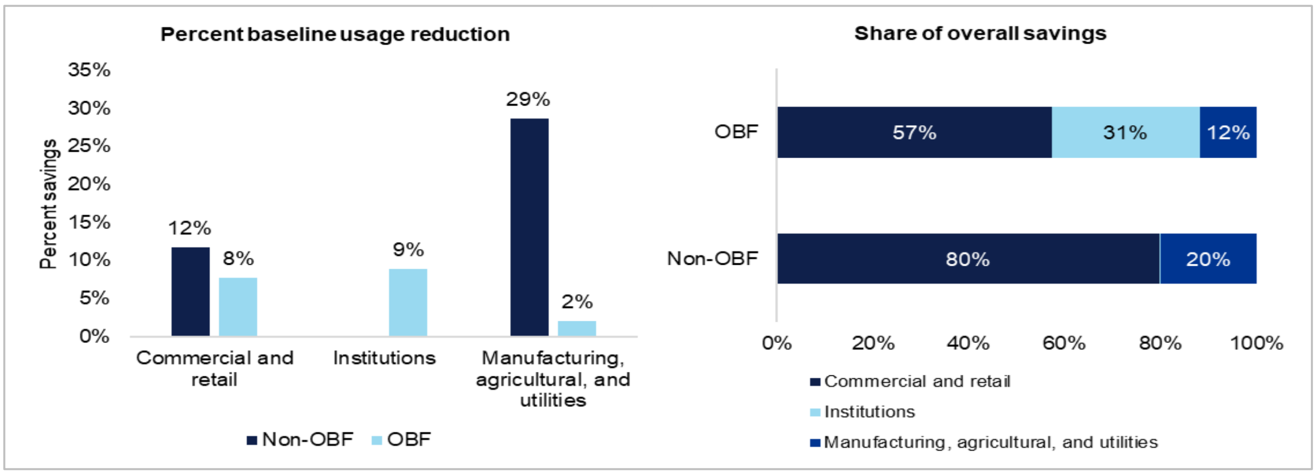
Non-MAP style programs. PG&E administers the statewide Institutional Partnerships program, which is open to state-owned facilities, serving hospitality & recreational services as well as institutional & public services customers. Prior to PY2024, only fairgrounds within the hospitality & recreational services sector participated in this program. Beginning in PY2024, additional state agencies in the institutional and public services sector began enrolling.

OBF. PG&E OBF (n=982) serves the highest number and the broadest mix of customers across all sectors. It has the highest representation of institutional & public service customers at 25%, a segment largely absent from most other programs. In this program, commercial & retail customers are only 21% of the total enrollment.

Savings distribution. Characterizing participant composition allowed us to evaluate how different organization types contributed to overall savings. Commercial and retail businesses had the largest share of overall savings and exhibited relatively high percent savings. Figure 4-14 shows the percentage of baseline usage reduction (left) and the overall savings by organization type (right). Because the OBF program is substantially larger and distinct from the others, we report OBF results separately.

For OBF, industrial, agricultural, and utility businesses, a sector that should not be present in NMEC programs, accounted for 12% of overall savings. For the non-OBF programs, most savings came from commercial and retail businesses (80%), with the remainder coming from the problematic industrial, agricultural, and utility (20%) sector.[43] The industrial, agricultural, and utility participants showed very high savings at 29% of baseline usage, though these estimates may reflect non-program-related variability rather than true program impacts, which is why these sectors are considered problematic for pop NMEC. This group was relatively small, composed of just 21 sites. All but two of the sites were from BayREN's PY2023 cohort and were labelled as manufacturing, construction, and transportation in the disaggregated data.

Figure 4-14. Percentage of baseline usage reduction and share of savings by organization type, PY2022 – PY2023



While the NAICS codes do not necessarily provide sufficient information to make definitive statements about whether or not the industrial, agricultural, utility and institutional customers qualify as commercial or commercial-like under the NMEC rulebook, these codes do raise serious questions about the NMEC eligibility of those customers. Industrial, agricultural, and utility businesses are more likely to have process loads that cannot be appropriately modeled at the population level. The challenge of assessing eligibility runs deeper, however. Sub-groups within commercial and retail (e.g., fairgrounds) may also represent sites ill-suited for NMEC methods. Detailed information or site visits could reduce this risk, allowing programs to screen sites for NMEC eligibility, taking into consideration the business type, presence of process load, and other factors. This approach could be further strengthened by applying more refined NAICS definitions or clustering methods to ensure eligibility is based on how typical a site is within its sector, rather than broad classifications alone.

4.7.2 Annual electricity usage

Because pop NMEC eligibility depends on grouping customers with similar energy consumption, we examined the distribution of annual baseline electricity usage within and across programs. One of the clearest sources of guidance on pop NMEC customer eligibility in the NMEC rulebook concerns the similarity of customers grouped into a program cohort. The rulebook states that sites should "be expected to have similar... levels of energy consumption." Our analysis also mapped large customers to sectors that are not commercial or commercial-like and do not share similar energy-use drivers with most customers.

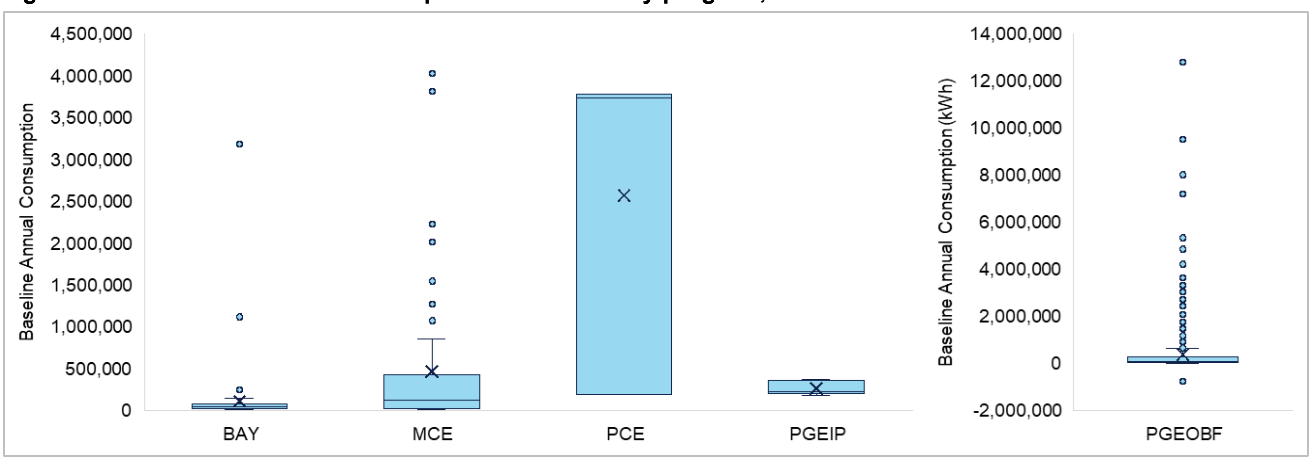
[43] Only one non-OBF project was completed by an institutional participant and so we exclude non-OBF institutions from the figure.

Table 4-21 presents the distribution of participants by annual electric usage size for each program. Figure 4-15 presents the baseline annual electricity usage in the form of a box plot. The box represents the interquartile range (25th to 75th percentile), the line inside the box is the median, the whiskers show the typical minimum and maximum values, and any points beyond the whiskers indicate outliers. Participation levels differ markedly across each electric usage size bin, highlighting distinct program profiles, though most programs are highly skewed. Overall, 125 out of 1474 sites from the PY2022 – PY2024 cohorts fell into the large electric usage bin.

Table 4-21. Annual electric usage for pop NMEC program participants, PY2022 – PY2024

Annual electric usage size	MAP-style				Non-MAP style		OBF
	BayREN Commercial FlexMarket (n=326)	MCE Commercial Efficiency Market (n=75)	PCE Commercial Efficiency Market (n=11)	SDGE Core Market Access (n=3)	PG&E NetOne (n=57)	PG&E IP (n=21)	PG&E OBF (n=981)
Small: Less than 150,000 kWh	98%	43%	55%	33%	23%	76%	70%
Medium: 150,000 – 999,999 kWh	1%	39%	9%	0%	40%	24%	22%
Large: 1,000,000 kWh or more	1%	19%	36%	67%	37%	0%	8%

Figure 4-15. Baseline annual consumption distribution by program, PY2022 – PY2023



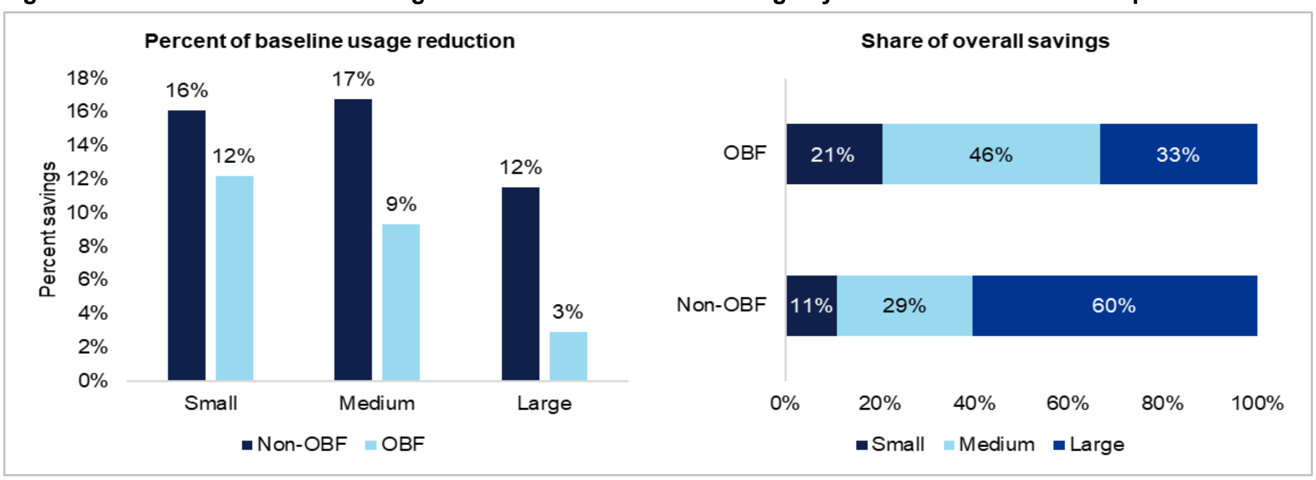
MAP-style programs. While small customers represented 98% of BayREN Commercial FlexMarket participants (n=326), the program had a few sites with much higher annual consumption than the rest of the population. The largest site had a manufacturing NAICS code and a baseline average annual consumption of over 3 million kWh. MCE's customers were evenly split between the small and medium usage bins, but also included several large sites, with a handful of outliers. The largest participant for MCE had a retail trade NAICS code and an average annual consumption of over 4 million kWh. PCE's program only had 11 participants, with over half being small and over a third large. SDGE's program is not shown in Figure 4-15 because it did not have any PY2022 or PY2023 cohorts.

Non-MAP style programs. The Institutional Partnerships program was primarily comprised of small sites, with an additional quarter of medium participants. The NetOne program participants tended to be larger, with medium sized customers making up the largest share at 40%. PG&E's NetOne program is not shown in Figure 4-15 because it did not have any PY2022 or PY2023 cohorts.

OBF. While small customers represented 70% of OBF's participants, the distribution of annual baseline consumption for OBF is highly variable, with many very large outliers. The largest site had an agricultural NAICS code and a baseline average annual consumption over 12 million kWh. The next largest site had a utilities NAICS code and a baseline average annual consumption of over 9 million kWh.

Savings distribution. Having established this distribution of electric consumption levels, we next examined how these differences manifested in energy savings outcomes. Figure 4-16 shows the percentage reduction in baseline usage per site (left) and the share of overall savings (right), grouped by small, medium, and large customers using the annual energy consumption bins defined in Table 4-5 and again reporting OBF separately given its distinct program size and characteristics. Smaller customers achieved higher savings rates, reducing a larger share of their baseline consumption than large customers, with small and medium customers achieving higher percentage reductions overall. However, because of the disproportionate size of large customers, this pattern is reversed when considering size group savings as a percentage of total savings. Although large customers represented only 8% of participants and had lower percentage usage reductions, they accounted for the largest share of overall savings (33% for OBF and 60% for non-OBF). These large sites may represent good energy efficiency program participants, but they are not good fits for pop NMEC. Industrial, agriculture, and utility customers should be able to pursue their efforts in CIAC or SEM programs, while the large customers that are commercial or commercial-like could pursue the site-level NMEC pathway. The savings produced by these large customers are too big and too risky for the low scrutiny environment of pop NMEC.

Figure 4-16. Percent of baseline usage reduction and share of savings by customer annual consumption size*

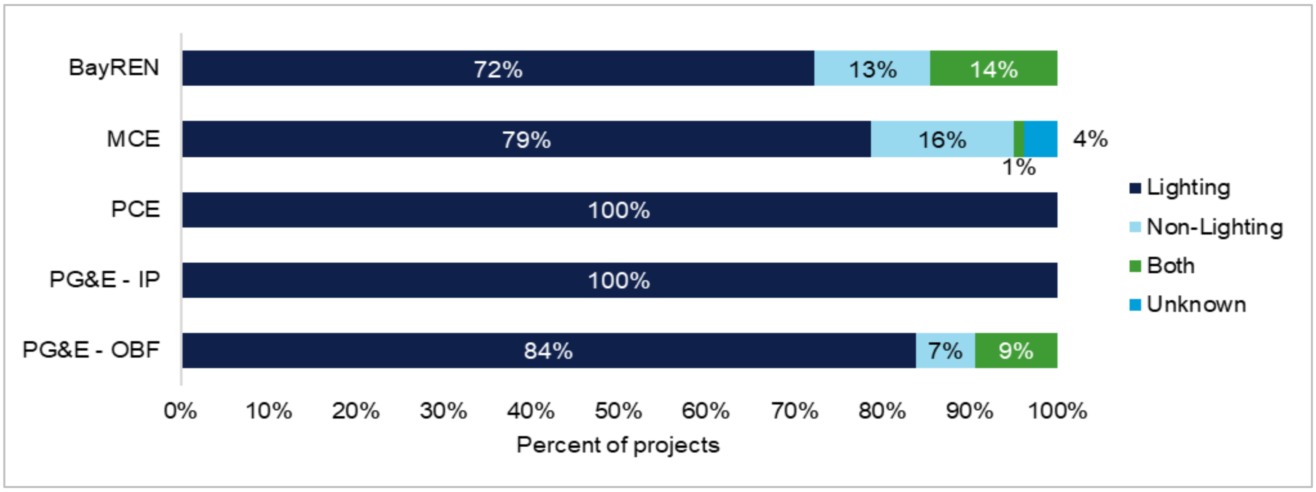


* This figure does not show sites that received assigned savings.

4.7.3 Measure type

This section examines the distribution of measure types and savings by program. Pop NMEC's whole-building, meter-based approach is well-suited for comprehensive energy-efficiency projects. However, most programs were dominated by lighting, and detail in measure descriptions varied substantially, with multiple programs only installing lighting measures. Figure 4-17 shows the measure type distribution by program for PY2022 – PY2023 cohorts. BayREN's program had the largest share of projects that included at least some non-lighting measures at 27%.

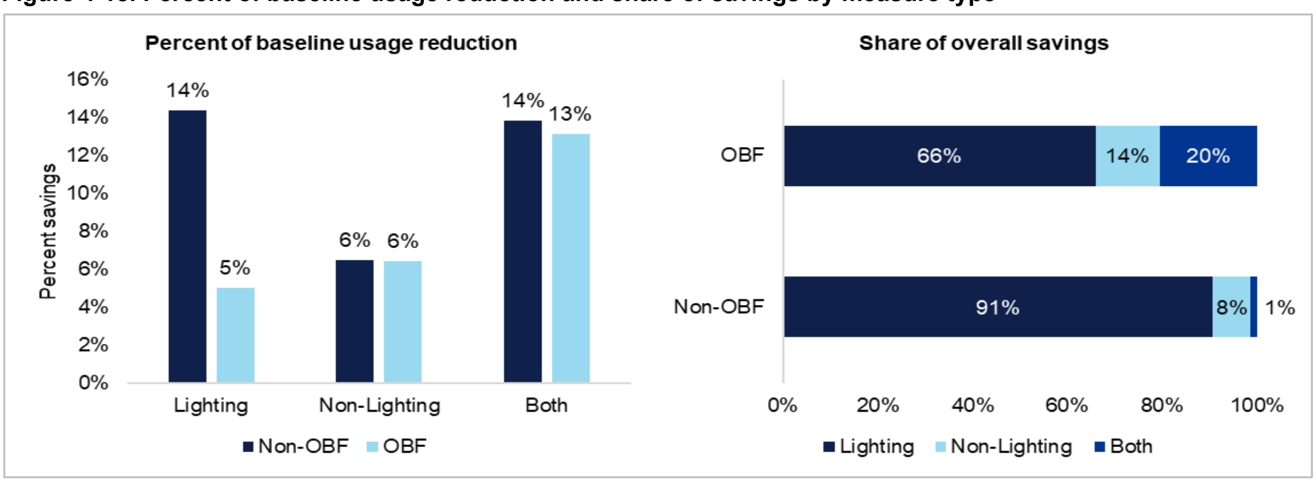
Figure 4-17. Measure type distribution by program, PY2022 – PY2023



Savings distribution. Figure 4-18 shows the percentage reduction in baseline usage (left) and the share of overall savings by measure type (right) for OBF and non-OBF programs. Per-site consumption reductions varied by measure mix. For OBF, sites that installed a combination of lighting and non-lighting measures (both) achieved the highest average reduction in baseline usage (13%), compared to lighting-only projects (5%) and non-lighting projects (6%). These results suggest that while lighting drives total OBF savings due to its prevalence and scale, mixed measure projects deliver greater usage reductions at the site level.

For the non-OBF programs, lighting-only projects and projects with a combination of lighting and non-lighting measures both saved 14% of baseline consumption. The higher lighting-only savings rate here could indicate lighting stranded savings, a plausible conclusion given that the BayREN program in particular targets HTR customer segments

Figure 4-18. Percent of baseline usage reduction and share of savings by measure type



While almost no participants called for an expansion of the number or types of measures available through the programs (see Section 4.7.2 and Table 4-16), the measure type summary presented here indicates that non-residential pop NMEC programs are mostly not completing comprehensive whole-building, deep retrofits but are instead operating largely as lighting programs.

4.7.4 Climate zones

While climate zone does not directly inform the question of stranded savings or eligibility, it is an important factor in savings potential as it directly influences site energy consumption. We categorize climate zone based on available location information and official California climate zones. BayREN and MCE serve customers adjacent to the Bay Area, but the eastern portions of their coverage areas are in climate zone 12, which is considered Central Valley. PCE serves one town in climate zone 12. PG&E, on the other hand, has substantial numbers of customers throughout the Central Valley. A quarter of the OBF participants are in zone 13, the southern portion of the Central Valley.

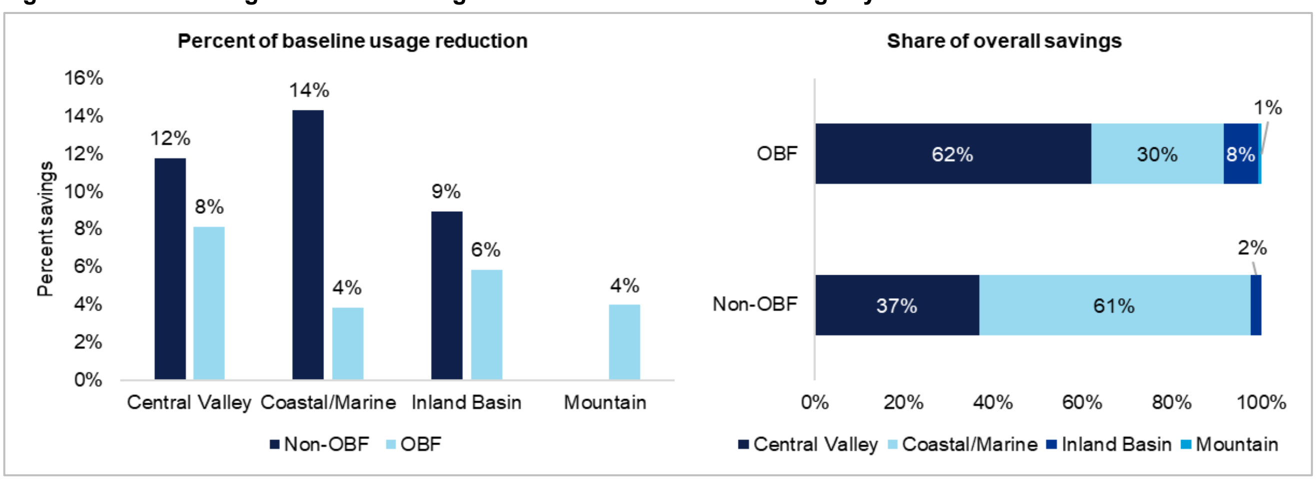
Across all programs, participants are concentrated primarily in the Central Valley and Coastal/Marine climate zones. OBF is the only program with participants located in the Mountain climate zone. No participants across any of the non-residential pop NMEC programs are located in the Desert climate zone (Table 4-22).

Table 4-22 Climate zone for pop NMEC program participants, PY2022 – PY2024

Climate zone	MAP-style				Non-MAP style		OBF
	BayREN Commercial FlexMarket (n=327)	MCE Commercial Efficiency Market (n=76)	PCE Commercial Efficiency Market (n=11)	SDGE Core Market Access (n=3)	PG&E NetOne (n=71)	PG&E IP (n=21)	PG&E OBF (n=982)
Central Valley	31%	59%	9%	0%	42%	76%	52%
Coastal/Marine	56%	41%	91%	33%	24%	5%	38%
Inland Basin	13%	0%	0%	67%	34%	19%	8%
Mountain	0%	0%	0%	0%	0%	0%	1%

After examining how participants are distributed across climate zones, we explored how these geographic differences relate to energy savings outcomes. Figure 4-19 shows the percentage reduction in baseline usage per site (left) and the share of overall savings by climate zone (right).

Figure 4-19. Percentage of baseline usage reduction and share of savings by climate zone



MAP-style and non-MAP style. For MAP-style programs, overall, customers are concentrated in the coastal and marine climate zones, though MCE had a substantial share of customers in the Central Valley, where both overall savings and percent reductions in baseline usage are higher. For the non-MAP style participants, most participants were in the Central Valley climate zone, though a substantial number of participants were also in the Inland Basin and Coastal/Marine climate zones.

OBF. For OBF, both the largest share of overall savings and the highest percentage reductions in baseline usage came among customers in the Central Valley. Central Valley climate zones accounted for 62% of OBF savings and represented roughly half of the participating customer population. Customers in the Central Valley also achieved higher per-site reductions in baseline usage (8%) than participants in other climate zones. The Inland Basin zone exhibited slightly smaller percentage reductions in baseline usage (6%), but accounted for only 8% of OBF savings and participating sites. Customers in the Mountain zone achieved lower percent reductions in baseline usage (4%), though this result is based on a small sample of 13 participants.

4.7.5 Program

Finally, we examined how the savings vary across the programs within the context of the participant profile summaries presented in Sections 4.7.1 to 4.7.4.

Figure 4-20 shows the percentage reduction in baseline usage per site by program (left) and each program's share of overall savings (right). At the program level, per-site usage reductions ranged from 6% for OBF to 23% for PG&E IP. However, OBF was the largest program during the evaluation window, with 905 participants across the two program years, compared to a combined total of 175 participants across the other four programs. As a result, OBF accounted for the majority of participants and savings and therefore disproportionately influenced overall program savings.

Figure 4-20. Percent of baseline usage reduction and share of savings by program, PY2022 – PY2023

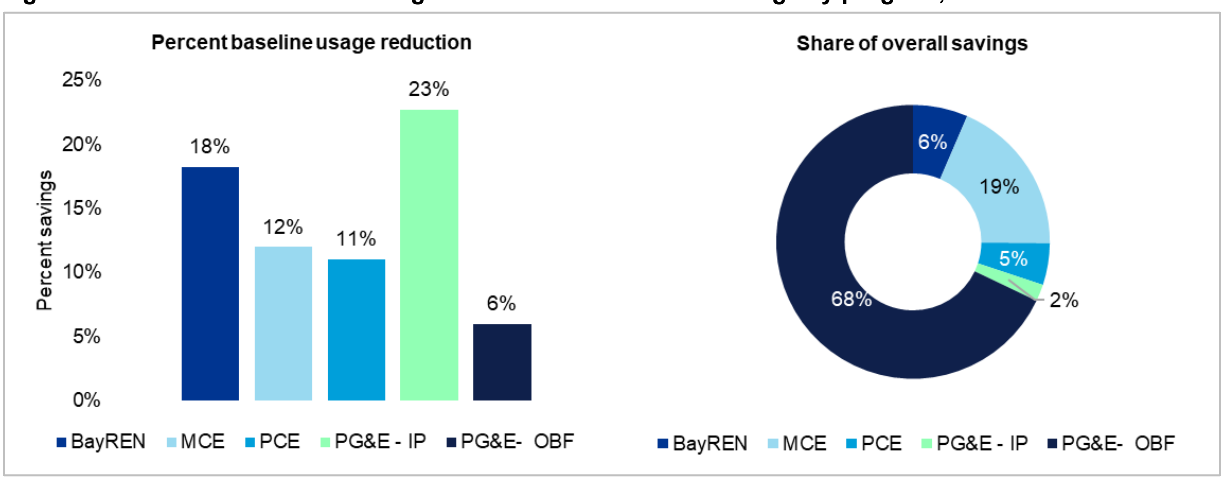
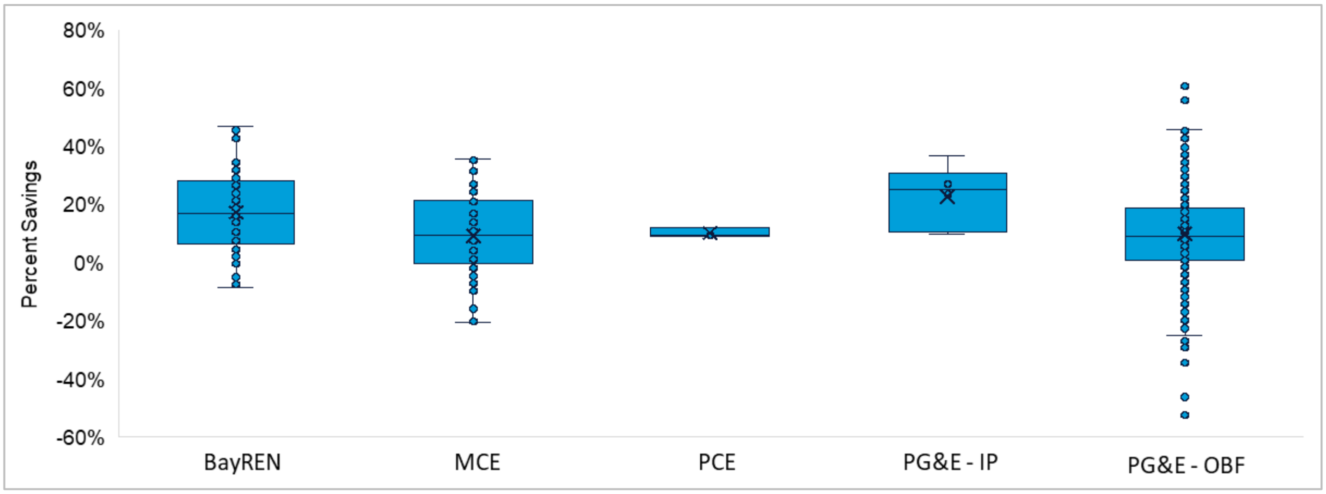


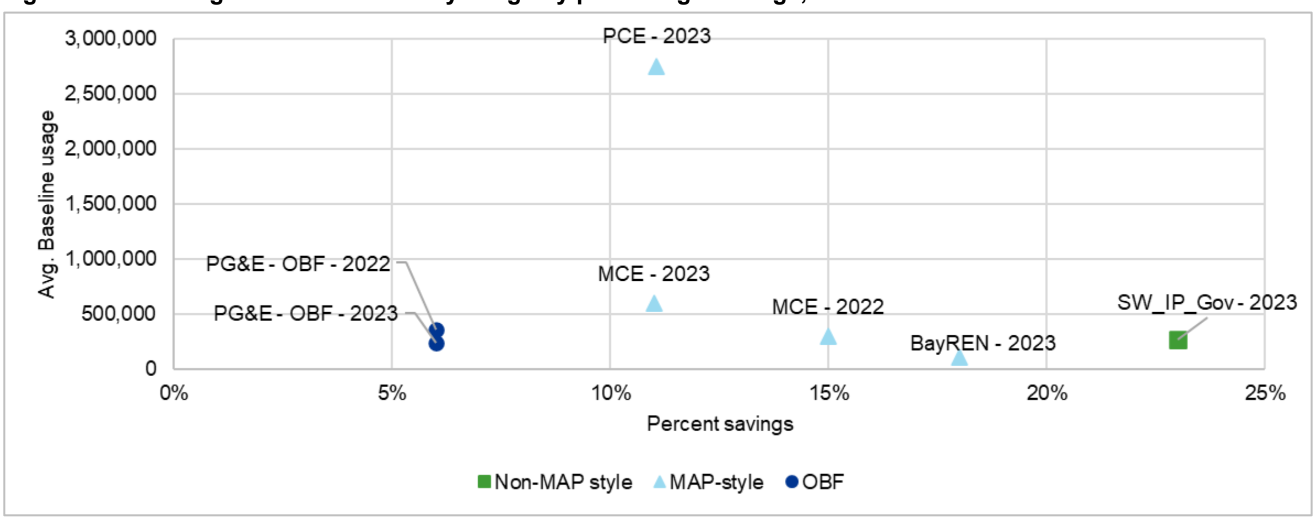
Figure 4-21 shows the savings distribution across programs for PY2022 – PY2023 participants. Positive values represent energy savings, while negative values indicate increased energy usage. OBF showed the widest range of outcomes, including the most outliers and the largest decreases and increases in energy usage, and consequently had the lowest median savings among the programs. Programs with a higher share of medium and large electric consumers, such as MCE and OBF, showed a lower median savings percentage than programs predominantly composed of low electric usage participants, such as BayREN (98%) and PG&E IP (76%).

Figure 4-21. Percentage savings distribution across programs, PY2022 – PY2023



The pop NMEC programs exhibited a wide range of participant sizes, defined by average baseline energy usage, and percentage savings achieved. Figure 4-22 plots the average annual electricity usage and the percentage savings for each cohort with final savings claims (PY2022 and PY2023).[44] The percentage of savings range from 6% to 23%, while the average annual electricity usage ranges from just over 100 thousand kWh to nearly 3 million kWh.

Figure 4-22. Average baseline electricity usage by percentage savings, PY2022 – PY2023 cohorts



BayREN Commercial FLEXMarket program. As noted earlier, BayREN's Commercial FLEXMarket program is a MAP-style program designed to reach small customers (≤150 thousand kWh) with a particular focus on HTR customers who may also be more likely to have stranded savings. The program was the second largest program by the number of participants, after OBF, during the evaluation period. It achieved an 18% reduction in baseline electricity usage, the second highest of any program. Customers participating in the program were mostly from the retail and commercial (86%) sectors, but even so, the program had a substantial share of industrial, agricultural, and utility participants (12%). Program participant average annual electricity usage mostly aligned with the program design, as most participants fell into the small energy usage (<100

44 PG&E NetOne Commercial Efficiency and SDGE CORE Market Access Program did not have any PY2022 or PY2023 cohorts so they are not included in the figure below or in tables or figures comparing savings.

thousand kWh) size bin, with a few notable exceptions. Additionally, BayREN's program had the largest share of projects including measures beyond lighting (27%).

MCE Commercial Efficiency Market. MCE's Commercial Efficiency Market program is a MAP-style program. It was the third largest non-residential pop NMEC program during the evaluation period. The program served a wide range of customer sizes, mostly below 1 million kWh average annual consumption, with notable outliers. MCE projects achieved a 12% baseline reduction on average, the third highest of any program. Most projects were completed at commercial and retail sites (87%), with the largest portion of this group coming from offices & information services (25%). Most projects installed only lighting measures, but the program had the second highest share of projects with some non-lighting components (17%).

PCE Commercial Efficiency Market. The PCE Commercial Efficiency Market is a small (n=11) MAP-style program with a wide range of baseline consumption. The program achieved 11% baseline reduction on average. According to the provided documentation, the program only installed lighting.

SDGE Core Market Access. The SDGE Core Market Access program is a very small MAP-style program (n=3). It was new in PY2024, so no savings were yet ready to be evaluated. The program was designed to target peak savings. Since the program was so small and we do not yet have evaluated savings, we were unable to draw conclusions about the customer profiles within this program.

PG&E NetOne. The pop NMEC pathway within PG&E's NetOne program is a non-MAP offering within a broader portfolio. This pathway was introduced in PY2024, so verified savings are not yet available. The program primarily served commercial and retail sites (92%), with a smaller share in industrial, agricultural, and utility sectors (7%).

PG&E Institutional Partnerships. The Institutional Partnerships program is a non-MAP style statewide program administered by PG&E. While the program was open to a wide range of institutional and public buildings, all of the sites included in this evaluation were fairgrounds that installed lighting. On average, the program achieved 23% baseline reduction.

PG&E OBF. PG&E's OBF program was by far the largest during the evaluation period, accounting for 66% of participants and 68% of savings. The program does not offer performance-based incentives; instead, it offers 0% financing for energy efficiency improvements. While small customers accounted for 70% of OBF's participants, the baseline annual consumption is highly variable, with many very large outliers. The program achieved the lowest average annual baseline reduction in electricity at 6%. Most projects involved lighting only (84%), while 16% included at least some non-lighting measures.

4.7.6 Grid value

Based on the verified hourly savings shapes, programs differed in how much savings aligned with periods of high system cost and carbon intensity, including summer months and late-afternoon peak hours. Overall, none of the programs' peak savings aligned with the DEER-defined late-afternoon peak period (4 PM – 7 PM), even though some programs, particularly OBF, had concentrated savings in higher-value seasons such as summer. The discussion below first examines the seasonality of program-type savings shapes to identify when verified savings occurred throughout the year and then looks at hourly shapes to understand when savings occurred within the day.

Seasonality of savings. Figure 4-23 illustrates the shapes of annualized savings by program type, highlighting differences in seasonal patterns and the regularity of savings throughout the year. These are unitized daily savings shapes representing what percentage of annual savings would fall on any particular day of the year. If savings were flat across the whole year, the value would be a constant 2.7e-3. MAP-style programs produced the greatest share of savings in late fall and early winter, with comparatively lower savings in the spring. In contrast, the OBF program exhibited its highest savings during the summer.

The non-MAP-style Institutional Partnerships program shows the most irregular savings shape among program types. This irregularity is due to a combination of the underlying load patterns of the fairground participants and the poor performance of counterfactuals constructed to assess performance for some sites. To provide an alternative view that does not rely on these model fits, we examined the savings shapes based on simple pre-post delta values. This alternative representation, which is included in Figure 4-23, shows savings peaking during each site's busy season, with large spikes in July and October at fairground sites and smaller spikes distributed throughout the year.

Figure 4-23. Annualized savings shapes by program type

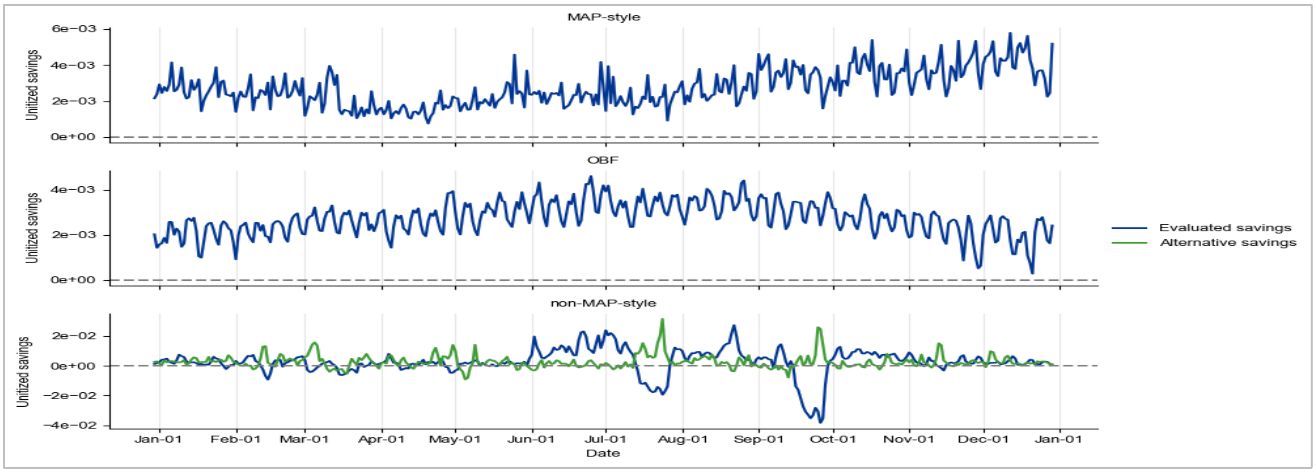
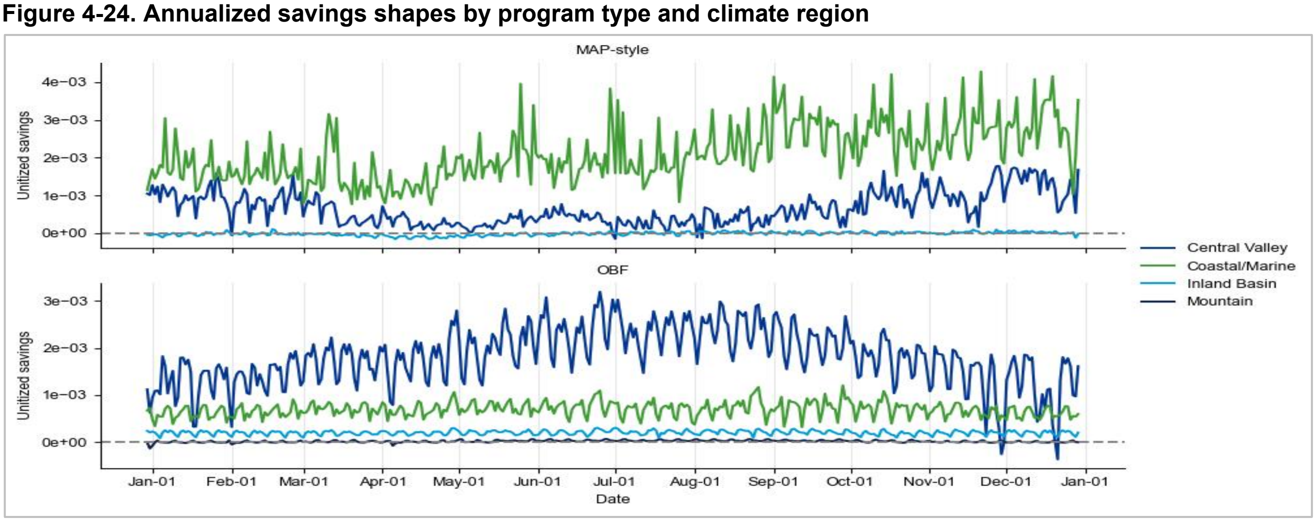


Figure 4-24 disaggregates annualized savings shapes by climate region, revealing that program-type seasonality is strongly influenced by the geographic distribution of projects. For the OBF program in particular, savings are concentrated in the Central Valley, where projects account for a disproportionately large share of total program savings. This regional concentration largely explains the pronounced summer peak observed in the aggregate OBF savings shape.

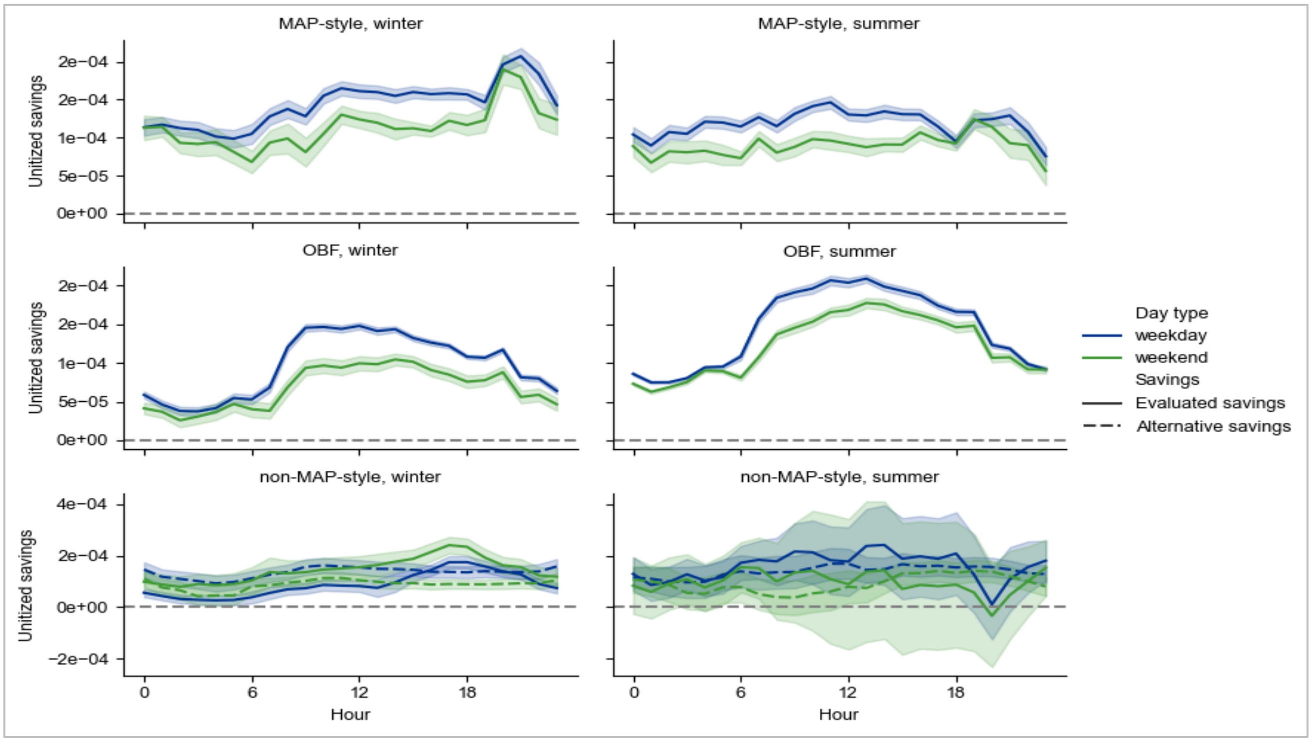
Central Valley projects also have a higher share of lighting measures than other climate regions. While the measure mix alone does not fully explain the pronounced summer concentration in OBF savings, the reduction in radiant heat from lighting upgrades could decrease summer HVAC loads, especially in the hot climate of the Central Valley.

Figure 4-24. Annualized savings shapes by program type and climate region



Timing of savings. We examined hourly savings shapes to identify when verified savings occurred during the day and how their timing varied by program type and season. Figure 4-25 presents average unitized hourly savings shapes, showing the share of annual savings that occurred in each hour of the year. The results indicate that program types exhibit distinct savings profiles. If savings were evenly distributed across all hours, each hour would account for a constant 1.14e-4 of annual savings.

Figure 4-25. Seasonal hourly shapes by program type



The OBF program achieved savings during typical daytime business hours in both seasons. These shapes are consistent with lighting-driven program savings. The summer midday hours achieved substantially higher savings than winter midday hours, possibly reflecting cooling-related savings either due to more efficient lighting or some cooling-related measures. That the savings tail off in the late afternoon supports the hypothesis of lighting-related interactive effects.

MAP-style program savings were surprisingly flat across most hours of both the summer and winter, with slightly greater savings during winter daytime and with unexpected peak savings in the late evening winter hours. Understanding the drivers of these savings will require more granular analysis.

The non-MAP-style Institutional Partnerships program exhibited highly irregular hourly savings patterns during the summer, with wide uncertainty bands that make it difficult to discern consistent hourly trends. In winter, the program exhibited higher early evening savings, likely driven by lighting measures during low daylight hours. Given the modeling challenges with this program, the figure also presents an alternative view of savings to support the interpretation of these patterns.

4.8 Program design and implementation

This section examines how PAs plan, implement, and refine population NMEC programs, focusing on three areas: (1) how PAs understand NMEC program logic, (2) how they use program experience and available evaluation feedback to adjust forecasting and incentive design, and (3) how they apply eligibility rules to align with NMEC rulebook expectations. We draw on PA interviews and program documentation to characterize program theory, operations, eligibility screening, and how PAs handle ineligible or non-enrolled projects. We complement this qualitative evidence with program performance analysis using program-level realization rates (FRR and GRR) and project-level comparisons of initial claimed and verified savings. Where relevant, we discuss OBF separately, given its distinct design and evaluation history, as well as its implications for forecasting, customer routing, and savings reliability.

4.8.1 Understanding of program logic

Program logic is an explicit theory explaining how and why a program is expected to produce energy savings and other benefits. It links program design choices to intended results in a manner that is transparent, testable, and evaluable. In the NMEC context, program logic outlines how a meter-based, measure-agnostic intervention, operating within defined roles, rules, and incentives, delivers measured savings via observable changes in whole-facility energy consumption.

In this evaluation, we focused on whether this logic was consistently understood and operationalized in ways that support the NMEC policy intent. Overall, we found that the PAs consistently articulated NMEC logic in principle, but understanding of its operational implications varied, as reflected in differing expectations about the administrative and analytical effort required to deliver meter-based savings and in implementation outcomes, including measure mix delivered and savings realized. Table 4-23 summarizes the relationship between pop NMEC program design features, their intended benefits, and observed implementation outcomes discussed in the findings below.

Table 4-23. Summary of the PY2022 – PY2024 pop NMEC program goals, designs, and implementation

Design feature	Intended benefit	Observed reality
NMEC policy logic	Clear theory of program benefits	Understood in principle
Administrative design	Simplified participation	Reported as complex and time-consuming
Meter-based savings	Verifiable performance	Implemented inconsistently
Measure-agnostic approach	Flexibility for site-specific solutions	Mostly lighting measures used

PA understanding of NMEC logic aligned with stated policy intent. Program implementation plans consistently cited the NMEC policy articulated by the CPUC, and PAs reinforced this logic in interviews: NMEC programs are intended to provide a flexible, measure-agnostic approach to delivering metered savings. PAs initially viewed these programs, particularly MAP-style programs, as more accessible, leveraging market incentives and aggregator networks to reach customers. Several PAs also highlighted the ability to capture stranded savings, support load shifting, and expand reach to smaller and HTR customers.

More than one PA reported that administrative burden and cost were major challenges for pop NMEC programs. Even with third party implementers handling day to day activities, PAs described long project timelines, complex modeling, and reporting demands as resource intensive. While they noted that project paperwork may be simpler for participating customers than in other energy efficiency programs, they highlighted ongoing reporting challenges for program staff. Staff from two programs reported entering site level claims into CEDARS rather than submitting an aggregate claim, and in one case, staff did not become familiar with the NMEC rulebook until well into program administration. Several PAs requested clearer guidance on reporting expectations and procedures.

PAs with MAP-style programs took a largely hands-off approach to their programs, trusting their third-party implementers with many elements of program design and execution. One PA expressed confidence that market incentives guide aggregators to make the program work effectively and adhere to its underlying logic. Because incentives are tied to metered performance, PAs believe aggregators are strongly motivated to recruit projects that are modelable, eligible, and likely to move successfully through the full NMEC process, including unlocking stranded savings. As a counterpoint, PG&E's OBF program relies on a contractor network that does not receive performance-based incentives, which the PA suggested may contribute to the program's consistently low realization rates. In that program, contractors are primarily focused on financing and installing measures, whereas the PA speculated that performance incentives in other programs may more actively encourage aggregators to vet and recruit projects that are well aligned with the NMEC model.

All PAs reported that installed measures skewed heavily toward lighting, with a few PAs acknowledging that lighting has a limited overlap with system peak hours and therefore may yield lower TSB than measures such as HVAC. Because programs were measure-agnostic and aggregators were driving recruitment and installation, PAs reported limited ability and opportunities to steer programs toward a more diverse range of measures.

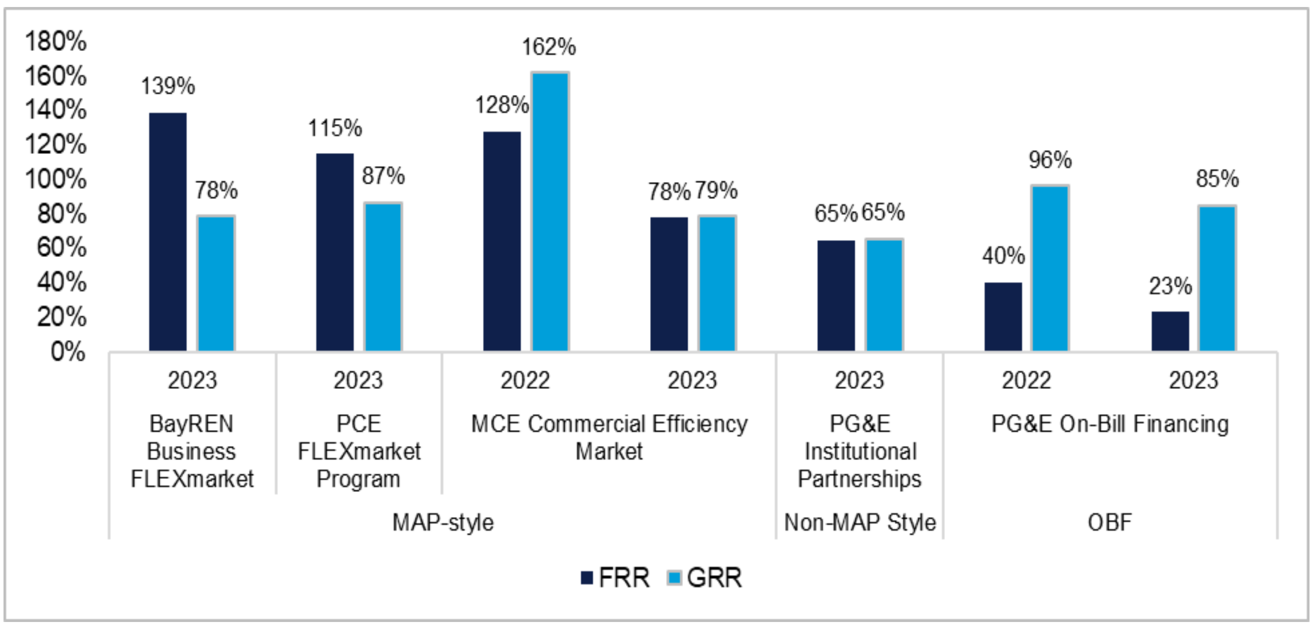
4.8.2 Program feedback and improvement

This section examines how PAs conducting pop NMEC programs are adapting their planning, design, and implementation over time based on experience and available feedback. We review program performance data alongside PA perspectives on how programs have adjusted operations in response to that experience.

Across program types, differences in realized outcomes were driven primarily by planning assumptions rather than by delivery effectiveness. We examined how initial planning assumptions, final claimed savings, and evaluated results align in practice based on realization rates to assess whether programs are delivering on their initial savings promises and the effectiveness of their delivery.

Figure 4-26 shows these realization rates by program year, program, and program type. Specifically, the figure presents FRR and GRR for cohorts with electric (kWh) savings, which are used here to distinguish differences due to planning assumptions (FRR) from those reflecting program delivery effectiveness (GRR).

Figure 4-26. Forecast (FRR) and gross realization rates (GRR) by program type and year



Across program types, initial forecasts imply different levels of savings to customers, with some understating and others overstating what is achievable, at times beyond physically plausible levels. FRR, which compares verified to initial claimed savings, indicates that MAP-style programs tended to plan conservatively, with verified savings generally exceeding initial claimed values. This pattern suggests that MAP-style programs were more likely to under promise relative to what they ultimately delivered. In contrast, OBF and the Institutional Partnerships programs exhibited lower FRRs, indicating more optimistic planning assumptions and initial savings claims that dramatically exceeded subsequently documented and verified savings.

To further examine these differences in planning assumptions, the figures below plot project-level forecasted energy consumption reductions relative to pre-installation consumption (forecast percentage savings) against verified energy consumption reductions relative to pre-installation consumption (verified percentage savings). In both figures, the 45-degree grey dashed line represents the point at which forecasted and verified percentage savings are equal. Projects plotted above the dashed line achieved higher savings than forecasted, while projects plotted below the dashed line achieved lower savings than forecasted.

Figure 4-27 shows this comparison for MAP-style and non-MAP-style programs. Across these projects, verified savings remained below 50%, even where forecasted savings were much higher. Figure 4-28 presents the same comparison for OBF projects. Here too, no projects achieved verified savings greater than 50%, despite several forecasts indicating much higher savings levels. Moreover, for some OBF projects, forecasted savings were not physically plausible as they exceeded total site energy consumption.

Figure 4-27. Project-level forecast savings as a percentage of baseline consumption vs. verified savings as a percentage of consumption, MAP-style and non-MAP style

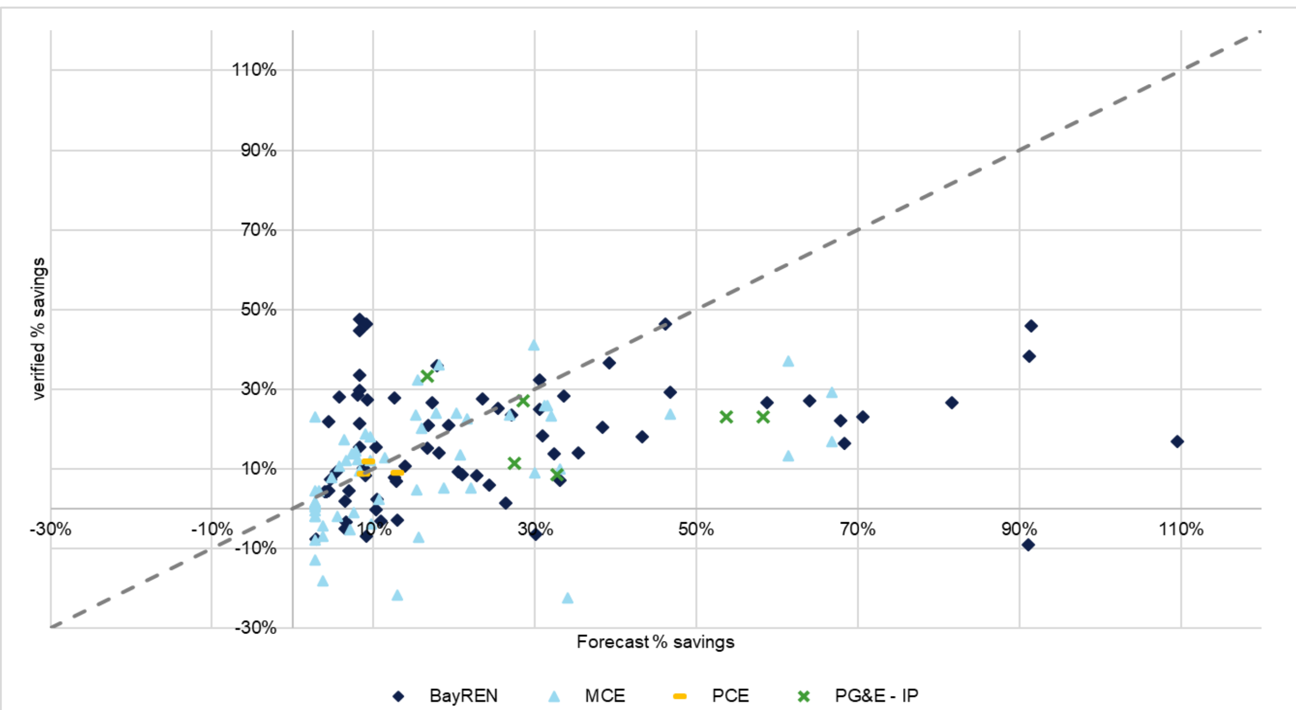
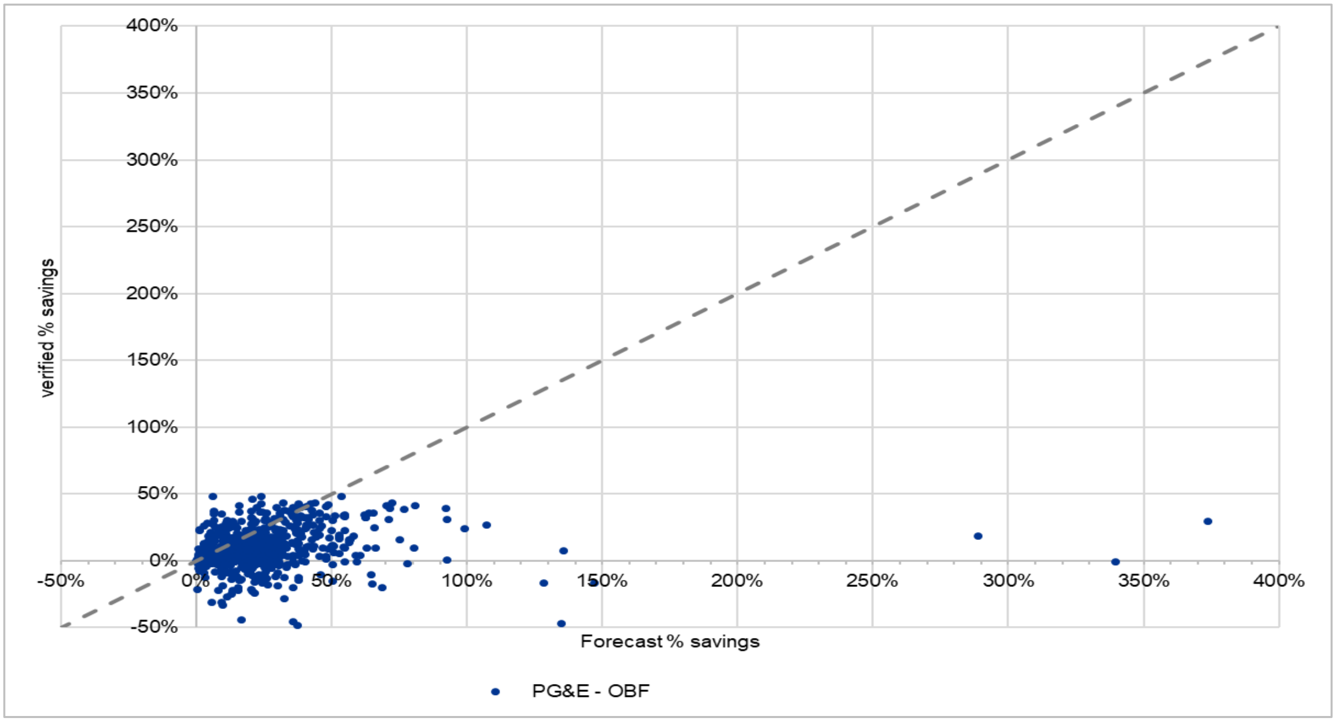


Figure 4-28. Project-level forecast savings as a percentage of baseline consumption vs. verified savings as a percentage of baseline consumption, OBF



These program-level findings suggest that setting a preliminary savings cap at 50% would substantially improve the accuracy of the preliminary savings estimates used to market projects to participating customers. While non-OBF programs showed smaller preliminary savings errors, the customers in these programs were still affected by those errors. For OBF customers, the effect can be larger because preliminary savings forecasts help determine loan repayment terms set prior to installation and reflected in customers' utility bills. When those forecasts materially overstate savings by more than 30%, they can make projected bill reductions appear larger than they are and shorten the implied payback period. Because OBF participation is tied to loan repayment terms set prior to installation and reflected in customers' utility bills, these forecasts help determine both the expected bill offset and the length and structure of the repayment period. As a result, forecast error can lead customers to accept repayment terms that assume faster payback than is ultimately realized, raising questions about how those repayment terms are determined.

PA interviews indicate that the pop NMEC programs actively adapted their offerings to improve program operations and savings. Only a few of the programs have previously received formal evaluations, but all the programs drew on their own experience to adjust and improve. PAs described several adjustments intended to strengthen program operations and address implementation challenges:

- **Enhanced monitoring of project data.** PAs reported that improvements in implementer tools enabled them to more actively monitor project-level data during the performance period. Several PAs described using these tools to track project performance and savings calculations and identify NREs or other anomalies.
- **Adjusted incentive structures.** PAs reported modifying incentive structures over time, both to respond to budget considerations and to target specific customer segments. In other cases, they adjusted incentive structures to ensure that aggregators received a prompt upfront payment (often 20% – 40% of forecasted savings), a critical consideration given how heavily the MAP-style programs relied on market incentives to steer the program.

- **Climbing the reporting learning curve.** Challenges with NMEC reporting requirements also prompted adaptation. Several PAs described improved understanding and handling of reporting over time, including making aggregate claims. One program staff reported learning about the NMEC rulebook through informal discussions with other PAs. All PAs described ongoing communication with peer programs to discuss implementation practices and share lessons learned.
- **Onboarding aggregators.** MAP-style programs relied on incentives to steer aggregators toward accurate forecasts and NMEC-appropriate projects. While some PAs reported early challenges in familiarizing aggregators with NMEC principles, they generally expressed satisfaction with aggregator realization rates over the course of the performance period.
- **Grappling with cohort formation.** PG&E's OBF program staff reported navigating tradeoffs among several competing objectives, including:
  - Building coherent cohorts that feature similar projects and are large enough to gain the certainties that come from pooling results across a large population.
  - Keeping implementation streamlined and providing an accessible energy efficiency option for smaller customers.
  - Encouraging comprehensive measures, reducing peak demand, and shifting load.

  Putting more requirements on participation shrinks populations. Lessening requirements means more lighting projects and a lack of measure diversity. To get to a population size that reduces uncertainty, PG&E pooled all OBF participants into one cohort but has studied other options, for example, considering how more rigorous FSU screening might affect uncertainty.
- **Improving recruitment.** The PG&E OBF program has updated project screening, to better screen out sites that are too large, industrial, or otherwise poor fits for NMEC. (See section 4.8.3.2 for more context.) Even the PAs that largely left recruitment to aggregators mentioned efforts to refine project screening. PCE reported that indoor agricultural facilities were a poor fit for NMEC and would be deprioritized in future cohorts. Similarly, PG&E's Institutional Partnerships program sought to shift recruitment away from facilities with highly variable or seasonal usage, such as fairgrounds, toward regular and comparable usage, such as Department of Motor Vehicles offices, that would make a strong pop NMEC cohort. Whether or not that effort pans out, it reflects an adaptation based on past experience and an attempt to better align with NMEC principles.

Evaluation feedback versus experience-based learning. While PAs described multiple adaptations to program design and implementation, these adjustments were generally informed by operational experience and implementation challenges rather than by formal evaluation findings. Non-residential pop NMEC programs have a limited evaluation history; among the evaluated programs, OBF is the only program with a multi-year record of evaluated results.

Implication for forecasting and incentive feedback. As indicated above, none of the programs have used evaluated savings results to adjust forecasting assumptions or incentive payment practices, largely because they have not yet received formal evaluations. The OBF program is the only pop NMEC program with a formal evaluation, but it does not include performance-based incentives. When asked how the OBF program was adjusted following a prior evaluation that highlighted low realization rates, program staff noted that they had reviewed their engineering calculations but had not "added a multiplier to cut all of our estimates by a set percentage like a safety factor." They also acknowledged that these adjustments had not improved outcomes and that a potential "next step would be to preemptively apply a safety factor." Realization rate patterns and project-level comparisons for OBF show little evidence of improvement in forecasting practices, indicating a continued need for stronger feedback mechanisms linking evaluated results to planning assumptions.

Table 4-24 summarizes OBF's multi-year initial claims, true-ups, verified savings, and realization rates, illustrating the persistence of forecast uncertainty over time.

Table 4-24. OBF multi-year forecast and gross realization rates

PY	Initial	True-up	FRR	Claimed	Verified	GRR
Electric (kWh)						
2020	27,012,587	-24,575,641	9%	2,436,946	2,466,287	101%
2021	33,048,458	-22,842,327	26%	10,206,131	8,596,574	84%
2022	28,141,549	-16,434,890	40%	11,706,659	11,282,202	96%
2023	22,770,868	-16,635,089	23%	6,135,779	5,216,773	85%
Gas (therm)						
2020	5,317	0	973%	5,317	51,741	973%
2021	123,282	0	343%	123,282	423,454	343%
2022	54,792	0	0%	54,792	0	0%
2023	58,851	-49,504	49%	9,347	29,014	310%

4.8.3 Application of eligibility rules for consistency with NMEC rulebook expectations

This section examines how pop NMEC programs applied eligibility rules in practice and whether their eligibility screening and project-handling processes align with NMEC rulebook expectations. The analysis focuses on PA visibility into screening decisions, consistency in eligibility applications, and the management of ineligible or non-enrolled projects.

Table 4-25 summarizes how pop NMEC programs applied NMEC eligibility requirements and is followed by a discussion of these processes and their implications.

Table 4-25. Application of NMEC eligibility requirements for PY2022 – PY2024 pop NMEC programs

Eligibility requirement	Program application
Pre-program data sufficiency	Checked via automated tools
Screening for NREs	Automated tools flag baseline anomalies
Avoiding overlap with other programs	Prior DSM/energy efficiency participation checked (tracking varies)
Handling ineligible customers	No process to redirect
Post-participation model failure	Assigned savings used instead of modeled savings

Eligibility processes were largely handled by implementers or aggregators, with PAs having only limited visibility into those processes. Turnkey operations were part of the appeal of the programs for most PAs, and they had almost no direct contact with customers. PAs were not involved in recruitment and programs did not track aggregators' interactions with potential customers. As a result, they had little involvement in eligibility decisions or the disposition of customers who were not good candidates for pop NMEC programs. The use of metered savings reassured PAs that aggregators were recruiting and enrolling customers in good faith and in line with NMEC principles.

The lack of relationship between PA and customer meant that, for good and bad, most PAs did not take a strategic or deliberate approach to customer participation. PAs could steer customers who are not a good fit into pop NMEC programs to benefit from lower administrative burden or to influence evaluation results, or instead direct them to Custom or site-level NMEC programs or other options that are a better fit. We found no evidence of either practice. Siloed program actors did not have both the motivation and the relationship with the customer to serve as a savings dispatcher, identifying opportunities beyond the specific program in question.

PG&E's NetOne program, which offered several savings pathways, did include this kind of approach in its implementation plan: "Customers who are not good candidates for pop NMEC treatment due to data insufficiency, prior or expected NREs, participation in other energy efficiency programs or other demand-side management offerings such as solar and EV *will be served by the program through Deemed, Custom, or site-based NMEC approaches instead*" (emphasis added). In their responses to us, however, they noted only, "Customers who do not qualify during the pre-installation screening are not enrolled in the population."

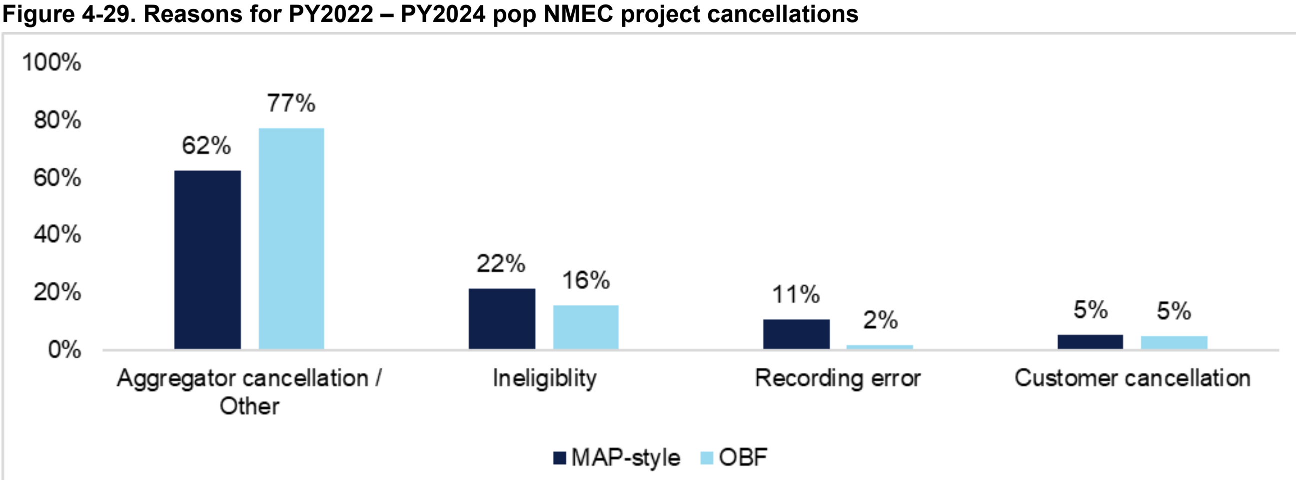
4.8.3.1 Ineligible and cancelled projects

PA pre-participation screening records offered limited insight into cancellation reasons and outcomes. Aggregators for MAP-style and implementers for other programs pre-screened potential projects to assess their fit for pop NMEC before submitting them for official review. No PA had a consistent system to track the disposition of customers whom aggregators or implementers pre-screened, and so there is no estimate of customers who aggregators considered but ultimately pre-screened out.

Once PAs received a project, they ensured that it met eligibility criteria. PAs tracked these projects, documenting the reasons some projects were deemed ineligible. BayREN, MCE, PCE, SDGE, and PG&E for OBF provided pre-participation ineligibility data, though how much detail they provided varied. Some programs had additional layers of review. For example, the San Francisco Environment Department (SFE), which helped administer BayREN's program, performed eligibility review of those projects.

Figure 4-29 summarizes the broad categories for project cancellations by program type based on PA-tracked ineligibility data. The data offered limited insight because the most common reason projects did not advance was cancellation by the aggregator for the MAP-style programs and "other" for OBF. PAs had little insight into the reasons behind those cancelations, but aggregators might, for example, submit a reservation letter to reserve an incentive, only to later conclude the site was ineligible.

Figure 4-29. Reasons for PY2022 – PY2024 pop NMEC project cancellations



PG&E's OBF program also excluded dozens of sites from PY2023 and PY2024 cohorts when the program discovered during true-up review that those sites were industrial or agricultural and should not have been enrolled. The program rejected a handful more projects because their load was too large or event-based schedules made modeling impractical. These post-participation exclusions did not necessarily produce more favorable results, but they should have been pre-participation eligibility exclusions. PG&E had already come to this conclusion on its own: "The commercial site type, maximum site size, and minimum savings depth were existing requirements in our program. These findings prompted us to significantly increase our pre-screening rigor to ensure compliance and prevent similar enrollment errors in future cohorts." PG&E also created a rule to bar from its OBF program sites with event-based schedules that do not model well.

4.8.3.2 OBF program alignment with NMEC rulebook

PG&E's OBF program warrants separate consideration because its design, incentive structure, and eligibility boundaries differ substantially from other pop NMEC programs. Unlike MAP-style programs, OBF relies on financing rather than performance-based incentives and originated prior to the NMEC rulebook. As a result, assessing OBF's alignment with NMEC expectations requires examining not only formal eligibility criteria, but also how program design, participant characteristics, and realized savings interact.

PG&E's OBF program is unique among the pop NMEC programs. It was originally established as a High Opportunity Energy Efficiency Program (HOPP) before the creation of the NMEC rulebook, and while the 2016 advice letter that first proposed the program implicitly or explicitly addresses many rulebook requirements,[45] OBF's novelty means adherence with the NMEC rulebook is not straightforward. In the intervening years, no one seems to have considered the issues that CPUC staff flagged in their commentary to that advice letter.

The zero-interest loan offered by the program is the sole incentive, and while the program uses measured savings to claim final savings, these play no role in determining the terms of the loan or performance-based incentives. The program's implementation plan details the history of this program design and its novelty: "[B]ecause many of OBF-AP's strategies are novel, and because PG&E had not claimed savings under the existing OBF program, there was no precedent for an OBF savings-estimation framework. (PG&E looked for impact evaluations for OBF programs run outside of California and was not able to find any.)"

The CPUC approved the 2016 advice letter that first proposed OBF after extensive discussion and the addition of supplementary materials, addressing a number of challenges presented by the unusual program design. CPUC staff wrote at the time, "[T]he OBF alternative pathway will test whether using a forecasted change in normalized consumption as a basis for savings, in exchange for losing the opportunity to receive a rebate will substantially broaden participation. *Commission Staff accepts the proposed alternative approach as an opportunity to test this idea, and accepts that the proposed program may not reach significant stranded potential*" (emphasis added). Likewise, Commission staff acknowledged in its commentary that upfront incentives may overrun the value of the final savings but determined that it was "willing to accept this proposal with the caveat that *data and information collected during implementation will be used to inform a follow up assessment before the program is offered at full scale*" (emphasis added).

While Cadmus and DNV performed evaluations of the OBF program in 2020 and 2023, respectively, neither evaluation addressed the issues flagged in the 2016 advice letter. To our knowledge, there has been no attempt to assess in an official evaluation whether the trade-off of forecasted savings, which in practice are based on engineering estimates and not measured savings, for rebates broadened participation.

The program's low realization rates means that loan repayment terms are overrunning the value of final savings, as theorized in 2016. While the program's loans may still be paid back and there may be no harm to the program or ratepayers, customers whose loan terms were calculated to be bill neutral may face higher than expected utility bills after participation. The OBF loan agreement states, *"Customer shall repay the Loan Balance to PG&E* as provided in this Loan Agreement irrespective of whether or when the Work is completed, or whether the Work is in any way defective or deficient, and *whether or not the Work delivers the energy management benefits to the Customer*" (emphasis added).

[45] Advice Letter, PG&E 3697-G /4812-E, 3697-G-A/4812-E-A ELEC_4812-E-A.pdf

However, responses to our survey suggest that for many pop NMEC OBF customers, access to financing was the chief motivation behind participation, and they may be willing to accept the trade-off of higher utility bills. This finding provides important context for interpreting program performance, as it highlights the extent to which OBF serves customers whose primary barrier is capital availability. The section below provides an analysis of participant funding considerations, participation drivers, and firmographic details.

The OBF program reaches segments that are structurally limited in their access to traditional financing. Table 4-26 illustrates the distribution of facility types among survey respondents. A significant share come from sectors that tend to have tight budgets, long capital cycles, or limited options for traditional loans. Many of the survey participants (62%) are municipal entities such as K-12 schools that otherwise would have to issue bonds, likely requiring a ballot measure, to access financing.[46] As a result, OBF fills a critical financing gap by offering an accessible, low-barrier alternative that enables these entities to pursue energy efficiency upgrades that would otherwise be delayed or abandoned.

Table 4-26. Main activity at the facility

Facility activity (n=53)	Unweighted %	Weighted %
*Education (primarily K-12 schools)	37%	22%
*Municipal & civic facilities	17%	19%
*Utilities & infrastructure (primarily water treatment plants)	15%	21%
Food & agricultural processing	13%	14%
Office	9%	13%
Commercial	4%	4%
Mixed use	2%	3%
Housing	2%	3%
Medical	2%	1%

*** Indicates sector with limited access to traditional financing**

Survey respondents also reported their motivations for undertaking projects, which center on sustainability, productivity, and organizational and policy compliance, as shown in Table 4-27. While energy reduction was the most frequently cited motivation cited, other drivers such as improving productivity and policy compliance also played a role, reflecting the multifaceted value proposition of energy efficiency investments for OBF participants.

Table 4-27. Participant motivations for completing projects

Motivations for undertaking projects (n=79)	Unweighted %	Weighted %
Energy reduction	48%	60%
Improve productivity	22%	13%
Compliance	9%	11%
Sustainability	12%	7%
Equipment failure	4%	3%
Financial reasons	3%	3%
Other	3%	3%

Three-quarters of OBF program participants reported that they would not have been able to complete their projects at all or as-is without the program. Table 4-28 shows that about a third of projects (35%) would not have happened at all.

[46] SoCalREN and TRC. “Public Sector Market Characterization Financing Study -Final Report.” 2021. https://pda.energydataweb.com/api/view/2468/SoCalREN%20Public%20Sector%20Market%20Characterization%20Financing%20Study%20Report_FINAL.pdf

Additionally, among participants who might have proceeded with their project without on-bill financing, most indicated they would have faced substantial delays (35 – 43%) or reduced scope (5%). In other words, OBF did not merely influence project financing; it significantly enabled and accelerated projects that would not otherwise have occurred. In general, participants emphasized that limited budgets, competing priorities, and the absence of alternative funding made upgrades difficult, yet the opportunity to reduce operating costs, address maintenance needs, and achieve a relatively quick return on investment made the project both feasible and valuable. Some illustrative participant comments include:

- "There was money available for us to make the complete upgrade; we are renters and our company is paycheck to paycheck."
- "We are a school. There isn't much disposable income and there are competing priorities"
- "There was no other alternative funding to get this job done and our goal was to lower our operations cost and maintenance."

Table 4-28. How the installed project would have changed without OBF program

Project changes without OBF (n=67)	Unweighted %	Weighted %
Would have taken longer / been done in phases	43%	35%
Would not have happened at all	35%	35%
Nothing would have changed	19%	26%
Reduced scope or lower efficiency	3%	5%

This is consistent with other survey findings: Table 4-29 demonstrates that nearly 71% of participants reported that they used no other funding source alongside OBF.

Table 4-29. Project funding sources used other than OBF

Funding sources used (n=78)	Unweighted %	Weighted %
No other source	80%	71%
Internal funds	13%	17%
Don't know	5%	8%
Financing (loan) other than the PG&E OBF	3%	4%

As shown in Table 4-30, OBF also covered the vast majority (98%) of project costs, while internal funding accounted for only 2% of the financial resources. Further, 98% said they did not seek other loan financing options at all and 68% reported that they would have had to seek out alternative financing without the OBF funding. Participants also expressed low perceived likelihood of qualifying for loans, with only 23% confident they could have secured other funding resources, reinforcing that OBF served customers who would otherwise not pursue or obtain debt financing.

Table 4-30 Reported project costs by sources used

Percent of project cost by source (n=78)	Unweighted %	Weighted %
PG&E OBF	98%	98%
Internal funds	2%	2%

OBF was, by far, the biggest pop NMEC program during the evaluation period. As discussed, its presence in the pop NMEC universe was initially debated and when it was allowed to go forward, the original intent was to assess whether it did, in fact, belong in pop NMEC. The program's success attracting customers and maintaining a degree of satisfaction while producing performance-based savings provides an encouraging piece of feedback. On the other hand, the challenges addressed in this section and throughout this report make clear that OBF can be a better pop NMEC program than it currently is.

- The dramatic overstatement of preliminary savings reflects a need for improvement in program savings forecasts. The poor FRRs have been present throughout the history of the program. They were identified in the prior pop NMEC report. The lack of progress in addressing this issue suggests the need for increased attention to how these forecasts affect customers and overall program performance.
- PG&E manages the OBF program and access to its financing, which gives it a strong position to address ongoing program issues. However, there is limited evidence that this position has been used to address persistent challenges. PG&E suggested that the lack of direct incentives to contractors may contribute to issues with FRRs and low savings as a share of consumption, but this potential driver has not yet been addressed. Several approaches could help, including tracking contractor-level FRRs and increasing scrutiny of contractors that consistently overestimate savings, capping preliminary savings at 50% to improve forecast accuracy, and refining eligibility criteria for participating customers. PG&E could also revisit earlier CPUC staff concerns about the role of financing in reaching stranded savings by strengthening reporting expectations for contractors, including information on participant barriers, self-reported acceleration, and system repair eligibility. Periodic audits or evaluator checks could reinforce these expectations and help ensure continued alignment between contractor practices and program goals.
- Now that OBF has been established for so many years, it is reasonable to reconsider CPUC staff's original concerns about the program. They were unsure that the program could successfully attract participants with only financing and no rebates. That question has clearly been answered in the affirmative. The obvious extension of that question is to go beyond loan terms that are based on demonstrably inflated preliminary savings. OBF could develop a two-stage process, updating the initial financing agreement with performance-based savings. Customers would then pay for their projects at a rate consistent with their meter-based savings. This change would force all parties to operate in the real world of meter-based savings, ensuring risk no longer lands solely on the customer. PG&E would have much greater motivation to manage the program carefully because poor performing projects would tie up funds for longer periods. Contractors would be motivated to identify good participants and serve them well because their activities will be tracked with respect to meter-based performance. Ultimately, the only way to fully justify OBF's presence in pop NMEC is to make it a performance-based program. The details of this transition will be complicated, but this improvement, if pursued thoughtfully and incrementally, should improve the program's service to both customer and ratepayer without meaningfully undermining the current program's stronger points.

4.9 Methodology assessment

This section provides a high-level assessment of the methods currently used to estimate performance of pop NMEC programs, with a focus on how well those methods support NMEC objectives. The discussion centers on site-level modeling and impact estimation practices derived from traditional billing analysis and examines how they have been adapted for pop NMEC applications.

The section is organized in two parts. Section 4.9.1 reviews site-level modeling as the foundation for pop NMEC, emphasizing its dual role in weather normalization and eligibility screening, and identifies synthetic control models as the primary departure from traditional specifications. Section 4.9.2 then describes and evaluates current pop NMEC impact estimation methods within the broader difference-in-difference (DID) framework used in billing analysis, focusing on how programs incorporate comparison information, address NREs, and quantify uncertainty.

4.9.1 Site-level modeling for pop NMEC

To date, all M&V providers implementing pop NMEC programs have relied on some form of site-level modeling as the foundation for estimating program impacts. Under site-level modeling, impacts are first estimated at the individual site level and then aggregated to the population, making the DID structure more explicit and transparent than it is in panel methods, under which participant and comparison-group customers are pooled into a single regression model.[47]

Traditionally, site-level modeling has served a single primary purpose: weather normalization. Relatively simple models explicitly separate weather-correlated from non-weather-correlated load, typically isolating heating and cooling effects from baseload. These model structures allow observed consumption in both the baseline and performance periods to be adjusted to a common weather basis, such as TMY temperatures. This adjustment isolates changes in energy consumption due to program activity from differences in realized weather, while reflecting typical conditions expected over the useful life of installed measures. An alternative approach more common in real-time meter-based assessment, avoided energy use (AEU), projects baseline consumption into the performance period using a pre-period model. While this AEU framework eliminates the need for a post-period model, it forces savings estimates to reflect actual performance-period weather, whether or not those conditions are typical, unless normalized in a separate procedure.

These traditional site-level models have a well-established engineering foundation rooted in PRISM analysis, originally developed in the 1980s, to establish the relationship between load and temperature.[48] Traditional site-level models are intentionally static. That is, their structure and parameters are assumed to be stable over time. The intent of these static models, which estimate heating and cooling slopes around model-optimized changepoint temperatures, is not to explain all observed consumption variability but to capture the primary drivers of energy consumption using defensible, engineering-based assumptions.

This design is conveniently consistent with NMEC, which assumes that absent program activity, underlying site consumption patterns remain stable across the baseline and performance periods. Accordingly, the role of site-level modeling in NMEC is not to detect underlying trends or structural changes in consumption, but to normalize for weather and thereby isolate program-related effects.

Thus, in the pop NMEC context, only independent variables that are expected to repeat consistently in future periods, such as day-of-week indicators, can be appropriately incorporated in site-level models. More flexible specifications, including monthly dummy variables or approaches that estimate separate models by month (such as the version of the TOWT model discussed in Section 4.2.6), can reduce regression error. However, they are not well aligned with NMEC objectives because they do not reflect static consumption across the year and can capture within-year trends or NREs that may be carried into counterfactual estimates.

NMEC's emphasis on interpretability means model selection should not hinge solely on the basis of fit statistics from any model, such as CVRMSE, but should favor models with stable structures that do not simply absorb irregular or unstable load but reflect it as increased impact uncertainty. This distinction is particularly important for pop NMEC eligibility screening, since sites that do not exhibit sufficiently regular and repeatable load patterns are unsuitable for inclusion, even if flexible models yield strong fit statistics.

Site-level modeling, therefore, serves a dual role in pop NMEC. In addition to supporting weather normalization, it supports eligibility determination through estimates of FSU, which reflects the precision of savings and represents an effective eligibility metric for pop NMEC programs. Section 4.2.6 illustrates the challenges of getting models to reveal instability rather than concealing it through excessive temporal flexibility.

---

[47] In either case, the NMEC rulebook states that "the specific measurement method(s) and calculation software must be determined before the program begins and applied uniformly to all sites in the program" (III.3.A.6, p. 18).

[48] Fels, M. F., and M. L. Goldberg. 1986. "Using the Scorekeeping Approach to Monitor Aggregate Energy Conservation." Energy and Buildings 9 (1–2): 161–68. doi:10.1016/0378-7788(86)90017-4

4.9.1.1 Synthetic control models

Section 4.2.6 discusses the two modeling approaches observed in this year's evaluation. These include traditional TOWT regression models and the synthetic control (SC) approach that incorporates granular load profiles as an independent variable as well as weather splines and time indicators.

The SC approach deserves additional attention because it represents the greatest departure from traditional site-level modeling approaches, primarily through its inclusion of granular profiles as explanatory variables. The granular profile is a mean hourly consumption series constructed from a sample of non-participant customers that share the modeled site's characteristics, including industry (NAICS), climate zone, and size. Site-level models typically include weather as the primary continuous independent variable. The SC model introduces the granular profile as an additional continuous predictor with comparable statistical properties.

In simple model-fit terms, adding a granular profile improves explanatory power by introducing a continuous variable that reflects hourly energy consumption. With this addition, the model is no longer fixed across the year; instead, it varies with the temporal patterns embedded in the granular profile while remaining stable with respect to the broader trends or aggregate NREs that the granular profile captures. Beyond improving the site-level model fit, the granular profile also plays a key role in the SC method by helping control for non-participant trends.

As a stand-alone site-level model, this SC approach does not obscure underlying consumption patterns in the same way that models permitting highly flexible temporal effects do. However, it gives a different FSU estimate than a standard static site-level model with weather alone. As a result, at a given FSU benchmark used to determine pop NMEC eligibility, some sites may incorrectly qualify under this modeling approach but not under the traditional modeling approach, or vice versa.

4.9.1.2 Modeling findings

As noted above, current pop NMEC programs employ different site-level modeling approaches. While each approach may be appropriate under certain conditions, they differ substantially in how well they support the pop NMEC process. In particular, the use of model-based results as a criterion for program eligibility is only defensible when eligibility thresholds have a consistent statistical interpretation across modeling approaches.

Some of the models currently in use mask NREs and irregular consumption patterns and systematically understate savings variability estimates. As a result, these models allow sites with unsuitable consumption characteristics to enter the program and contribute to greater variability in realized savings estimates.

Moreover, while the approaches are unlikely to materially affect annual savings magnitudes, they can distort estimated 8760 impact shapes. Pop NMEC programs would therefore be better served by limiting eligible modeling approaches to those that reveal NRE-related variance in FSUs, rather than attenuating or obscuring it.

4.9.2 Pop NMEC impact methods

This section describes the methods used by the different programs to develop impact estimates. It begins with background on traditional approaches, providing the foundation from which these newer methods have emerged. This context establishes a basis for critically evaluating the newer approaches, some of which offer promising alternatives, while others may represent a step backward relative to established practice.

4.9.2.1 Basic billing analysis DID

Pop NMEC methods that aggregate site-level impact estimates to produce program-level savings build on decades of work in billing analysis. Traditional billing analysis relies on a DID approach that makes comparisons within and across subjects to provide the most accurate estimates of impacts. In particular, it compares pre-post changes in participants' energy use to the same pre-post changes for sites in the comparison group.

In billing analysis, each customer provides its own within-subject baseline for estimating program impact at the site level. The site-level impact estimate captures all observed change between the baseline and performance periods: program-related savings *and* non-program-related change. That non-program-related change represents the savings estimation error. A comparison-group customer, selected for similarity to a participating customer, undergoes an identical pre-post comparison, producing a parallel estimate of change that is intended to reflect, and thereby isolate, non-program-related changes in consumption embedded in the participating customer's estimate.

Program-level savings are then estimated by taking the difference between the mean pre-post change observed among program participants and the mean pre-post change observed among comparison-group customers. This DID estimator adjusts participant savings for non-program-related change that would otherwise bias the impact estimate.

The DID approach also provides a straightforward way to estimate the uncertainty of the impact estimate. Each group's mean pre-post change has an associated standard error that reflects variation across customers. When aggregating site-level results, both the participating customers and the comparison groups are treated as samples drawn from the full populations of interest. This treatment is required because the comparison group, regardless of how it is constructed, represents only one possible selection within the broader population from which an appropriate comparison group could be drawn.

Since the final impact estimate is the difference between the mean estimates of the two groups, the standard error for this estimate is based on the combination of the standard errors of the two groups, produced by quadrature (the square root of the sum of the squared standard errors). This combined standard error includes both across- and within-subject variation[49] and is roughly 1.4 times the size of each group's standard error.[50] The associated reduction in the impact estimate's precision is essentially the cost of the bias adjustment offered by the comparison group.

This approach to estimating billing analysis impact precision underlies the NMEC rulebook's restatement of the long-standing recommendation to evaluate relatively homogeneous populations. As with larger sample sizes, greater similarity among participants reduces variability in observed pre-post changes and improves impact estimate precision, all else being equal.

4.9.2.2 Variance calculations and NREs

Standard error calculations are critical in pop NMEC because they indicate whether an impact estimate has adequately accounted for NREs. NREs pose a fundamental challenge for NMEC and other meter-based savings methods because they are embedded in site-level pre-post consumption changes. As a result, the observed difference between baseline and performance period consumption reflects both true program-related savings and unrelated changes in consumption and is not, by itself, a reliable estimate of savings.

Some non-program-related changes are routine, such as those caused by weather, and can typically be addressed through regression modeling. In contrast, many non-program-related changes in consumption are idiosyncratic and cannot be readily explained or forecast as part of a stable, ongoing process.

The site-level NMEC pathway often conceives of NREs as visually apparent, discrete events, such as equipment failures or operational changes that produce sharp shifts in consumption. In pop NMEC, however, NREs are addressed in aggregate by the comparison group. As a result, they encompass all unexplained changes in consumption, whether they appear as

[49] This is in contrast to a quadrature approach where individual site standard errors are squared and summed to estimate the combined variance. If the site-level model error was the only source of error, then this aggregation would be suitable. But the non-program-related changes that are included in participant impacts and are the sole source of comparison group impacts are clearly not zero and must be accounted for.

[50] Assuming the groups' standard errors are equal, which they are unlikely to be, the combination adds a factor of square root of two. As the comparison group's standard error would be expected to be smaller, as it lacks any additional variability caused by program savings, the combined standard error will be less than 1.4 times the participant group standard error but there will still be an adder.

abrupt deviations at individual sites or as more subtle trends. Across many comparison group or granular profile customers, NREs appear as trend over time in aggregated data.

Addressing NREs is essential because their magnitude can be large relative to the savings programs aim to measure. At the site level, dramatic NREs can temporarily alter consumption by 50% or more. More commonly, for groups of customers, consumption may trend upward or downward from year to year, with fluctuations on the order of 5% not uncommon, especially with smaller sized groups. These levels of variation are particularly important in the context of savings estimates that are often expected to fall within a range of roughly 5% to 25%.

Aggregate NREs therefore have the potential to significantly increase or decrease savings estimates derived from meter-based methods. While site-level approaches can sometimes address NREs directly through individual data review and targeted modeling adjustments, pop NMEC approaches do not permit that degree of site-specific intervention. Instead, pop NMEC relies on comparison groups to mitigate the bias introduced by NREs, and it relies on large sample sizes to manage the additional variance that results from that bias adjustment.[51] Traditional billing analysis provides a well-established framework for understanding how these tradeoffs are handled, and it offers important context for evaluating how newer pop NMEC methods attempt to achieve the same objective.

4.9.2.3 Comparison of current methods

The methods used in the current pop NMEC programs differ from traditional methods across multiple dimensions. Although all the approaches incorporate some form of a comparison group, these groups are defined and incorporated in different ways. Table 4-31 maps the program variations.

Table 4-31. Program method variation

PA	Program	Comparator	RHS regressors	Model	Savings	Weather
MAP-style						
BayREN	**Business FLEXmarket**	Granular profile	Time, Temp	TOWT	AEU	Actuals
PCE	**FLEXmarket Program**	Matched	Time, Temp	TOWT	AEU	Actuals
MCE	**Commercial Efficiency Market**	Matched	Time, Temp	TOWT	AEU	Actuals
Non-MAP Style						
PG&E	**Institutional Partnerships**	Granular profile	Time, Temp	TOWT	AEU	Actuals
OBF						
PG&E	**On-Bill Financing (OBF)**	Granular profile	Time, Temp, Granular profile	SC	AEU	TMY

- Two programs used matched customers to construct comparison groups; the remaining programs relied on granular profiles. Granular profiles represent the hourly mean load shape of populations within segments defined by industry (NAICS), climate zone, and customer size. They reflect the average consumption pattern of non-participants within each segment. Granular profiles eliminate the need to supply M&V providers with individual non-participating customer data.
- The programs used two different site-level modeling approaches. Under traditional billing analysis, the aggregate DID estimate is largely insensitive to the specific site-level model used. This does not hold for the OBF program, however, where the synthetic control method incorporates the comparison logic directly into the site-level model. Each participant's baseline-period model, which includes the granular profile along with weather and time variables, is

[51] Larger populations or sample sizes included in a pop NMEC analysis generally improve precision by reducing the influence of site-level variability, model error, and comparison-group adjustment uncertainty on aggregate savings estimates; however, population size is one of several factors that affect whether results meet a specific precision threshold.

projected forward into the performance period, and that projection then serves as the source of bias adjustment typically provided by the comparison group.

- All programs used an AEU approach to estimate site-level impacts. This contrasts with the standard approach, which compares baseline and performance period consumption estimates normalized to TMY weather. In fact, only one program, OBF, puts the final impact estimates on a TMY basis, and that is achieved with a separate regression where savings are modeled as a function of weather. Because AEU relies on one specific year's weather, which may vary widely from a typical year, final results produced on an AEU basis are not consistent with lifecycle savings and TSB, which are calculated over the EUL of the program.
- All programs applied comparison-group adjustments at the site level rather than aggregating comparison groups separately and adjusting results afterward. As a result, each participating site is adjusted individually, even though the adjustment is only meaningful in the aggregate. This structure allows the use of multiplicative, rather than the additive adjustments of the type implicit in traditional DID methods. For programs that do not use the synthetic control modeling approach, the granular profile adjustment is implemented by calculating performance-period ratios of actual to counterfactual granular profile hourly values and applying the ratios to the participant's counterfactual hourly values, which are the baseline model projected under performance-period weather. The synthetic control approach achieves the same adjustment through a single site-level model that incorporates the granular profile directly into the performance-period baseline.

These permutations introduce a complex set of interacting variations that will be impossible to fully explore without extensive test datasets across multiple temperature scenarios and customer edge cases. Even then, being able to make apples-to-apples comparisons will not always be possible.

Even so, following the structure of the earlier discussion, there are three stages at which to compare the methods on a more qualitative basis: site-level impacts, the incorporation of the comparison group, and the appropriate consideration of errors in the aggregation. The methods do not all cleanly fit these three stages, but it is still possible to make some comparisons.

Site-level impact estimates

Differences in participant site-level impact estimates across methods primarily reflect how each approach handles weather correlation. Because the weather specifications used in the models differ notably, the implications for estimated impacts could only be evaluated across a range of weather conditions and participant characteristics. While weather-based consumption models tend to be stable, differences in impacts can emerge when weather conditions differ markedly between the baseline and performance periods.

Two modeling choices illustrate how differences in weather specification can affect site-level impact estimates. In TOWT, the flexibility that allows the model to accommodate within-year trends may also influence how the relationship between consumption and weather is estimated. In the synthetic control approach, the site-level model incorporates the granular profile, which captures weather-driven consumption variation common to both the participant and the granular profile, while only the remaining participant-specific weather sensitivity is captured by the explicit weather terms.

Overall, differences in weather specification across modeling approaches are unlikely to produce large variation in annual impact estimates. However, the implications of these methods for hourly impacts over the year's 8,760 hours remain relatively unexplored. As a result, the effects on peak-period load reductions and TSB are difficult to predict. While some variation across methods is reasonable to expect, there is a lower likelihood of a systematic directional bias associated with any single approach.

DID aggregation

The combination of participating customer and comparison customer information, implemented in a DID-like framework, introduces additional variability that has not been well explored. A key novelty is that some of the programs apply comparison adjustments at the site level using a multiplicative, ratio-based approach combined with an AEU framework that relies on performance-period actual consumption.

While AEU results reflect performance-period weather, weather can be subsequently normalized to TMY, as demonstrated by the OBF program. Of more significant concern, AEU-based methods carry full performance-period hourly variability directly into the 8760 impact shapes. This preserves realistic variation but also embeds performance-period NREs, whether discrete deviations or trend-related shifts, into the final impacts. In contrast, traditional weather-normalized DID approaches tend to assign NREs to model error, limiting their influence on impact shapes.

The multiplicative ratio approach further amplifies the effect of performance period NREs. Additive adjustments are impractical because granular profile values can differ substantially in magnitude from an individual participating customer's actual consumption. While ratio-based scaling solves that problem, it can magnify variability, particularly when ratio denominators are small or when NREs differ between baseline and performance periods, hardly uncommon given that NREs are, by definition, *non-routine*.[52] The synthetic control approach avoids an explicit ratio adjustment by embedding the granular profile directly in the baseline model, but the resulting impacts still reflect performance-period hourly variability and NREs.

Across all approaches, AEU-based performance periods can lead to disjointed annual savings shapes that are inconsistent with the steady-state assumptions underlying NMEC savings and TSB.

Accumulation of errors

In addition to the differences in their savings estimates, current methods handle uncertainty in potentially unreliable ways. Under a standard billing analysis DID framework, the comparison-group adjustment is applied at the aggregate level, and the variance associated with that adjustment is explicitly carried as the cost of reducing bias. This makes transparent the relationship between bias reduction and increased uncertainty.

Approaches that rely on granular profiles as a comparator complicate this accounting. Because granular profiles are constructed by averaging the hourly consumption of many non-participants, they smooth over variation across individual customers and therefore exhibit less variance than the underlying population. Matched-customer approaches, such as MCE's use of four close neighbors, already improve precision relative to a single comparator, but granular profiles further reduce estimated standard errors by averaging over many additional, less relevant customers. Additionally, when granular profile-based results are aggregated, the same granular profile information is effectively counted multiple times, creating redundancy that can bias standard error estimates downward. These same concerns are present in the SC approach but through the novel inclusion of the granular profile as an independent variable.

The incorporation of the comparison group adjustment at the site-level prior to aggregation further complicates the comparison of variance estimates. Rather than treating the comparison adjustment as a separate component whose variance can be tracked and added explicitly, the adjustment is embedded within each site-level estimate and then aggregated. Once this is done, it is not possible to quantify the uncertainty associated with the comparison group. While some program implementations appear to attempt to counteract this artificial variance reduction, there is insufficient evidence to conclude that the resulting standard error estimates reliably reflect true uncertainty. Given the likelihood of meaningful variability across impact measures, this raises concerns about over-confidence in reported precision.

[52] It's unclear how these adjustment deal with negatives in the denominator. This raises particular concerns regarding population groups with solar generation.

4.9.2.4 Conclusions regarding methodological variation

Pop NMEC relies on complex billing analysis methods to estimate meter-based savings, and these methods must meet an unusually high standard of transparency to allow independent validation. Unlike traditional evaluation settings, all modeling steps, assumptions, and implicit decisions must be sufficiently documented so that results can be replicated and validated by a third party. This requirement substantially raises the technical bar for both method design and implementation and places greater weight on clarity, interpretability, and reproducibility of the relatively new methods offered.

At the same time, NMEC guidance on acceptable methods is broad. In the absence of prescriptive standards, programs have adopted a range of modeling approaches. While innovative, the statistical properties and suitability of many of these methods for NMEC purposes remain insufficiently understood. Several of these methods were developed in response to the practical challenges of having M&V providers, rather than evaluators, conduct the analysis. Granular profiles, for example, are intended to reduce reliance on individual non-participant consumption data when constructing comparison groups, but their use has also driven additional methodological variation. Multiplicative adjustments and approaches such as synthetic control models that directly embed granular profiles within site-level models add methodological variation.

As a result, current pop NMEC methods have produced savings and uncertainty estimates that vary substantially across programs and approaches, both in magnitude and precision. These differences are not always transparent and have not been systematically evaluated across a full range of conditions, including different business types, weather patterns, and NREs. This level of variability does not serve most stakeholders well, particularly customers and ratepayers, and complicates interpretation of results. Addressing this issue will require greater consistency, either through convergence in program methods, the establishment of formal validation criteria, or the application of a more standardized and consistent evaluation approach. Without improved alignment, the credibility and long-term viability of pop NMEC will remain at risk.

5 CONCLUSIONS AND RECOMMENDATIONS

Pop NMEC programs aim to deliver credible, performance-based savings while expanding access to energy efficiency to customers for whom traditional program approaches may be challenging or overly cumbersome. While pop NMEC can clearly motivate meaningful savings, two primary areas need substantial attention: 1) strengthening savings measurability and 2) achieving stranded savings that are consistent with core NMEC objectives. Savings measurability can be improved through a combination of stronger adherence to program rules regarding the appropriate type of load for pop NMEC meter-based methods and using technical methods that provide transparent and consistent results. Stranded savings, the up-to-code savings for which NMEC methods are designed, can be realized by focusing on customers with demonstrated technical or financial barriers and supporting comprehensive upgrades that would otherwise be deferred or forgone.

5.1 Savings claims accuracy recommendations

Ensuring savings measurability is a top priority as pop NMEC programs mature. To meet savings measurability and accuracy expectations, pop NMEC programs need to recruit and serve customers whose meter data will support standardized population, meter-based approaches. Additionally, pop NMEC methods need to be streamlined so that they provide consistent final estimates of savings across PAs and implementers.

5.1.1 Program design improvements

Savings accuracy in pop NMEC programs depends substantially on participant eligibility criteria and screening, as well as program documentation management.

As many as 15% of participants in PY2022 – PY2023 came from sectors that cannot be modeled accurately under pop NMEC. The NMEC rulebook limits pop NMEC participation to commercial or commercial-like sites and excludes sites with industrial or process-driven loads. These limits are consistent with site-level NMEC programs and support standardized, pop NMEC modeling. Nevertheless, approximately 15% of participants in the PY2022 – PY2023 programs came from sectors that commonly exhibit process- or event-driven energy consumption (manufacturing, agriculture, infrastructure and utilities, etc.), which cannot be modeled accurately under pop NMEC. The participation of these ill-suited sites affects savings accuracy because their energy consumption patterns are often so variable they cannot provide a reliable baseline from which to estimate savings.

Program savings accuracy can be improved with more homogenous populations. The NMEC rulebook stipulates pop NMEC cohorts should include sites with similar building types, equipment, drivers of energy consumption, and levels of energy use. BayREN's program was largely composed of retail sites but also included one large manufacturing site, which was enough to dramatically increase savings uncertainty. PG&E's OBF program reported difficulty forming cohorts that were both sufficiently similar in project characteristics and large enough to support pooled population level analysis but was the only program to meet 90/25 precision for its impact estimates. By addressing highly variable levels of energy consumption among their populations, all programs can better align with the NMEC rulebook and improve their savings precision, both of which are essential program goals.

Insufficient pre-participation screening allowed participation of unsuitable sites, leading to non-meter-based savings assignments, affecting more than 60% of gas OBF sites. Pop NMEC programs require effective eligibility screening to ensure that participating sites have sufficient and appropriate data to support meter-based measurement. These eligibility requirements are detailed in program M&V plans. When PAs do not identify or screen out ineligible sites during enrollment, programs cannot develop robust, transparent, and defensible final savings claims. The OBF pre-screening, for example, did not exclude all sites with inadequate pre-installation data. As a result, approximately 7% of electric sites and more than 60% of gas OBF sites ultimately received non-meter-based savings estimates. These kinds of lapses in eligibility screening directly affect savings accuracy, compromising the analysis population and undermining the use of performance-based estimates to support program-level claims.

Nearly all programs faced challenges maintaining clear, consistent, and transparent documentation from enrollment through final M&V. Because savings claims are made at the pop NMEC program-cohort level and not at the individual project level, the PAs and program implementers alone are currently responsible for gathering, organizing, and documenting data to support the pop NMEC savings claims. However, the PAs and implementers did not maintain clear records connecting individual projects to the CEDARS claims that PAs make at the program-cohort level. Only BayREN could produce documentation that reconciled with initial CEDARS claims without requiring substantial investigation from the PA and the evaluation. One PA deleted the documentation that could have supported its performance-based results, and we could not verify the savings claimed. Another PA provided two inconsistent datasets for the same purpose, neither of which included projects whose documented savings matched the final claimed savings.[53]

Recommendations 1 & 2	Eligibility rigor and auditable records

- PAs should strengthen participant eligibility criteria by enforcing existing M&V eligibility criteria—including the requirement for sufficient pre-installation data—and tightening upfront screening rules to exclude sites with known process-, event-, or production-driven loads. Eligibility screening should also include size limits, restricting participation to sites with annual energy consumption below a pre-defined threshold (for example, 1,000,000 kWh). These criteria will ensure enrollment is limited to sites suitable for standardized pop NMEC modeling, homogenous cohort formation, and compliance with NMEC rulebook precision requirements.
- PAs should propose a solution to the reporting PCG that defines a customer project-level pop NMEC dataset that links with 100% accuracy to both initial and true-up aggregate CEDARS claims by mid-September 2026. The solution should reflect an official, date-stamped set of required project claim elements, including, but not be limited to, claim and parent claim IDs; project identifiers; participant (end-user) IDs and contact information; projected[54] and true-up savings for all fuel types (kWh, kW, therms) and TSB; installed measures; EULs; realization rates; NTG values; and implementation partners (implementer, aggregator or contractor, and M&V vendor).

5.1.2 Technical improvements

Savings accuracy in pop NMEC programs also depends on the selection and application of consistent modeling methods and on consistent and appropriate characterization of uncertainty.

The use of different methods to calculate savings increased variability and risk in savings results. Because program implementers used different modeling and non-participant selection approaches, which are required to account for non-program effects, different assumptions shape savings estimates, and those estimates are sensitive to different sources of error. As a result, differences in final claimed and verified savings may reflect methodological choice as well as underlying program performance. This reduces the confidence of stakeholders, including participating customers and contractors, in cross-program comparisons and runs counter to pop NMEC's standardization intent.

[53] For pop NMEC, PAs claim aggregate program-level final savings for projects that occur in a particular program year. Documented savings refer to the meter-based savings calculated by the PAs at the end of the reporting period, which serve as the basis for final claimed savings. We refer to these as documented savings, as the PAs provide us with the supporting data, models, and results as part of the evaluation.

[54] Projected savings refer to savings forecasts made at the individual project level prior to implementation.

Approaches to modeling energy use at the site level varied in their alignment with NMEC rulebook expectations. While all evaluated program implementers used modeling approaches to account for routine, recurring drivers of consumption, such as weather, and non-program-related factors, methodological choices did not always best serve savings accuracy. For example, only one program normalized savings to typical weather conditions, consistent with NMEC rulebook intent. Other programs used reporting-year weather, which increases savings errors in multiple ways. Further, aspects of the modeling approaches used extend beyond capturing routine drivers of consumption and can also incorporate within-year trends and NREs into modeled consumption. This reduces the ability to distinguish routine consumption patterns from non-routine variation, with implications for eligibility screening and the interpretation of hourly savings profiles

Most program implementers' uncertainty methods and reporting do not provide sufficient evidence to support the credibility of claimed pop NMEC savings under the NMEC rulebook's precision expectations. Characterizing uncertainty is essential for establishing claimed pop NMEC savings are valid. Despite this, only two programs reported their calculation of uncertainty metrics. In addition, common uncertainty practices (e.g., reliance on FSU[55]) may not fully reflect sources of program-level uncertainty beyond model error, such as the uncertainty introduced by the selection of comparison sites to account for non-program effects. By this evaluation's uncertainty calculations, most programs did not meet the NMEC rulebook's 90/25 precision standard. It may be appropriate to relax this standard to a 90/50 benchmark for these non-residential cohorts with relatively small sample sizes, but this will only be reasonable if the PAs take uncertainty characterization seriously as a key indicator of program success.

Recommendations 3 & 4	Consistent methods uncertainty treatment

- The PAs and their implementers should participate in a stakeholder process to develop more consistent modeling approaches for pop NMEC programs that prioritize clear representation of weather and schedule effects, facilitate identification of NREs that influence eligibility decisions, and support accurate TSB estimates
- The PAs should apply program-level pop NMEC savings uncertainty approaches that reflect both modeling error and other sources of cross-site variability.

5.2 Stranded savings recommendations

NMEC is most effective when it targets customers with demonstrated barriers to upgrading aging, below-code equipment. Its treatment of below-code savings is justified when projects would not have occurred otherwise, whether barriers are technical or financial. By tying meter-based savings to tools such as on-bill financing, NMEC can directly address capital constraints while ensuring ratepayer support delivers genuine benefits.

The evaluated pop NMEC programs motivated a substantial share of customer savings, on average 70%, and were well received by participating customers. Overall, pop NMEC programs achieved NTGRs comparable to or higher than most non-residential program pathways, such as the Commercial, Industrial, Agricultural, and Custom pathway (CIAC) , with variation across programs suggesting that targeting hard-to-reach (HTR) customers is associated with higher influence. BayREN's Business FLEXmarket program, which focuses on HTR customers, achieved particularly high NTGRs, including 90% in one program year.

OBF survey results provide strong evidence that the program enabled or accelerated upgrades when capital barriers were central, but the program would provide more value if outcomes reflected performance-based results. OBF reaches customers who face barriers in accessing traditional financing, including municipal entities such as K-12 schools and civic facilities. Three-quarters of OBF survey respondents said they would not have been able to complete the

55 Fractional savings uncertainty (FSU) measures the uncertainty around program energy savings. As the value of FSU decreases, confidence in the estimated savings level increases.

project or do the same project at the same time without the program. That said, the program's current payback process is based on exaggerated preliminary savings estimates and undermines the program's ability to benefit more of these kinds of customers. Updating payback plans to reflect meter-based savings will fully address customer barriers while also incentivizing implementers and contractors to develop projects that deliver full value to customers.

Recommendations 5 & 6	Target customers with high stranded savings potential

- PAs should prioritize customers with demonstrated barriers (e.g., limited access to financing or constrained capital cycles), as NMEC uniquely credits "stranded savings" from to-code upgrades that are more common in these segments.
- PG&E should offer a plan that achieves program-level forecast realization rates (initial claimed savings relative to verified savings) greater than 90% by mid-September 2026. This approach will better align program incentives with realized performance, address the current misalignment between initial claimed savings and actual outcomes, and align payment schedules with actual bill-level savings.

Eighty-four percent of OBF and 77% of non-OBF interventions included only lighting upgrades, which do not fully support pop NMEC's focus on whole-building savings and delivering grid value. Despite pop NMEC's whole-building, measure-agnostic design being well suited to comprehensive retrofits, lighting upgrades dominated most programs, with multiple programs installing only lighting measures. Lighting-only projects accounted for the majority of savings (66% of OBF savings and 91% of non-OBF savings in the evaluated cohorts), while mixed-measure projects sometimes delivered higher site-level reductions (for OBF, projects with both lighting and non-lighting measures achieved higher percent reductions than lighting-only). Thus, most of the pop NMEC programs functioned more like lighting programs than comprehensive whole-building programs, which suggests that many pop NMEC programs are not fully realizing their intended role in capturing stranded savings through broader modernization.

Recommendation 7	Broader technologies for greater whole-building upgrades

PAs should look beyond lighting-only projects and increase the share of mixed-measure or comprehensive projects in pop NMEC, such as those including HVAC and refrigeration-related upgrades. Pop NMEC is designed for whole-building savings, and mixed-measure projects can deliver larger site-level reductions.

Customers deemed ineligible or with canceled projects were rarely directed to other savings pathways or programs. Eligibility processes were largely handled by implementers or aggregators, with PAs having only limited visibility into those processes. As a result, the PAs had little involvement in eligibility decisions or the disposition of customers who were not good candidates for pop NMEC programs. Generally, implementers and aggregators lacked motivation to find other opportunities for customers because they were enrolled in and incentivized by their specific pop NMEC programs.

Large customers and "non-commercial-like" sectors account for substantial savings, but they also increase pop NMEC risk and are often better suited to other pathways. Across cohorts, large sites (≥1,000,000 kWh) are a small share of participants but account for a disproportionately large share of savings. While these large sites may represent good energy efficiency program participants, they are not good fits for pop NMEC because the savings are too big and too risky for the pop NMEC environment, which uses consistent model structures across all projects. Industrial, agricultural, and utility-type customers are also better served by CIAC and strategic energy management (SEM) programs, while large commercial-like customers may be better aligned with site-level NMEC, SEM, or CIAC programs. Pop NMEC should be reserved for customers where the approach adds unique value that addresses stranded savings, rather than absorbing high-risk savings from customers that have better-fit alternatives.

Recommendation 8	Clear referral pathways for poor-fit projects
PAs should implement a consistent tracking and referral process so customers/projects screened out of pop NMEC (e.g., industrial/process loads, very large sites) are redirected to appropriate alternatives (e.g., SEM, CIAC, or site-level NMEC), and PAs should report those dispositions.	

APPENDIX A. STANDARD HIGH-LEVEL SAVINGS TABLES

This section provides tables with the gross and net lifecycle and first year savings of the evaluated non-residential pop NMEC programs.

Table A-1. Gross Lifecycle Savings (MWh)

Report name	PA	Standard report group	Ex ante gross	Ex post gross	GRR	% Ex ante gross pass through	Eval GRR
POP NMEC	BAYREN	Commercial	24,296	19,051	0.78	0.0%	0.78
POP NMEC	**BAYREN**	**Total**	**24,296**	**19,051**	**0.78**	**0.0%**	**0.78**
POP NMEC	MCE	Commercial Efficiency Market	61,895	49,984	0.81	0.0%	0.81
POP NMEC	**MCE**	**Total**	**61,895**	**49,984**	**0.81**	**0.0%**	**0.81**
POP NMEC	PCE	PCE FLEXmarket	16,720	14,498	0.87	0.0%	0.87
POP NMEC	**PCE**	**Total**	**16,720**	**14,498**	**0.87**	**0.0%**	**0.87**
POP NMEC	PGE	On-Bill Financing Alternative Pathway	212,038	185,209	0.87	0.0%	0.87
POP NMEC	PGE	Institutional Partnerships	8,920	5,842	0.65	0.0%	0.65
POP NMEC	**PGE**	**Total**	**220,958**	**191,051**	**0.86**	**0.0%**	**0.86**
POP NMEC	SCP	SCP FLEXmarket	12,414	0	0.00	0.0%	0.00
POP NMEC	**SCP**	**Total**	**12,414**	**0**	**0.00**	**0.0%**	**0.00**
POP NMEC		***Statewide***	***336,283***	***274,584***	***0.82***	***0.0%***	***0.82***

Table A-2. Net Lifecycle Savings (MWh)

Report name	PA	Standard report group	Ex ante net	Ex post net	NRR	% Ex ante net pass through	Ex ante NTG	Ex post NTG	Eval ex ante NTG	Eval ex post NTG
POP NMEC	BAYREN	Commercial	24,296	17,144	0.71	0.0%	1.00	0.90	1.00	0.90
POP NMEC	**BAYREN**	**Total**	**24,296**	**17,144**	**0.71**	**0.0%**	**1.00**	**0.90**	**1.00**	**0.90**
POP NMEC	MCE	Commercial Efficiency Market	61,895	38,804	0.63	0.0%	1.00	0.78	1.00	0.78
POP NMEC	**MCE**	**Total**	**61,895**	**38,804**	**0.63**	**0.0%**	**1.00**	**0.78**	**1.00**	**0.78**
POP NMEC	PCE	PCE FLEXmarket	16,720	7,249	0.43	0.0%	1.00	0.50	1.00	0.50
POP NMEC	**PCE**	**Total**	**16,720**	**7,249**	**0.43**	**0.0%**	**1.00**	**0.50**	**1.00**	**0.50**
POP NMEC	PGE	On-Bill Financing Alternative Pathway	209,918	122,835	0.59	0.0%	0.99	0.66	0.99	0.66
POP NMEC	PGE	Institutional Partnerships	8,920	3,699	0.41	0.0%	1.00	0.63	1.00	0.63
POP NMEC	**PGE**	**Total**	**218,838**	**126,533**	**0.58**	**0.0%**	**0.99**	**0.66**	**0.99**	**0.66**
POP NMEC	SCP	SCP FLEXmarket	12,414	0	0.00	0.0%	1.00		1.00	
POP NMEC	**SCP**	**Total**	**12,414**	**0**	**0.00**	**0.0%**	**1.00**		**1.00**	
POP NMEC		***Statewide***	***334,162***	***189,730***	***0.57***	***0.0%***	***0.99***	***0.69***	***0.99***	***0.69***

Table A-3. Gross Lifecycle Savings (MW)

Report name	PA	Standard report group	Ex ante gross	Ex post gross	GRR	% Ex ante gross pass through	Eval GRR
POP NMEC	BAYREN	Commercial	3.4	2.6	0.79	0.0%	0.79
POP NMEC	**BAYREN**	**Total**	**3.4**	**2.6**	**0.79**	**0.0%**	**0.79**
POP NMEC	MCE	Commercial Efficiency Market	5.6	3.8	0.69	0.0%	0.69
POP NMEC	**MCE**	**Total**	**5.6**	**3.8**	**0.69**	**0.0%**	**0.69**
POP NMEC	PCE	PCE FLEXmarket	1.9	0.4	0.22	0.0%	0.22
POP NMEC	**PCE**	**Total**	**1.9**	**0.4**	**0.22**	**0.0%**	**0.22**
POP NMEC	PGE	On-Bill Financing Alternative Pathway	25.1	14.5	0.58	0.0%	0.58
POP NMEC	PGE	Institutional Partnerships	1.5	0.5	0.33	0.0%	0.33
POP NMEC	**PGE**	**Total**	**26.6**	**15.0**	**0.56**	**0.0%**	**0.56**
POP NMEC	SCP	SCP FLEXmarket	1.4	0.0	0.00	0.0%	0.00
POP NMEC	**SCP**	**Total**	**1.4**	**0.0**	**0.00**	**0.0%**	**0.00**
POP NMEC		***Statewide***	***38.9***	***21.9***	***0.56***	***0.0%***	***0.56***

Table A-4. Net Lifecycle Savings (MW)

Report name	PA	Standard report group	Ex ante net	Ex post net	NRR	% Ex ante net pass through	Ex ante NTG	Ex post NTG	Eval ex ante NTG	Eval ex post NTG
POP NMEC	BAYREN	Commercial	3.4	2.4	0.71	0.0%	1.00	0.90	1.00	0.90
POP NMEC	**BAYREN**	**Total**	**3.4**	**2.4**	**0.71**	**0.0%**	**1.00**	**0.90**	**1.00**	**0.90**
POP NMEC	MCE	Commercial Efficiency Market	5.6	3.8	0.68	0.0%	1.00	0.98	1.00	0.98
POP NMEC	**MCE**	**Total**	**5.6**	**3.8**	**0.68**	**0.0%**	**1.00**	**0.98**	**1.00**	**0.98**
POP NMEC	PCE	PCE FLEXmarket	1.9	0.2	0.11	0.0%	1.00	0.50	1.00	0.50
POP NMEC	**PCE**	**Total**	**1.9**	**0.2**	**0.11**	**0.0%**	**1.00**	**0.50**	**1.00**	**0.50**
POP NMEC	PGE	On-Bill Financing Alternative Pathway	24.8	9.6	0.39	0.0%	0.99	0.67	0.99	0.67
POP NMEC	PGE	Institutional Partnerships	1.5	0.3	0.21	0.0%	1.00	0.63	1.00	0.63
POP NMEC	**PGE**	**Total**	**26.4**	**10.0**	**0.38**	**0.0%**	**0.99**	**0.66**	**0.99**	**0.66**
POP NMEC	SCP	SCP FLEXmarket	1.4	0.0	0.00	0.0%	1.00		1.00	
POP NMEC	**SCP**	**Total**	**1.4**	**0.0**	**0.00**	**0.0%**	**1.00**		**1.00**	
POP NMEC		***Statewide***	***38.6***	***16.3***	***0.42***	***0.0%***	***0.99***	***0.75***	***0.99***	***0.75***

Table A-5. Gross Lifecycle Savings (MTherms)

Report name	PA	Standard report group	Ex ante gross	Ex post gross	GRR	% Ex ante gross pass through	Eval GRR
POP NMEC	BAYREN	Commercial	0	0			
POP NMEC	**BAYREN**	**Total**	**0**	**0**			
POP NMEC	MCE	Commercial Efficiency Market	23	0	0.00	0.0%	0.00
POP NMEC	**MCE**	**Total**	**23**	**0**	**0.00**	**0.0%**	**0.00**
POP NMEC	PCE	PCE FLEXmarket	0	0			
POP NMEC	**PCE**	**Total**	**0**	**0**			
POP NMEC	PGE	On-Bill Financing Alternative Pathway	565	347	0.61	0.0%	0.61
POP NMEC	PGE	Institutional Partnerships	0	0			
POP NMEC	**PGE**	**Total**	**565**	**347**	**0.61**	**0.0%**	**0.61**
POP NMEC	SCP	SCP FLEXmarket	0	0			
POP NMEC	**SCP**	**Total**	**0**	**0**			
POP NMEC		***Statewide***	***588***	***347***	***0.59***	***0.0%***	***0.59***

Table A-6. Net Lifecycle Savings (MTherms)

Report name	PA	Standard report group	Ex ante net	Ex post net	NRR	% Ex ante net pass through	Ex ante NTG	Ex post NTG	Eval ex ante NTG	Eval ex post NTG
POP NMEC	BAYREN	Commercial	0	0						
POP NMEC	**BAYREN**	**Total**	**0**	**0**						
POP NMEC	MCE	Commercial Efficiency Market	26	0	0.00	0.0%	1.11		1.11	
POP NMEC	**MCE**	**Total**	**26**	**0**	**0.00**	**0.0%**	**1.11**		**1.11**	
POP NMEC	PCE	PCE FLEXmarket	0	0						
POP NMEC	**PCE**	**Total**	**0**	**0**						
POP NMEC	PGE	On-Bill Financing Alternative Pathway	682	237	0.35	0.0%	1.21	0.68	1.21	0.68
POP NMEC	PGE	Institutional Partnerships	0	0						
POP NMEC	**PGE**	**Total**	**682**	**237**	**0.35**	**0.0%**	**1.21**	**0.68**	**1.21**	**0.68**
POP NMEC	SCP	SCP FLEXmarket	0	0						
POP NMEC	**SCP**	**Total**	**0**	**0**						
POP NMEC		***Statewide***	***708***	***237***	***0.33***	***0.0%***	***1.20***	***0.68***	***1.20***	***0.68***

Table A-7. Gross First Year Savings (MWh)

Report name	PA	Standard report group	Ex ante gross	Ex post gross	GRR	% Ex ante gross pass through	Eval GRR
POP NMEC	BAYREN	Commercial	2,025	1,588	0.78	0.0%	0.78
POP NMEC	**BAYREN**	**Total**	**2,025**	**1,588**	**0.78**	**0.0%**	**0.78**
POP NMEC	MCE	Commercial Efficiency Market	4,902	4,581	0.93	0.0%	0.93
POP NMEC	**MCE**	**Total**	**4,902**	**4,581**	**0.93**	**0.0%**	**0.93**
POP NMEC	PCE	PCE FLEXmarket	1,393	1,208	0.87	0.0%	0.87
POP NMEC	**PCE**	**Total**	**1,393**	**1,208**	**0.87**	**0.0%**	**0.87**
POP NMEC	PGE	On-Bill Financing Alternative Pathway	17,842	16,499	0.92	0.0%	0.92
POP NMEC	PGE	Institutional Partnerships	743	487	0.65	0.0%	0.65
POP NMEC	**PGE**	**Total**	**18,586**	**16,986**	**0.91**	**0.0%**	**0.91**
POP NMEC	SCP	SCP FLEXmarket	1,014	0	0.00	0.0%	0.00
POP NMEC	**SCP**	**Total**	**1,014**	**0**	**0.00**	**0.0%**	**0.00**
POP NMEC		***Statewide***	***27,920***	***24,362***	***0.87***	***0.0%***	***0.87***

Table A-8. Net First Year Savings (MWh)

Report name	PA	Standard report group	Ex ante net	Ex post net	NRR	% Ex ante net pass through	Ex ante NTG	Ex post NTG	Eval ex ante NTG	Eval ex post NTG
POP NMEC	BAYREN	Commercial	2,025	1,429	0.71	0.0%	1.00	0.90	1.00	0.90
POP NMEC	**BAYREN**	**Total**	**2,025**	**1,429**	**0.71**	**0.0%**	**1.00**	**0.90**	**1.00**	**0.90**
POP NMEC	MCE	Commercial Efficiency Market	4,902	3,510	0.72	0.0%	1.00	0.77	1.00	0.77
POP NMEC	**MCE**	**Total**	**4,902**	**3,510**	**0.72**	**0.0%**	**1.00**	**0.77**	**1.00**	**0.77**
POP NMEC	PCE	PCE FLEXmarket	1,393	604	0.43	0.0%	1.00	0.50	1.00	0.50
POP NMEC	**PCE**	**Total**	**1,393**	**604**	**0.43**	**0.0%**	**1.00**	**0.50**	**1.00**	**0.50**
POP NMEC	PGE	On-Bill Financing Alternative Pathway	17,664	10,935	0.62	0.0%	0.99	0.66	0.99	0.66
POP NMEC	PGE	Institutional Partnerships	743	308	0.41	0.0%	1.00	0.63	1.00	0.63
POP NMEC	**PGE**	**Total**	**18,407**	**11,243**	**0.61**	**0.0%**	**0.99**	**0.66**	**0.99**	**0.66**
POP NMEC	SCP	SCP FLEXmarket	1,014	0	0.00	0.0%	1.00		1.00	
POP NMEC	**SCP**	**Total**	**1,014**	**0**	**0.00**	**0.0%**	**1.00**		**1.00**	
POP NMEC		***Statewide***	***27,741***	***16,785***	***0.61***	***0.0%***	***0.99***	***0.69***	***0.99***	***0.69***

Table A-9. Gross First Year Savings (MW)

Report name	PA	Standard report group	Ex ante gross	Ex post gross	GRR	% Ex ante gross pass through	Eval GRR
POP NMEC	BAYREN	Commercial	0.3	0.2	0.79	0.0%	0.79
POP NMEC	**BAYREN**	**Total**	**0.3**	**0.2**	**0.79**	**0.0%**	**0.79**
POP NMEC	MCE	Commercial Efficiency Market	0.5	0.3	0.69	0.0%	0.69
POP NMEC	**MCE**	**Total**	**0.5**	**0.3**	**0.69**	**0.0%**	**0.69**
POP NMEC	PCE	PCE FLEXmarket	0.2	0.0	0.22	0.0%	0.22
POP NMEC	**PCE**	**Total**	**0.2**	**0.0**	**0.22**	**0.0%**	**0.22**
POP NMEC	PGE	On-Bill Financing Alternative Pathway	2.0	1.3	0.65	0.0%	0.65
POP NMEC	PGE	Institutional Partnerships	0.1	0.0	0.38	0.0%	0.38
POP NMEC	**PGE**	**Total**	**2.1**	**1.3**	**0.63**	**0.0%**	**0.63**
POP NMEC	SCP	SCP FLEXmarket	0.1	0.0	0.00	0.0%	0.00
POP NMEC	**SCP**	**Total**	**0.1**	**0.0**	**0.00**	**0.0%**	**0.00**
POP NMEC		***Statewide***	***3.1***	***1.9***	***0.61***	***0.0%***	***0.61***

Table A-10. Net First Year Savings (MW)

Report name	PA	Standard report group	Ex ante net	Ex post net	NRR	% Ex ante net pass through	Ex ante NTG	Ex post NTG	Eval ex ante NTG	Eval ex post NTG
POP NMEC	BAYREN	Commercial	0.3	0.2	0.71	0.0%	1.00	0.90	1.00	0.90
POP NMEC	**BAYREN**	**Total**	**0.3**	**0.2**	**0.71**	**0.0%**	**1.00**	**0.90**	**1.00**	**0.90**
POP NMEC	MCE	Commercial Efficiency Market	0.5	0.3	0.68	0.0%	1.00	0.98	1.00	0.98
POP NMEC	**MCE**	**Total**	**0.5**	**0.3**	**0.68**	**0.0%**	**1.00**	**0.98**	**1.00**	**0.98**
POP NMEC	PCE	PCE FLEXmarket	0.2	0.0	0.11	0.0%	1.00	0.50	1.00	0.50
POP NMEC	**PCE**	**Total**	**0.2**	**0.0**	**0.11**	**0.0%**	**1.00**	**0.50**	**1.00**	**0.50**
POP NMEC	PGE	On-Bill Financing Alternative Pathway	1.9	0.8	0.43	0.0%	0.99	0.67	0.99	0.67
POP NMEC	PGE	Institutional Partnerships	0.1	0.0	0.24	0.0%	1.00	0.63	1.00	0.63
POP NMEC	**PGE**	**Total**	**2.1**	**0.9**	**0.42**	**0.0%**	**0.99**	**0.66**	**0.99**	**0.66**
POP NMEC	SCP	SCP FLEXmarket	0.1	0.0	0.00	0.0%	1.00		1.00	
POP NMEC	**SCP**	**Total**	**0.1**	**0.0**	**0.00**	**0.0%**	**1.00**		**1.00**	
POP NMEC		***Statewide***	***3.1***	***1.4***	***0.45***	***0.0%***	***0.99***	***0.74***	***0.99***	***0.74***

Table A-11. Gross First Year Savings (MTherms)

Report name	PA	Standard report group	Ex ante gross	Ex post gross	GRR	% Ex ante gross pass through	Eval GRR
POP NMEC	BAYREN	Commercial	0	0			
POP NMEC	**BAYREN**	**Total**	**0**	**0**			
POP NMEC	MCE	Commercial Efficiency Market	5	0	0.00	0.0%	0.00
POP NMEC	**MCE**	**Total**	**5**	**0**	**0.00**	**0.0%**	**0.00**
POP NMEC	PCE	PCE FLEXmarket	0	0			
POP NMEC	**PCE**	**Total**	**0**	**0**			
POP NMEC	PGE	On-Bill Financing Alternative Pathway	64	29	0.46	0.0%	0.46
POP NMEC	PGE	Institutional Partnerships	0	0			
POP NMEC	**PGE**	**Total**	**64**	**29**	**0.46**	**0.0%**	**0.46**
POP NMEC	SCP	SCP FLEXmarket	0	0			
POP NMEC	**SCP**	**Total**	**0**	**0**			
POP NMEC		***Statewide***	***69***	***29***	***0.42***	***0.0%***	***0.42***

Table A-12. Net First Year Savings (MTherms)

Report name	PA	Standard report group	Ex ante net	Ex post net	NRR	% Ex ante net pass through	Ex ante NTG	Ex post NTG	Eval ex ante NTG	Eval ex post NTG
POP NMEC	BAYREN	Commercial	0	0						
POP NMEC	**BAYREN**	**Total**	**0**	**0**						
POP NMEC	MCE	Commercial Efficiency Market	5	0	0.00	0.0%	1.00		1.00	
POP NMEC	**MCE**	**Total**	**5**	**0**	**0.00**	**0.0%**	**1.00**		**1.00**	
POP NMEC	PCE	PCE FLEXmarket	0	0						
POP NMEC	**PCE**	**Total**	**0**	**0**						
POP NMEC	PGE	On-Bill Financing Alternative Pathway	63	20	0.31	0.0%	0.99	0.68	0.99	0.68
POP NMEC	PGE	Institutional Partnerships	0	0						
POP NMEC	**PGE**	**Total**	**63**	**20**	**0.31**	**0.0%**	**0.99**	**0.68**	**0.99**	**0.68**
POP NMEC	SCP	SCP FLEXmarket	0	0						
POP NMEC	**SCP**	**Total**	**0**	**0**						
POP NMEC		***Statewide***	***69***	***20***	***0.29***	***0.0%***	***0.99***	***0.68***	***0.99***	***0.68***

APPENDIX B. STANDARD PER-UNIT SAVINGS TABLES

This section provides tables with the per unit (quantity) gross and net energy savings of the evaluated non-residential pop NMEC programs.

Table B-1. Per Unit (Quantity) Gross Energy Savings (kWh)

Report name	PA	Standard report group	Pass through	% ER ex ante	% ER ex post	Average EUL (yr)	Ex post lifecycle	Ex post first year	Ex post annualized
POP NMEC	BAYREN	Commercial	0	0.0%	0.0%	12.0	9,525,546.0	793,795.5	793,795.5
POP NMEC	MCE	Commercial Efficiency Market	0	0.0%	0.0%	12.0	4,147,458.0	345,621.5	345,621.5
POP NMEC	MCE	Commercial Efficiency Market	0	0.0%	0.0%	10.4	6,678,927.0	639,616.6	639,616.6
POP NMEC	PCE	PCE FLEXmarket	0	0.0%	0.0%	12.0	2,899,526.4	241,627.2	241,627.2
POP NMEC	PGE	On-Bill Financing Alternative Pathway	0	0.0%	0.0%	11.0	30,885,040.8	2,820,550.6	2,820,550.6
POP NMEC	PGE	On-Bill Financing Alternative Pathway	0	0.0%	0.0%	11.8	15,417,135.7	1,304,193.3	1,304,193.3
POP NMEC	PGE	Institutional Partnerships	0	0.0%	0.0%	12.0	730,260.0	60,855.0	60,855.0
POP NMEC	SCP	SCP FLEXmarket	0	0.0%	0.0%	0.0	0.0	0.0	0.0

Table B-2. Per Unit (Quantity) Gross Energy Savings (Therms)

Report name	PA	Standard report group	Pass through	% ER ex ante	% ER ex post	Average EUL (yr)	Ex post lifecycle	Ex post first year	Ex post annualized
POP NMEC	BAYREN	Commercial	0	0.0%	0.0%	12.0	0.0	0.0	0.0
POP NMEC	MCE	Commercial Efficiency Market	0	0.0%	0.0%	12.0	0.0	0.0	0.0
POP NMEC	MCE	Commercial Efficiency Market	0	0.0%	0.0%	10.4	0.0	0.0	0.0
POP NMEC	PCE	PCE FLEXmarket	0	0.0%	0.0%	12.0	0.0	0.0	0.0
POP NMEC	PGE	On-Bill Financing Alternative Pathway	0	0.0%	0.0%	11.0	0.0	0.0	0.0
POP NMEC	PGE	On-Bill Financing Alternative Pathway	0	0.0%	0.0%	11.8	86,847.7	7,296.5	7,346.8
POP NMEC	PGE	Institutional Partnerships	0	0.0%	0.0%	12.0	0.0	0.0	0.0
POP NMEC	SCP	SCP FLEXmarket	0	0.0%	0.0%	0.0	0.0	0.0	0.0

Table B-3. Per Unit (Quantity) Net Energy Savings (kWh)

Report name	PA	Standard report group	Pass through	% ER ex ante	% ER ex post	Average EUL (yr)	Ex post lifecycle	Ex post first year	Ex post annualized
POP NMEC	BAYREN	Commercial	0	0.0%	0.0%	12.0	8,572,038.0	714,336.5	714,336.5
POP NMEC	MCE	Commercial Efficiency Market	0	0.0%	0.0%	12.0	4,147,458.0	345,621.5	345,621.5
POP NMEC	MCE	Commercial Efficiency Market	0	0.0%	0.0%	10.4	4,442,822.6	425,473.0	425,473.0
POP NMEC	PCE	PCE FLEXmarket	0	0.0%	0.0%	12.0	1,449,763.2	120,813.6	120,813.6
POP NMEC	PGE	On-Bill Financing Alternative Pathway	0	0.0%	0.0%	11.0	20,192,639.3	1,844,076.0	1,844,076.0
POP NMEC	PGE	On-Bill Financing Alternative Pathway	0	0.0%	0.0%	11.8	10,516,028.2	889,590.2	889,590.2
POP NMEC	PGE	Institutional Partnerships	0	0.0%	0.0%	12.0	462,327.6	38,527.3	38,527.3
POP NMEC	SCP	SCP FLEXmarket	0	0.0%	0.0%	0.0	0.0	0.0	0.0

Table B-4. Per Unit (Quantity) Net Energy Savings (Therms)

Report name	PA	Standard report group	Pass through	% ER ex ante	% ER ex post	Average EUL (yr)	Ex post lifecycle	Ex post first year	Ex post annualized
POP NMEC	BAYREN	Commercial	0	0.0%	0.0%	12.0	0.0	0.0	0.0
POP NMEC	MCE	Commercial Efficiency Market	0	0.0%	0.0%	12.0	0.0	0.0	0.0
POP NMEC	MCE	Commercial Efficiency Market	0	0.0%	0.0%	10.4	0.0	0.0	0.0
POP NMEC	PCE	PCE FLEXmarket	0	0.0%	0.0%	12.0	0.0	0.0	0.0
POP NMEC	PGE	On-Bill Financing Alternative Pathway	0	0.0%	0.0%	11.0	0.0	0.0	0.0
POP NMEC	PGE	On-Bill Financing Alternative Pathway	0	0.0%	0.0%	11.8	59,238.8	4,976.9	5,011.2
POP NMEC	PGE	Institutional Partnerships	0	0.0%	0.0%	12.0	0.0	0.0	0.0
POP NMEC	SCP	SCP FLEXmarket	0	0.0%	0.0%	0.0	0.0	0.0	0.0

APPENDIX C. BILL IMPACTS

This bill impact analysis estimates how changes in electricity consumption associated with pop NMEC program participation affect participants' energy costs. Rather than reproducing the exact bill changes customers would observe on their utility statements, the analysis provides insight into the direction and approximate magnitude of energy-cost impacts due to program-related changes in energy consumption.

All the non-residential pop NMEC participants with trued-up claims (PY2022 and PY2023) operate within PG&E's service territory and are subject to PG&E's non-residential tariffs. We obtained the applicable rates from the PA M&V documents for OBF participants and from PG&E billing data for other program participants.

Participant rate codes map to a limited set of PG&E tariff families that differ by TOU structure and the presence of demand charges, with assignment primarily determined by monthly demand (kW) thresholds. In general, small customers faced energy-only TOU rates while medium and large customers faced both. We used PG&E's tariff sheets to obtain the rate components required for bill calculations, including energy and demand charges, seasonal definitions, and TOU period windows.[56] Table C-1 summarizes the applicable tariff families for pop NMEC participants, showing the demand thresholds, customer size categories, and the number of participants subject to each tariff type.[57]

Table C-1. Applicable tariff details and site characteristics for pop NMEC participants

Tariff family	Tariff name	Monthly demand threshold (kW)	Customer size category	Number of participants
B-1	Small General Service	< 75 kW	Small	412
B-6	Small General Service (TOU)	< 75 kW	Small	138
B-10	Medium General Demand-Metered	< 500 kW	Medium	242
B-19	Medium General Demand-Metered (TOU)	< 1,000 kW	Medium	206
B-20	Large Customer (Demand ≥ 1,000 kW)	≥ 1,000 kW	Large	6
AG-A1 / AG-A2	Agricultural Power (TOU)	< 35 kW	Small	5
AG-B	Agricultural Power (TOU)	≥ 35 kW	Medium/Large*	2
AG-C	Agricultural Power (TOU)	≥ 35 kW	Medium/Large*	23
OL	Outdoor Area Lighting	N/A	N/A	6
SB	Standby Service	N/A	N/A	7
Unclassified	Missing/unknown rate code	N/A	N/A	33

*AG-B vs AG-C depends on usage intensity (load factor and operating hours), not kW alone

Figure C-1 highlights clear differences in the rate families served by each program to help explain observed differences in per participant bill impacts. The PCE and MCE programs primarily serve customers on demand metered TOU rates (notably B-19), while the BayREN and PG&E Institutional Partnerships programs include a larger share of customers on small general service rates (B-1 and B-6) without demand charge components. The PG&E OBF electric program spans the broadest mix of rate families, reflecting its wide customer reach. Because demand metered TOU rates expose customers to

56 PG&E Electric Rate Schedules.

57 To focus on the bill impacts of consumption changes rather than reconstructing full customer bills, the analysis applied several simplifying assumptions. We applied the same tariff-based rates to all customers, regardless of whether they received bundled service, Direct Access, or Community Choice Aggregation supply. For tariffs with demand components, we approximated monthly demand using the highest hourly kW observed in each month, rather than 15-minute intervals, because 15-minute interval demand data were not available. Where tariffs vary by service voltage, we applied secondary-voltage rates because the available data did not identify participant voltage levels. The bill impact calculations reflect energy and demand charges only and exclude power factor adjustments and Peak Day Pricing components, since information on customer enrollment in those tariffs was not available. We also did not calculate bill impacts for customers with tariffs that include usage-independent charges, such as outdoor lighting tariffs with per-lamp pricing.

demand charges, programs with greater participation on these tariffs tend to realize higher bill reductions per participant for comparable energy savings.

Figure C-1. Share of customers across rate families by program

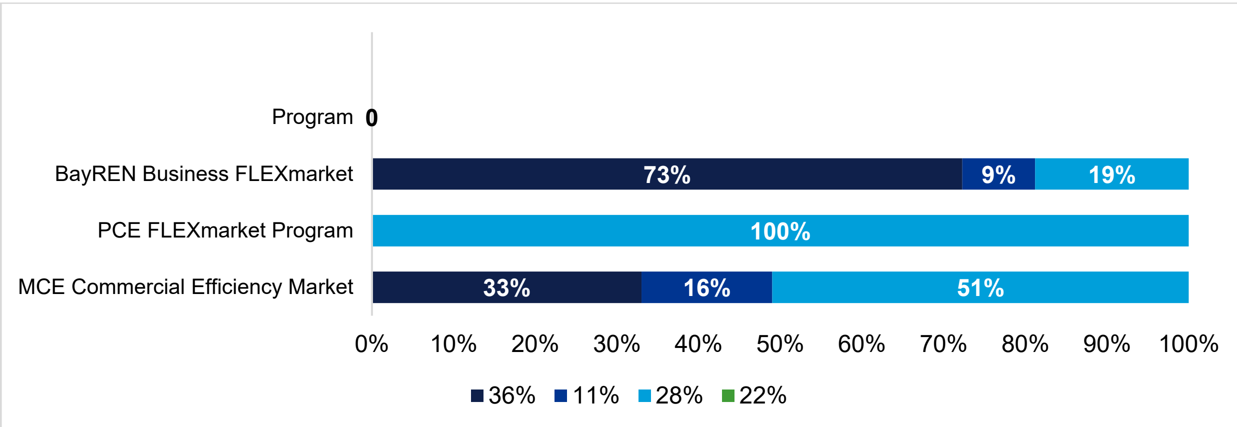


Table C-2 shows that total and average per site bill savings vary substantially across program styles, reflecting differences in program scale, participant mix, and underlying rate exposure. OBF generates the largest aggregate bill savings because it is the largest program by participation, while average per site savings remain moderate, indicating that total impacts accrue primarily through scale rather than large site-level reductions.

In contrast, MAP-style programs exhibit a wide range of average per site outcomes. The PCE program achieves very high average per site bill savings, consistent with participation concentrated among larger customers on demand-metered TOU rates, while the BayREN program shows lower average per site bill impacts due to a greater share of smaller customers on small general service rates. The MCE program achieves average per-site bill savings between the two, reflecting a participant mix dominated by medium-sized customers with exposure to demand charges. Combined with relatively high total bill savings, this pattern suggests a balance between program scale and site-level bill reductions.

Overall, the results indicate that programs differ not only in the total bill savings they deliver, but also in the distribution of bill savings across sites based on participants' rate exposure.

Table C-2. Total and average per-site annual bill savings by program and program year

Style	PA	Program	PY	Total bill savings ($/year)	Average bill savings ($/site/year)
MAP-style	BayREN	BayREN Business FLEXmarket	2023	442,726	6,325
	PCE	PCE FLEXmarket Program	2023	220,774	73,591
	MCE	MCE Commercial Efficiency Market	2022	481,740	17,842
			2023	548,810	16,631
	SCP	SCP FLEXmarket	2023	0	0
Total				**1,694,049**	**21,705**
Non-MAP Style	PG&E	PG&E Institutional Partnerships	2023	143,525	23,921
Total				**143,525**	**23,921**
OBF	PG&E	PG&E On-Bill Financing	2022	3,673,571	9,371
			2023	3,157,273	12,430
Total				**6,830,845**	**10,785**

Table C-3 shows that average participant bill savings vary systematically with customer characteristics, reinforcing the role of customer size, location, and measure mix in shaping bill impacts across program types. Across all programs, larger customers consistently realize higher average per-site bill savings than smaller customers, reflecting larger absolute savings and greater exposure to demand charges. MAP-style programs exhibit the widest spread in participant outcomes, with particularly large bill savings among medium and large customers, while smaller customers experience lower impacts. OBF follows a similar size-based pattern but with generally lower average per-site bill savings. Differences by location suggest that climate and load characteristics also influence bill outcomes, while variation by measure group indicates that the type of installed measures affects how energy savings translate into bill reductions.

Table C-3. Average per-site bill savings by customer characteristics

Characteristic	MAP-style	Non-MAP Style	OBF
Customer Size			
Small	2,030	-	2,586
Medium	25,891	23,921	17,361
Large	81,269	-	46,260
Location			
Coastal/Marine	13,589	-	8,129
Inland basin	5,878	31,064	8,155
Central Valley	13,161	16,778	12,959
Mountain	-	-	4,182
Measure Group			
Lighting	15,693	23,921	9,446
Non-Lighting	6,583	-	16,726
Both	2,886	-	15,425
Organization Type			
Commercial and retail	12,406	23,921	11,242
Institutions	(5,441)	-	9,541
Industrial, agricultural, and utilities	20,464	-	11,132

APPENDIX D. DETAILS ON PEAK DEMAND IMPACTS

The PY2023 MCE Commercial Efficiency program exhibited the lowest peak-impact realization rate among the programs evaluated. To illustrate how peak impacts are calculated for this program, the figures below present peak-period weather conditions and corresponding hourly savings by climate zone. In each figure, the top panel shows daily temperatures from June through September for the calendar years overlapping participating sites' performance periods, with blue callout boxes identifying the consecutive three-day heatwave defined using the DEER peak period methodology and site-specific weather data. The bottom panel shows aggregate hourly savings during the identified peak period for all sites with overlapping performance periods, with gray shaded regions indicating the 4 PM – 9 PM peak hours.

Figure D-1 illustrates the identified peak period and corresponding aggregate hourly savings for MCE PY2023 sites in climate zone 2, showing savings under observed heatwave conditions. During most of the peak period hours, the figure indicates savings in this climate zone, consistent with the overall positive peak demand impact estimated in the evaluation for this climate zone.

Figure D-1. Peak period weather and hourly impacts for MCE PY2023, climate zone 2

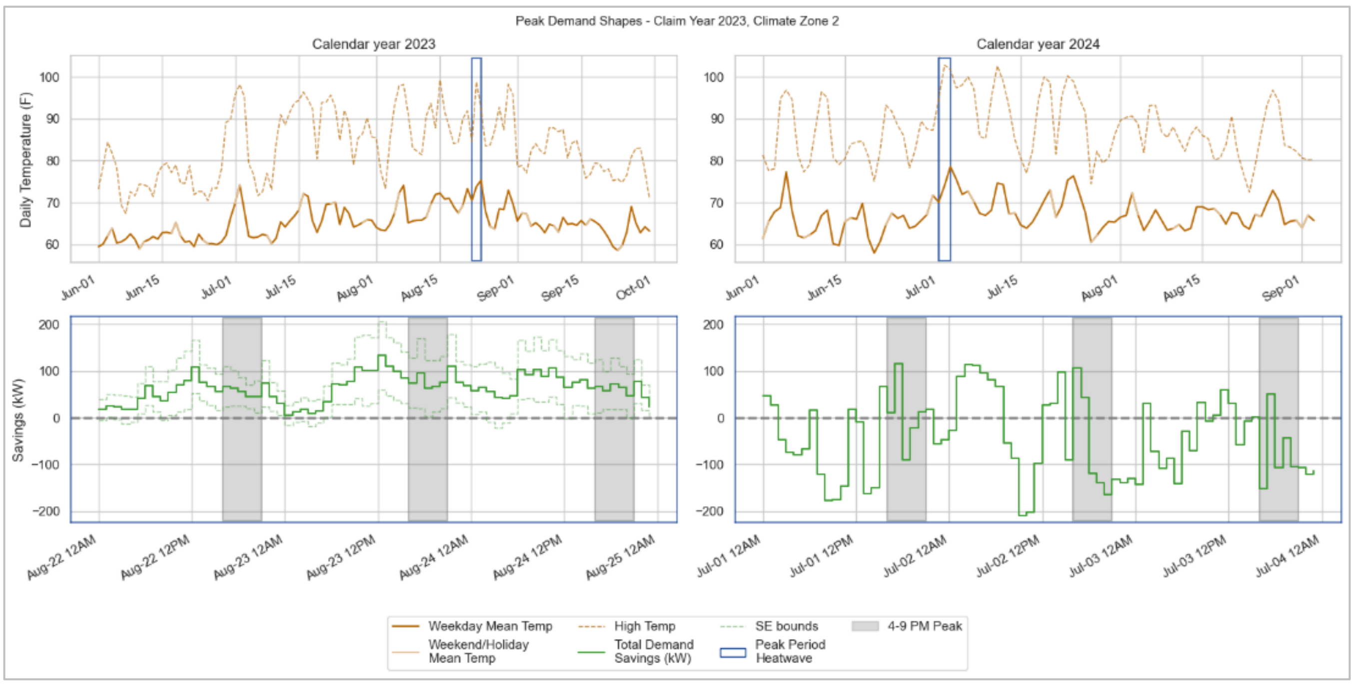


Figure D-2 presents the same peak-period weather and hourly savings analysis for climate zone 3, highlighting differences in peak-hour savings patterns relative to temperature conditions. The figure shows savings during most of the identified peak period hours, consistent with the overall positive peak demand impact estimated in the evaluation for this climate zone.

Figure D-2. Peak period weather and hourly impacts for MCE PY2023, climate zone 3

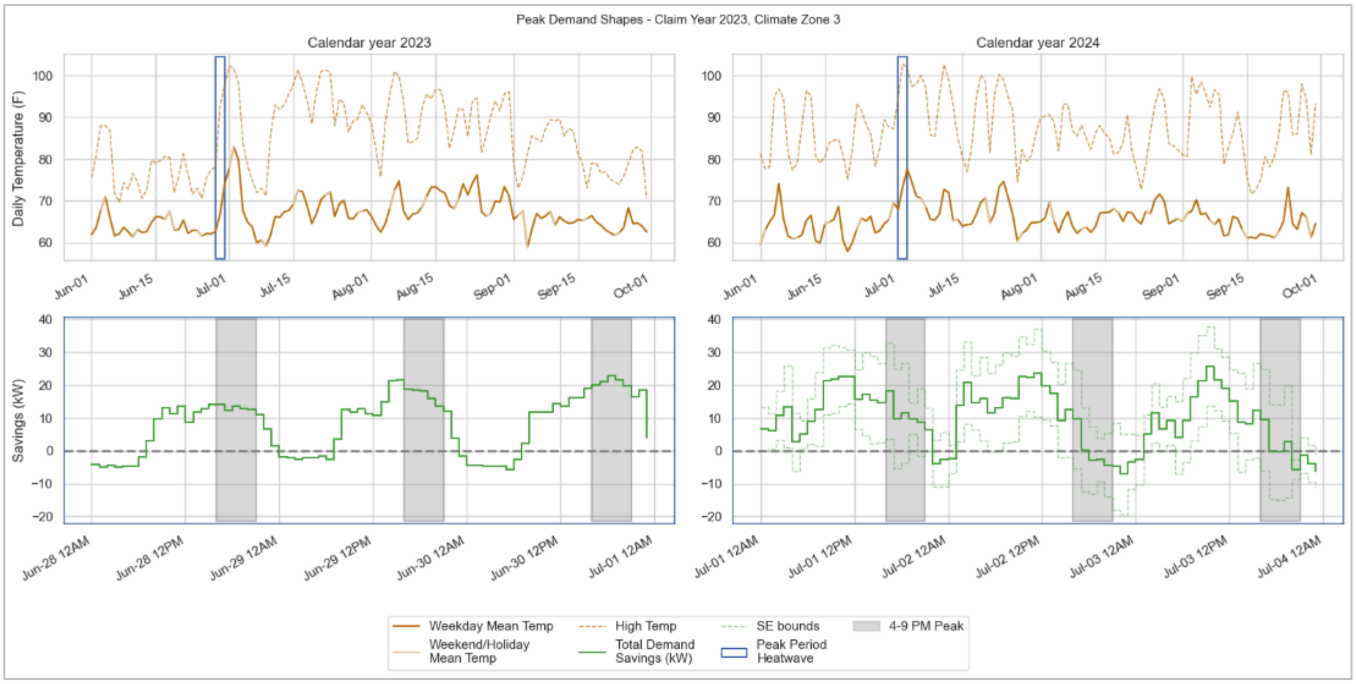
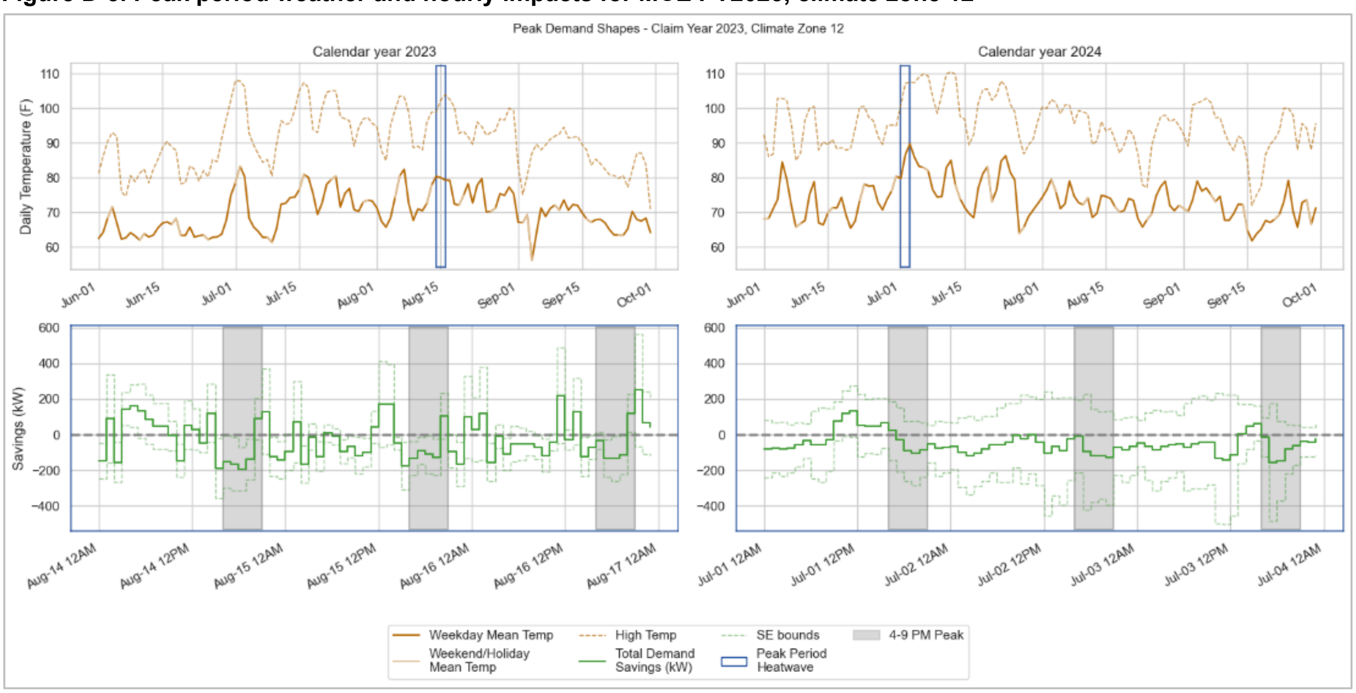


Figure D-3 shows results for climate zone 12, where aggregate savings during the identified peak period hours are negative, consistent with the overall negative peak impact estimated in the evaluation for this climate zone. Given the larger per-site impacts in climate zones 2 and 3 than in this climate zone, this program cohort shows overall peak demand savings.

Figure D-3. Peak period weather and hourly impacts for MCE PY2023, climate zone 12



APPENDIX E. NTGR SAMPLE DESIGN AND EXPANSION

This section summarizes our NTGR sample design, response rates, and expansion methodology.

Sample design

In contrast to other commercial programs, pop NMEC claims are made at the program cohort level. Consequently, we could not use the tracking data to produce the NTGR sample design and instead needed to rely on the project-level participant information provided by the PAs. The project-level documentation the PAs provided included only first-year electric savings (kWh) and did not include other fuel types, so we allocated sample points using electric savings only. While preparing the data, we aggregated the information to the project level and identified participants with multiple projects (clusters). We stratified the sample based on program, initial claim year, and project size, resulting in 26 strata.

The first step was to define the population, which was all projects from program years 2022 – 2024 from seven distinct non-residential pop NMEC programs. Table E-1 presents the stratified sample design including the number of accounts in the populations, number of accounts in the sample frame, and total projected kWh savings for each stratum.

Table E-1. Sample design

PA	Program	PY	Size Stratum	Population	Sample	Total Projected kWh
BayREN	Business FLEXmarket	2023	1	64	13	847,227
		2023	2	1	1	292,665
		2024	1	218	16	1,453,012
PCE	FLEXmarket Program	2023	1	1	1	24,810
		2023	2	3	3	1,027,351
		2024	1	5	3	111,208
		2024	2	2	2	119,424
MCE	Commercial Efficiency Market	2022	1	8	4	416,989
		2022	2	1	1	289,331
		2023	1	51	14	3,830,555
		2023	2	1	1	422,943
		2024	1	9	4	1,577,083
		2024	2	1	1	473,476
SDGE	CORE Market Access Program	2024	1	1	1	64,332
		2024	2	2	2	356,940
PG&E	Institutional Partnerships	2023	1	6	4	509,853
		2023	2	1	1	242,310
		2024	1	5	3	219,877
		2024	2	1	1	90,056
PG&E	NetOne Commercial Efficiency	2024	1	36	15	8,042,136
		2024	2	1	1	728,131
PG&E	OBF	2022	1	384	21	28,352,540
		2022	2	1	1	2,491,316
		2023	1	278	19	21,052,631
		2024	1	231	18	15,673,840
		2024	2	1	1	1,945,328

Sample design expansion

Weights were developed after collapsing stratum with no completes. Stratums were combined with the next closest option, usually other stratums under the same PA, program, and project year. Table E-2 presents the final stratification including the number of participants in the population and sample, and weight.

Table E-2. Post-expansion results

PA	Program	PY	Stratum	Population	Sample	Weights
BayREN	**Business FLEXmarket**	**2023**	1	65	15	4.3
		2024	1	218	16	13.6
PCE	**FLEXmarket Program**	**2023**	1	4	3	1.3
		2024	1	5	1	5.0
		2024	2	2	2	1.0
MCE	**Commercial Efficiency Market**	**2022**	1	9	1	9.0
		2023	1	51	13	3.9
		2023	2	1	1	1.0
		2024	1	9	4	2.3
		2024	2	1	1	1.0
SDGE	**CORE Market Access Program**	**2024**	1	3	2	1.5
PG&E	**Institutional Partnerships**	**2023**	1	6	5	1.2
		2023	2	1	1	1.0
		2024	1	5	1	5.0
		2024	2	1	1	1.0
PG&E	**NetOne Commercial Efficiency**	**2024**	1	37	15	2.5
PG&E	**OBF**	**2022**	1	385	22	17.5
		2023	1	278	21	13.2
		2024	1	232	37	6.3

APPENDIX F. DETAILED NTGR SCORING AND RESULTS

In this appendix, we present the detailed NTGR scoring methodology and detailed NTGR results.

NTGR Scoring

The NTGR is calculated by averaging three program attribution indexes (PAI_1, PAI_2, PAI_3). Table F-1 presents the methodology used to calculate each PAI, which mirrors the approach used in the site-level NMEC evaluation. The approach builds on a well-established method used for nearly a decade to evaluate commercial programs.

Table F-1. NTGR scoring methodology

Score	Description	Calculation
Program attribution index 1 (PAI_1)	• PAI_1 reflects two different ways of measuring the relative influence of program and non-program factors on project decision making. • Respondents rate each program influence factor and nonprogram influence factor using a 0-to-10 scale, where 0 means “not at all important” and 10 means “extremely important.” PIF is the maximum score given to any program influence. NPIF is the maximum score given to any non-program influence. • Respondents also divvy up ten points between their collective program factors and their collective non-program factors. • PIP is the number of points given to the collective program factors. NPIP is the number of points given to the collective non-program factors. • This approach is the same of OBF and non-OBF programs.	$PAI_1 = 10 * \frac{PIF * PIP}{((PIF * PIP) + (NPIF * NPIP))}$
Program attribution index 2 (PAI_2)	• PAI_2 reflects the prior plans of a participant, namely if a site had plans in place and budget set aside for the capital project prior to interacting with the NMEC program. • If the respondent says their organization did not know of the opportunities included in their project before interacting with the program, PAI_2 is set to 10. • If they say they knew of the opportunity before interacting with the program they are asked if they made the decision to do the project before or after beginning discission with the program. For OBF, their timing of the decision is based on when they included the project cost in their capital budgets. If they indicate the decision was made before discussions with the program, PAI_2 is set to 0. If they indicate “after” PAI_2 is set to 10. If they say, “mixed” PAI_2 is set to 5. • If they give a mixed response, PAI_2 is set to 5. • This approach is slightly different for the non-OBF participants. If they indicated the decision to complete the project was made before discussions around the NMEC program, PAI_2 is set to 0. If the decision was made after, PAI_2 is set to 10. There is no option for a “mixed” PAI_2 score of 5.	Learned of opportunity **after** program contact: $PAI_2 = 10$ Learned of opportunity **before** program contact, but made decision after contact: $PAI_2 = 10$ Learned of opportunity **before** program contact and made decision before contact: $PAI_2 = 0$ Learned of opportunity **before** program contact, but made the decision for part of the project before contact: $PAI_2 = 5$

Score	Description	Calculation
Program attribution index 3 (PAI_3)	• PAI_3 reflects program influence on project scope and timing. • The score is calculated by subtracting from 10 the respondent's ranked likelihood that they would have installed a project of the same scope in the same time frame even if the program had not been available. • S is the respondent's ranked likelihood that they would have installed a project *of the same scope* even if the program had not been available. • T is the respondent's ranked likelihood that they would have installed a project *in the same time frame* even if the program had not been available. • PAI_3 is calculated with the greater of S or T. • This approach is the same of OBF and non-OBF programs.	$PAI_3 = 10 - Max(S, T)$
Customer-level net-to-gross ratio (NTGR)	• The NTGR is calculated as the average of the three program attribution index scores.	$NTGR_{cust} = \frac{Average(PAI_1, PAI_2, PAI_3)}{10}$

Detailed NTG results

In this section we provide additional details about the program attribution indexes and how it impacted the NTGR. We first provide an overview of the influence ratings by program and program year and follow this with additional details on each influence rating score.

Table F-2 provides the program attribution index ratings by program and program year.

Table F-2. PAI scores

PA	Program	PY	PAI_1	PAI_2	PAI_3
BayREN	**Business FLEXmarket**	**2023**	9.7	10.0	7.3
		2024	9.0	10.0	7.5
PCE	**FLEXmarket Program**	**2023**	5.0	10.0	0.0
		2024	5.3	10.0	9.0
MCE	**Commercial Efficiency Market**	**2022**	10.0	10.0	10.0
		2023	6.3	10.0	6.0
		2024	8.7	10.0	6.2
SDGE	**CORE Market Access Program**	**2023**	5.0	10.0	5.0
PG&E	**Institutional Partnerships**	**2023**	7.3	5.1	6.1
		2024	9.5	10.0	9.7
	NetOne Commercial Efficiency	**2024**	6.0	9.3	6.1
PG&E	**OBF**	**2022**	7.2	9.3	6.3
		2023	4.8	10.0	6.8
		2024	5.4	9.7	8.6

Program attribution index score 1 (PAI_1) individual influence ratings

PAI_1 captures the importance of program and non-program influence. Interviewers asked respondents to rate how important various potential influences were on their decision to implement their project when they did. Table F-3 lists the four influences we asked respondents to rate, with each categorized as either a program or a non-program influence.

Table F-3. PAI_1 influences

Influence Type	Influence Factor	Description
Program	**Financial incentives**	Financing, rebates, or performance payments
Program	**Technical assistance**	Program-provided feasibility study or facility or system energy audit
Non-program	**Company policies or mandates**	Maintenance/replacement schedules or regulatory requirements; note that this excludes corporate sustainability goals
Non-program	**Other payback/return on the project**	Excluding program incentives, such as utility bill savings

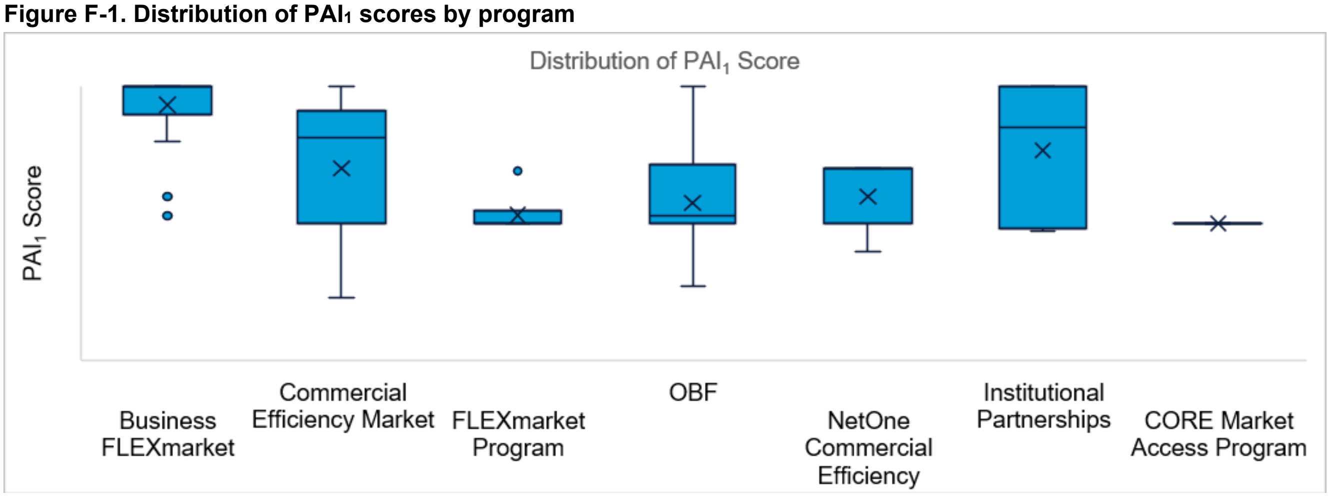
Respondents provided a rating on a 0-to-10 scale, where 0 means "Not at all important" and 10 means "Extremely important." The average ratings for each of the influence factors is presented in Table F-4 by PA and program.

Table F-4. Influence ratings for program and non-program influences

PA	Program	Financial incentives	Technical assistance	Company policies	Other payback/ return on the project
BayREN	Business FLEXmarket	9.4	9.3	2.9	9.3
PCE	FLEXmarket Program	9.5	1.5	3.8	8.5
MCE	Commercial Efficiency Market	9.3	5.4	4.9	7.1
SDGE	CORE Market Access Program	10.0	8.0	10.0	10.0
PG&E	Institutional Partnerships	9.4	9.7	5.3	9.6
PG&E	NetOne Commercial Efficiency	9.2	7.8	9.2	9.0
PG&E	OBF	8.9	7.7	8.2	8.0

As shown in Figure F-1, PAI_1 scores ranged from 5.0 to 9.1. The scores around 5 were mainly due to respondents giving a roughly equal number of points to program and non-program influences. Programs with a higher PAI_1 tended to have respondents give more points to program influences compared to non-program influences, and give higher ratings to those program influence factors.

Figure F-1. Distribution of PAI_1 scores by program



Program attribution index score 2 (PAI_2) individual influence ratings

PAI_2 reflects the prior plans of a participant, namely if they had plans in place and budget set aside for the project prior to interacting with the program. If the respondent reported that they made the decision to complete the project **after** interacting with the program, PAI_2 was set to 10. If the respondent reported that they made the decision to complete the project **before** interacting with the program, PAI_2 was set to 0. If they gave a mixed response, for example if there were prior plans for some measures but not others, PAI_2 was set to 5.

Table F-5 shows the count of respondents indicating if the project decision was made before or after interacting with the program. The majority of respondents (95%) indicated the decision was made after. PG&E was the only PA that had respondents indicate the decision was made before discussions about the program or incentive.

Table F-5. Respondent counts by project decision timing relative to program interaction

PA	Program	Before	After	Mixed
BayREN	**Business FLEXmarket**	0	31	0
MCE	**Commercial Efficiency Market**	0	20	0
PCE	**FLEXmarket Program**	0	6	0
PG&E	**OBF**	2	77	1
	NetOne Commercial Efficiency	1	14	0
	Institutional Partnerships	3	5	0
SDGE	**CORE Market Access Program**	0	2	0

Program attribution index score 3 (PAI_3) individual influence ratings

PAI_3 captures respondents' estimates of the program's influence on project scope and timing. Interviewers asked respondents to rate the likelihood their project would have had the same scope without the program, using a scale of 0 to 10, where 0 is "not at all likely" and 10 is "extremely likely." Using the same scale, we also asked respondents to rate the likelihood that, even without the program, they would have implemented the project at the same time.

Figure F-2 shows most respondents indicated the likelihood of completing the same project without the program was very low.

Figure F-2. **Likelihood of completing the same project**

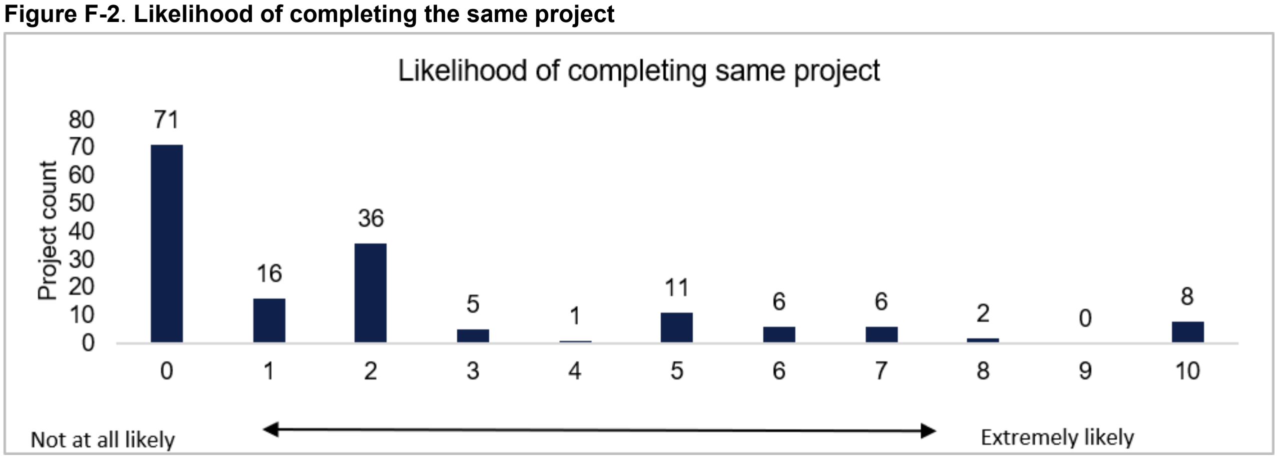


Figure F-3 shows the likelihood of completing the same project at the same time was more varied, with most respondents giving a score of 5.

Figure F-3. Likelihood of completing the project at the same time

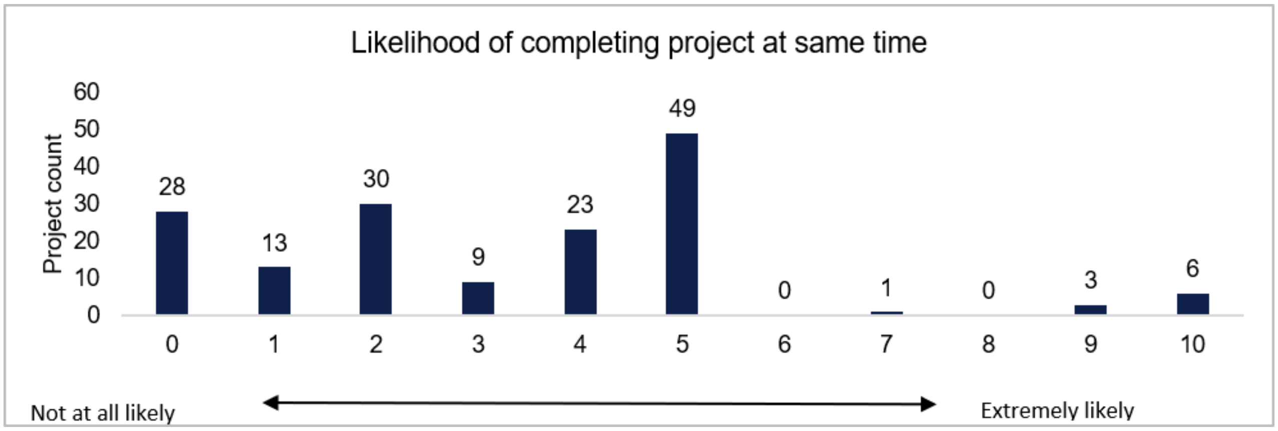
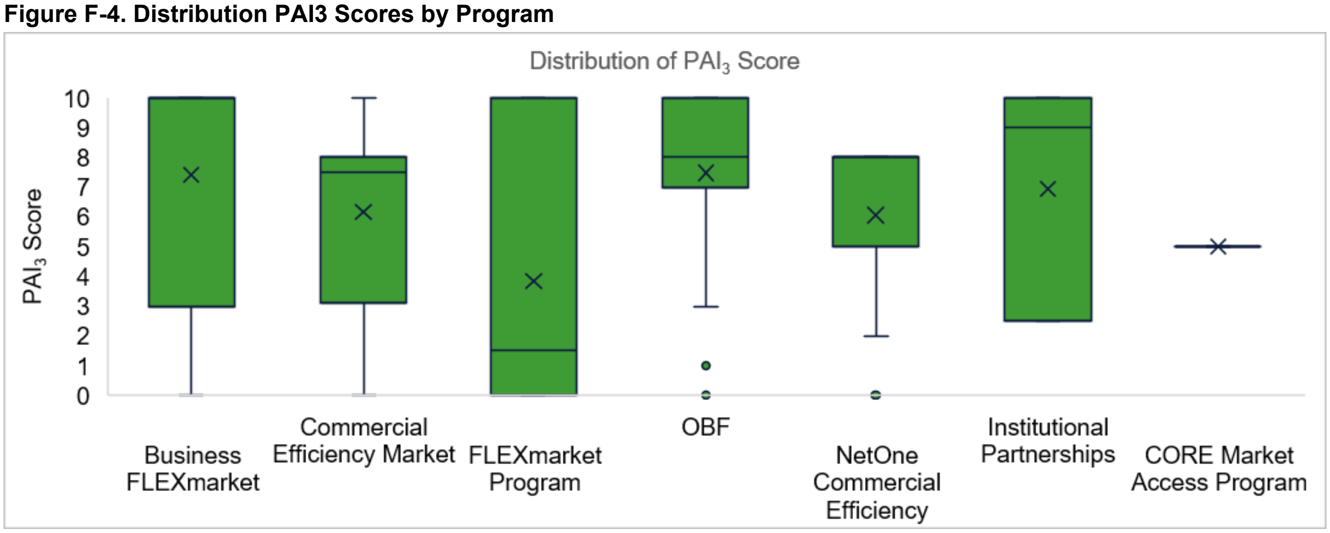


Figure F-4 shows the distribution of PAI_3 scores by program. The wide distribution of scores within each program is reflective of Figure F-2 and Figure F-3, which also show variations in scope and timing.

Figure F-4. Distribution PAI3 Scores by Program



APPENDIX G. PARTICIPANT SITE ATTRIBUTES

We developed a general business site profile of program participants to capture the history and scale of their site operations. We collected information on the year participants established operations at the site, the number of employees at the site, the square footage of the site, and the location type compared with other facilities they operate, and whether they own or lease their buildings. Table G-1 presents a summary of the findings by program and by each characteristic.

Site tenure. Overall, most participating sites were relatively established, having operated for several years prior to program enrollment. Across all programs, the majority of participants established their businesses at the site between 2000 and 2019. The second largest group began operations between 1980 and 1999. No more than 10% of a program's participants established operations in 2020 or later, indicating that newly established sites were less likely to participate.

Workforce scale. Across programs, most sites employed 25 or fewer individuals, reflecting a relatively small scale of site workforce. This is especially pronounced for BayREN Commercial FlexMarket (95%, n=229), which targets HTR small businesses.

Square footage. Similarly, sites under 10,000 sq ft represent the majority of participants in most programs, particularly for BayREN Commercial FlexMarket (98%, n=227). Mid-sized facilities between 10,000 and 99,999 sq ft show mixed engagement, with participation varying considerably by program. Large facilities over 100,000 sq ft are uncommon overall, enrolling only occasionally.

Business location type. DNV classified each participating site based on its role within the participant's broader organization: single locations, company headquarters, branches, or subsidiaries. Single locations and company branches comprised the majority of program participants. Company headquarters and subsidiaries rarely enrolled in the programs.

Building ownership. The majority of participants owned the buildings at their sites for PG&E OBF (70%, n=322) and PCE Commercial Efficiency Market (83%, n=6). In the other programs, participants were more evenly divided between owning and leasing their facilities, suggesting that building ownership did not strongly influence program participation overall.

Table G-1. Business site profile of pop NMEC program participants, PY2022 – PY2024

Attribute	Group	BayREN Business FLEXMarket (n = 230)	MCE Commercial Efficiency Market (n = 36)	PCE FLEXmarket Program (n = 11)	SDGE Core Market Access (n=1)	PG&E NetOne (n=39)	PG&E IP (n=10)	PG&E OBF (n=617)
Year established at site	Before 1980	0%	0%	0%	0%	0%	60%	7%
	1980 to 1999	32%	17%	9%	0%	15%	0%	30%
	2000 to 2019	58%	81%	91%	100%	79%	40%	57%
	2020 or later	10%	3%	0%	0%	5%	0%	5%
Number of site employees	25 or less	95%	50%	55%	0%	41%	60%	63%
	26 to 100	5%	39%	18%	100%	44%	40%	24%
	101 to 500	0%	11%	9%	0%	15%	0%	10%
	501 or more	0%	0%	18%	0%	0%	0%	3%
Square footage	Less than 10,000	98%	60%	55%	100%	58%	40%	87%
	10,000 to 99,999	2%	34%	18%	0%	42%	10%	10%
	100,000 or more	0%	6%	27%	0%	0%	50%	4%
	Single Location	88%	25%	0%	0%	8%	70%	87%

Attribute	Group	BayREN Business FLEXMarket (n = 230)	MCE Commercial Efficiency Market (n = 36)	PCE FLEXmarket Program (n = 11)	SDGE Core Market Access (n=1)	PG&E NetOne (n=39)	PG&E IP (n=10)	PG&E OBF (n=617)
Business location type	Branch	12%	72%	91%	100%	92%	30%	10%
	Headquarter	0%	3%	0%	0%	0%	0%	2%
	Subsidiary	0%	0%	9%	0%	0%	0%	1%
Own or lease	Own	54%	44%	83%	-	50%	57%	70%
	Lease	46%	56%	17%	-	50%	43%	30%

The participant data counts listed at the top of the table largely apply to all elements included in this table except the “square footage” and “own or lease” attribute, which includes information for only about half of the size.

APPENDIX H. SURVEY INSTRUMENTS AND INTERVIEW GUIDES

PY2022 – PY2024 POPULATION NMEC NON-RESIDENTIAL PROGRAM SURVEY INSTRUMENT

Survey Overview

Objective: Survey participants to inform Net-to-Gross (NTG) and process-related research objectives.

NTG Methodology Overview: The NTG evaluation for this population NMEC program follows the same methodology as for the Group D Custom programs, including site-level NMEC. Changes have been made to better align with pop NMEC program delivery. While the Custom programs evaluation addressed questions at an individual measure level, for pop NMEC projects, the questions were adjusted to refer to each project as a whole. Still, the guide contains questions asking if the scope of individual measures in the project would have been different without the NMEC program. The results from this interview guide were used to calculate NTG using the updated methodology developed and approved in conjunction with this guide.

Survey length: 25-30 minutes

Timing: November 2025 – February 2026

Method of data collection: Phone survey

Table 1. Research Objectives Mapped to Questions in this Instrument

Research Objectives	Survey Questions Address the Objectives
Introduction: Identify the appropriate contact to answer survey questions, introduce the purpose of the study, and confirm familiarity with the project/program	Q1 – Q6
Timing Questions: Explore whether program support influenced project timing by identifying when the project was introduced, when incentive discussions began, and how the absence of support might have changed the timeline.	Q7 – Q14
Scope Questions: Assess whether the project exceeded code requirements, the likelihood of implementing the same project without program support, and what alternative actions would have been taken.	Q15 – Q19
Influence Questions: Assess the importance of program incentives, technical assistance, and other factors (such as organization policies and financial considerations) in the decision to implement the project.	Q20 – Q23
Process Evaluation Battery: Gather feedback on program strengths, non-energy benefits, areas for improvement, overall satisfaction.	Q24 – Q27
Firmographics: Collect data on characteristics or organization	Q28 – Q34

Table 2. Overview of Data Collection Approach

Data Collection	Description
Population Description	To be updated following completion of sample design.
Population Size/Sample Frame	To be updated following completion of sample design.
Type of Sampling	Stratified Random
Target Sample - Survey Completions	To be updated following completion of sample design.
Instrument Type	Phone Survey
Anticipated Survey Length	20-25 minutes
Description of Contact Sought	Program Participants

Survey

Programming instructions are CAPITALIZED.

The evaluation team will input the following data from the PG&E OBF database(s) to reference the information during the interview. Throughout this instrument, pipe in fields are denoted by % within brackets and capital letters: [%EXAMPLE].

Table 3. Database Information Piped into the Survey Instrument

Variable Name	Variable Description and Values
ADDRESS	Address of the organization participating in the program
PA	Name of the program administrator
PA CONTACT NAME	Name of the PA program lead contact
PROGRAM	Name of the program
AGGREGATOR	Aggregator or contractor that enrolled and provided the upgrade to the customer
YEAR	Year of the project
MEASURE(s)	List of installed measures

The table below captures essential details about interviews with participating organizations, including contact information and survey completion metrics. It also provides call tracking records to document communication history.

Table 4. Interview Information

Variable Name	Variable Description and Values
Organization	Name of the organization participating in the program
Contact Name	Name of contact at the organization participating in the program
Contact Phone	Phone number of the contact at the organization participating in the program
Contact Email	Email of the contact at the organization participating in the program
MEASURE_N	Number of measures
Completion Date	Address of the organization participating in OBF

Variable Name	Variable Description and Values
Survey Length (MIN)	Length of survey in minutes

Introduction

[NOTE: THE QUESTIONS IN THIS INTERVIEW GUIDE WILL NOT NECESSARILY BE READ VERBATIM BUT MAY BE MODIFIED TO SUIT THE INTERVIEW]

1. Hi, my name is [CALLER NAME] OF DNV. We are calling on behalf of the California Public Utilities Commission (CPUC) and [%PA]. According to our records, the facility at [%ADDRESS] made energy efficiency upgrades which received financial incentives through [%PA]'s [%PROGRAM] in [%YEAR] [IF PA = MCE or BAYREN or PCE or SDGE or SCP] administered by [%AGGREGATOR]]. You are listed as the lead contact for this project. Are you able to answer questions about the decision-making process and motivations for doing this project? Any information that you provide will remain strictly confidential. We will not identify or attribute any of your comments or organization information. To thank you for your participation, upon successful completion, you may choose to be entered into a drawing to receive one of five available $100 Amazon e-gift cards.
    a. [IF YES, SKIP TO Q2]
    b. [IF YES, BUT THEY CAN'T DO THE INTERVIEW AT THAT TIME, SCHEDULE ANOTHER TIME]
    c. [IF NO, OBTAIN ALTERNATE CONTACT INFO AND SCHEDULE INTERVIEW WITH NEW PERSON]
    d. [IF REFUSED] Just to be clear, by receiving financial incentives through this program, your organization agreed to participate in this follow-up study on your experiences with this program. Considering that information, can we go ahead with this interview?
        i. [IF YES. GO TO Q2]
        ii. [IF NO] Okay, you may be contacted directly by the CPUC to complete this process.
    e. [IF THEY ASK HOW LONG THE INTERVIEW WILL TAKE] I will do my best to keep the survey under 30 minutes.
    f. [IF THEY ASK TO VERIFY WITH THE PA THAT THE RESEARCH IS LEGITIMATE, GIVE THEM [PA CONTACT NAME]:

[CONFIRMATION OF CORRECT RESPONDENT, PROJECT BACKGROUND]

2. [A1] According to our records, your organization implemented the following energy efficiency improvements at [%ADDRESS] in [%YEAR]. Is this correct? [READ THE CHOICES FROM THE TABLE BELOW AND MAKE NEEDED CORRECTIONS]

Project Information	Information from Tracking Data (pre-entered)	Corrected information (if relevant)
Measure	[%MEASURE(s)]	
Address	[%ADDRESS]	
Install Date	[%YEAR]	

3. [C3] What was your specific role and involvement in the project?
4. [C3a] Which other contacts were involved in moving this project forward?
    a. [OBTAIN NAME[S] AND CONTACT INFO]
5. [A1b] [IF MEASURE_N > 1] Our records show that your organization implemented or installed more than one energy efficiency measure as part of a project through the [%PA]'s [%PROGRAM] Program [[IF PA = MCE or BAYREN or PCE or SDGE or SCP] administered by [%AGGREGATOR]. Was the decision-making process for the implementation and installation of these measures made at the whole project level? Or was there a separate decision-making process for each measure?
    a. [Decision-making done at project level]
    b. [separate decision-making process for each measure]
    c. [Don’t know]
    d. [Refused]
6. Which of the following support did you receive from the program? [SINGLE RESPONSE]
    a. Cash or cash-equivalent incentive
    b. Equipment discount
    c. Rebate
    d. Financing
    e. Promise of energy and utility bill savings
    f. Other reward: Please specify

Timing Questions

For the rest of this interview, we will refer to all the upgrades we just discussed, collectively, as “the project.”

7. [A5b_2] Who first brought this project to your organization’s consideration? [DON’T READ THE LIST BELOW; SELECT THE CLOSEST RESPONSE OR NOTE THE ANSWER IN THE "OTHER" FIELD]
    a. Program staff
    b. Utility staff
    c. Non-program vendor

d. Organization staff

e. Other

8. [A5b] About when was this project first put forward for your organization's consideration?

    a. [RECORD YEAR AT MINIMUM, QUARTER/MONTH GRANULARITY IF POSSIBLE] ___

    b. [Don't know]

    c. [Refused]

9. [A5d] About when did your organization first begin discussions with [PA] regarding funding/incentives and/or possible technical assistance for this project?

    a. [RECORD YEAR AT MINIMUM, QUARTER/MONTH GRANULARITY IF POSSIBLE] ___

    b. [Don't know]

    c. [Refused]

10. [N2] Was the decision to do this project made before or after you began discussions with the project developer or installer regarding the availability of incentives or technical assistance for this project? [IF NEEDED: technical assistance may include identifying the project, providing project financials, technical support during contractor selection, installation and/or project commissioning, or verification of energy savings]

    a. Before

    b. After

    c. Don't know

    d. Refused

11. **CONSISTENCY CHECK** [DISPLAY IF (Q6 = a OR b) AND (Q9=a)]: IF RESPONDENT INDICATED PROJECT WAS FIRST PUT FORWARD BY THE PROGRAM BUT THAT THEY MADE THE DECISION TO DO THE PROJECT BEFORE LEARNING ABOUT INCENTIVES OR TECHNICAL ASSISTANCE FOR THIS PROJECT] Earlier you indicated that the project was first put forward to your organization by [A5b_2 response]. However, in your response to the previous question, you indicated the decision to go ahead with the project was made before learning about the incentives or technical assistance from the project developer or installer. To clarify, technical assistance from the project developer or installer can include presenting you with a project that your organization previously had not considered. Should I revise your answer about the timing of your decision to go ahead with the project before or after learning about incentives or technical assistance from the project developer or installer?

12. [N5b] If the program elements had not been available, including incentives or other types of program assistance, what is the likelihood that you would have conducted the project at the same time as you did?

    e. Very likely

    f. Somewhat likely

    g. neither likely nor unlikely

    h. Somewhat unlikely

i. Very unlikely

13. Without the assistance received from the [PROGRAM] (including any incentive funds, program information, energy audits, technical assistance, and any other support) would your organization have completed the whole project:

    j. About the same time or earlier than you did

    k. About a year later than you did

    l. About two years later than you did

    m. 3 or more years later than you did

    n. Or never

    o. Don't know

14. [IF A5b_2 = a or b] Were you aware of the specific opportunities included in this project before [A5b_2 response] suggested the project?

    p. [IF YES] Which parts of the project did you already know were potential opportunities?

Scope Questions

15. [N3a] Did the project increase energy efficiency beyond what was required by code?

    a. Yes

        i. [IF YES] Was any of the increase in efficiency above code due to the program?

    b. No [SKIP QUESTIONS 22 and 23]

    c. Don't know [SKIP QUESTIONS 22 and 23]

    d. Not applicable [SKIP QUESTIONS 22 and 23]

16. [N5] Using a likelihood scale from 0 to 10, where 0 is "Not at all likely" and 10 is "Extremely likely", if elements of this PROGRAM had not been available, including incentives or other types of program assistance, what is the likelihood that you would have implemented the exact same project that you did? Record 0 to 10 score (_______)

Actions Without the Program

17. [Revised N6] I would like you to think about what actions you would have taken regarding the upgrades included in this project if the program had not been available. Which of the following alternatives would you have been MOST likely to choose:

    a. Install the same equipment

    b. Continue to operate the existing equipment as-is [IF NEEDED: done nothing]

    c. Install fewer units

    d. Repair existing equipment

    e. Install equipment with a different efficiency level

    f. Something else (specify _____________)

18. On a 0-to-10 scale of likelihood, how likely it is that you would have been to [DISPLAY RESPONSE FROM Q17] if the program had not been available?
    a. Record 0 to 10 score (_______)
19. [A3a] Has your organization done similar types of projects at this or other California locations in the past?
    a. [IF YES] Which programs incentivized the project?
    b. [A3a] [IF YES] What equipment were included in those projects?
    c. [A3aa] [IF YES] Did your experience participating in [PA] [PROGRAM] have any impact on your installation decision for these similar projects?
        i. [IF YES] What impacts did they have?

Influence Questions

Now I'm going to ask you a few questions about possible factors that may have influenced your decision to go ahead with this project.

I'm going to ask you to rate the importance of the program as well as other factors that might have influenced your decision to conduct this project. Think of the degree of importance as being shown on a scale with equally spaced units from 0 to 10, where 0 means not at all important and 10 means extremely important, so that an importance rating of 8 shows twice as much influence as a rating of 4.

20. [N3] Now, using this 0-to-10 rating scale, where 0 means "Not at all important" and 10 means "Extremely important," please rate the importance of each of the following in your decision to implement this project at the time you did. [RANDOMIZE CHOICE]
21. [DISPLAY IF Q14 = WENT ABOVE CODE, N3a] Now I'd like to ask you the same question but thinking only about the portions of your project UP TO CODE requirements, using this 0-to-10 rating scale, where 0 means "Not at all important" and 10 means "Extremely important," please rate the importance of each of the following in your decision to implement the portion of the project to bring measures UP TO CODE at the time you did. [RANDOMIZE CHOICE]
22. [DISPLAY IF Q14 = WENT ABOVE CODE, N3a] Lastly, I'd like to ask you the same question but thinking only about the portions of your project that WENT ABOVE CODE requirements, using this 0-to-10 rating scale, where 0 means "Not at all important" and 10 means "Extremely important," please rate the importance of each of the following in your decision to implement the portions of the project that WENT ABOVE CODE at the time you did. [RANDOMIZE CHOICE]

Factor	[DISPLAY FOR Q19] Scoring (Overall)	[DISPLAY FOR Q20] Scoring (To Code)	[DISPLAY FOR Q21] Scoring (Above Code)
b. [N3b] The availability of the [PA] [PROGRAM] incentives, financing, or performance payments	Record 0 to 10 score (___). -97. Don't know -98. Not applicable -99. Refused	Record 0 to 10 score (___). -97. Don't know -98. Not applicable -99. Refused	Record 0 to 10 score (___). -97. Don't know -98. Not applicable -99. Refused

Factor	[DISPLAY FOR Q19] Scoring (Overall)	[DISPLAY FOR Q20] Scoring (To Code)	[DISPLAY FOR Q21] Scoring (Above Code)
c. [N3c] [IF RELEVANT] Information provided by the program such as facility or system energy audit or technical assistance	Record 0 to 10 score (___). -97. Don't know -98. Not applicable -99. Refused	Record 0 to 10 score (___). -97. Don't know -98. Not applicable -99. Refused	Record 0 to 10 score (___). -97. Don't know -98. Not applicable -99. Refused
j. [N3j] Organizational practices such as internal policies or goals, maintenance or replacement schedules, and regulatory requirements	Record 0 to 10 score (___). -97. Don't know -98. Not applicable -99. Refused IF > 5 What specific organization policies or practices influenced your decision to do the project?	Record 0 to 10 score (___). -97. Don't know -98. Not applicable -99. Refused	Record 0 to 10 score (___). -97. Don't know -98. Not applicable -99. Refused
n. [N3n] Payback or returns on the project without financial assistance from the program	Record 0 to 10 score (___). -97. Don't know -98. Not applicable -99. Refused	Record 0 to 10 score (___). -97. Don't know -98. Not applicable -99. Refused	Record 0 to 10 score (___). -97. Don't know -98. Not applicable -99. Refused

Comparing Program and Non-Program Influences

Next, I would like you to rate the importance of the PROGRAM in your decision to do this project as opposed to other non-program factors that may have influenced your decision such as:

Non-Program Factor
[N3j] Standard practice at your organization such as complying with organization policies or goals, maintenance/replacement schedules, and regulatory requirements
[N3n] Payback or return on the PROJECT without factoring in program incentives

As another reminder, program-related factors are:

Program Factors
[N3b] The availability of the [PA] [PROGRAM] incentives

[N3c] Information provided by the program such as facility or system energy audit or technical assistance

23. [N41] If you were given 10 points to award in total, how many points would give to the importance of the program and how many points would you give to these other factors?

    a. [N41] How many of the 10 points would you give to the importance of the PROGRAM in your decision? # ________rating of the importance of PROGRAM

    b. [N42] And how many points would you give to these other factors? # ________rating of the importance of all Other Factors

Process Evaluation Battery

24. [PP1] What do you believe the PROGRAM'S primary strengths are?

25. [PP2] What concerns do you have about the PROGRAM, if any?

    a. [IF THERE ARE CONCERNS ASK] What do you view as the primary features that need to be improved?

26. [PP4] On a scale of 0 - 10, where 0 is completely dissatisfied and 10 is completely satisfied, how would you rate your OVERALL satisfaction with the PROGRAM?

    a. [IF SATISFACTION RATING LESS THAN 7] Why do you say that?

27. Aside from saving energy, which of the following benefits did your organization realize from completing this project? [MULTIPLE RESPONSES]

    a. Regulatory or corporate compliance

    b. Improved user comfort

    c. Greater competitiveness (e.g., brighter retail lighting)

    d. Increased productivity

    e. Lower maintenance costs

    f. None [MAKE EXCLUSIVE]

    g. Other, please specify: [OPEN-END]

    h. Don't know [MAKE ESCLUSIVE]

Firmographics

And finally, I have a few questions about the characteristics of your business.

28. [CC12A, CC12B] In about what year was this [business/organization/entity] established at this location? Would you say it was

    a. After 2020

    b. In the 2010s

    c. In the 2000s

d. In the 1990s

e. In the 1980s

f. In the 1970s

g. In the 1960s or

h. Before 1960

i. Don't know

j. Refused

29. [CO] About what percentage of your operating costs does energy account for? PAUSE....Would you say...

    a. Less than 1 percent

    b. 1 to 2 percent

    c. 3 to 5 percent

    d. 6 to 10 percent

    e. 11 to 15 percent

    f. 16 to 20 percent

    g. 21 to 50 percent

    h. Over 51 percent

    i. Don't know

    j. Refused

30. [CCC1, CCC3] Do you know how many square feet the facility covers? PAUSE.... Would you say that the area the facility covers is…

    a. 1,500 sq. feet or less

    b. 1,500 to 5,000 sq. ft

    c. 5,001 to 10,000 sq. ft

    d. 10,001 to 25,000 sq. ft

    e. 25,001 to 50,000 sq. ft

    f. 50,001 to 75,000 sq. ft

    g. 75,001 to 100,000 sq. ft

    h. Over 100,000 sq. ft

    i. Don't know

    j. Refused

31. [C1] What is the main activity conducted at this facility?

    a. [RECORD RESPONSE]: ___

32. [C3] Approximately how many people are currently working at the facility where the project was conducted, including both full and part time?

    a. Ten or less
    b. Between 11 and 25
    c. 26 to 50
    d. 51 to 75
    e. 76 to 100
    f. 101 to 250
    g. 251 to 500
    h. 501 to 1000
    i. 1001 to 2500
    j. 2501 to 5000 or
    k. 5000 or more
    l. Don't know
    m. Refused

33. [C4] Does your [business/organization/entity] own, lease or manage this facility?

    a. Own
    b. Lease/Rent
    c. Manage
    d. Don't know
    e. Refused

34. [C5] How many locations does your organization have? Is it....

    a. 1
    b. 2 to 4
    c. 5 to 10
    d. 11 to 25
    e. Over 25

f. Don’t know

g. Refused

Closing

35. Do you have any reflections you wish to share about what was asked in this survey? [SINGLE RESPONSE]

    a. Yes [RECORD RESPONSE]: ___

    b. No

[DISPLAY TEXT ON TERMINATION SCREEN: You have completed the survey and your responses have been submitted. Your contribution to this survey helps CPUC and [%PA] to evaluate and improve program offerings. Thank you for your participation and time.]

PY2022 – PY2024 POPULATION NMEC OBF-AP SURVEY INSTRUMENT

Survey Overview

Objective: Survey PG&E OBF-AP participants to inform Net-to-Gross and process-related evaluation objectives.

Survey length: 20-25 minutes

Timing: November 2025 – February 2026

Method of data collection: Phone survey

Table 1. Research Objectives Mapped to Questions in This Instrument

Research Objectives	Survey Questions Address the Objectives
Identify the appropriate contact to answer survey questions, introduce the purpose of the study, and confirm familiarity of the program/OBF loan	S1 – S4
Determine whether other sources of payment or financing were used for projects beyond the OBF loan, and explore whether participants sought or would have sought other non-program (non-OBF) financing, as well as the likelihood of using such alternative	Q1 – Q7
Explore sources of program awareness and understand motivation for completing a project	Q8 – Q11
Understand the influence of general financing, collect data to inform NTG calculations	Q12 – Q16
Understand the influence of OBF, collect data to inform NTG calculations	Q17 – Q22

Research Objectives	Survey Questions Address the Objectives
Determine general program and non-program influence on OBF program participation to inform NTG based on the CA Non-residential NTG protocol	Q23 – Q30
Understand satisfaction with the program and its various elements and source of non-energy benefits	Q31 – Q33
Collect data on characteristics or organization	Q34 – Q41

Table 2. Overview of Data Collection Approach

Data Collection	Description
Population Description	To be updated following completion of sample design.
Population Size/Sample Frame	To be updated following completion of sample design.
Type of Sampling	Stratified Random
Target Sample - Survey Completions	To be updated following completion of sample design.
Instrument Type	Phone Survey
Anticipated Survey Length	20-25 minutes
Description of Contact Sought	OBF-AP Program Participants

Programmer Information

Programming instructions are CAPITALIZED.

The evaluation team will input the following data from the PG&E OBF database(s) to reference the information during the interview. Throughout this instrument, pipe in fields are denoted by brackets and capital letters: [EXAMPLE].

Table 3. Database Information Piped into the Survey Instrument

Variable Name	Variable Description and Values
ADDRESS	Address of the organization participating in OBF
YEAR	Year of OBF project
MEASURE_1	Lighting
MEASURE_2	Non-Lighting

Interview Information

This section captures essential details about interviews with participating organizations, including contact information and survey completion metrics. It also provides call tracking records to document communication history.

Table 4. Interview Information

Variable Name	Variable Description and Values
Organization	Name of the organization participating in the program
Contact Name	Name of contact at the organization participating in OBF
Contact Phone	Phone number of the contact at the organization participating in the program
Contact Email	Email of the contact at the organization participating in the program
Completion Date	Completion date of the survey
Survey Length (MIN)	Length of survey in minutes

Survey

Phone Introduction Text

Hello, my name is [CALLER NAME] from DNV calling on behalf of the California Public Utilities Commission (CPUC). According to our records, your organization participated in the Pacific Gas & Electric (PG&E) on-bill financing program. We are conducting some research on this program and wanted to ask you a few questions about it. The information gathered will be used solely for research purposes and your individual responses will be kept confidential.

[IF THEY ASK HOW LONG THE SURVEY WILL TAKE: 20-25 minutes]

[IF NEEDED: This call is regarding the PG&E On-Bill Financing program. Through this program, PG&E provided financing for energy-efficient upgrades to your facility. Our records show that your organization completed this project in [YEAR]. This call is for research purposes and is not a marketing or sales call].

Screening

S1. According to our records, your organization received on-bill financing from Pacific Gas & Electric (PG&E) for an energy efficiency project at [ADDRESS]. Is this correct? [SINGLE RESPONSE]

Response ID	Response	Action
01	Yes	S2
02	No	[THANK AND TERMINATE]

S2. Are you the best person to speak about your organization's decision-making process related to developing and implementing this project, including the decision to apply for an on-bill financing? [SINGLE RESPONSE]

Response ID	Response	Action
01	Yes	S3
02	No, but correct contact can come to the phone	[ASK TO SPEAK TO CORRECT CONTACT AND REPEAT QUESTION]
03	No, but correct contact is not available	[GET NAME AND PHONE NUMBER OF CORRECT CONTACT TO SCHEDULE CALL BACK]
-97	Don't know	[ASK TO SPEAK TO SOMEONE WHO KNOWS AND REPEAT QUESTION]
-98	Refused	[THANK AND TERMINATE]

S3. Our records indicate your organization used on-bill financing to install the following equipment: [DISPLAY IF MEASURE_1 IS NOT BLANK: MEASURE_1
[DISPLAY IF MEASURE_2 IS NOT BLANK: MEASURE_2].
Does that sound right? [SINGLE RESPONSE]

Response ID	Response	Action
01	Yes	S4
02	No, please explain: [INSERT OPEN-ENDED RESPONSE]	S4
-97	Don't know	S4
-98	Refused	[THANK AND TERMINATE]

S4. What specific equipment type did your organization install at [ADDRESS] using PG&E's on-bill financing? [MULTIPLE RESPONSE]

Response ID	Response	Action
01	Lighting	RESPONSE ID 01
02	HVAC	RESPONSE ID 02
03	Refrigeration	RESPONSE ID 03
04	Insulation	RESPONSE ID 04
05	Water Heating	RESPONSE ID 05
06	Pool Pump Smart Controller	RESPONSE ID 06
07	Food service	RESPONSE ID 07
-96	Other, please specify: [OPEN-END]	RESPONSE ID -96
-97	Don't know	[THANK AND TERMINATE]
-98	Refused	[THANK AND TERMINATE]

Payment Method

Q1. Which of the following payment methods would your organization have been most likely to use, if PG&E financing was not available? Please select all that apply. [ALWAYS READ ALL OPTIONS EXCEPT -97 or -98. ALLOW RESPONDENT TO CITE MULTIPLE RESPONSES]

Response ID	Response	Action
01	Internal funds / organization's budget	RESPONSE ID 01
02	Credit card	RESPONSE ID 02
03	Grant	RESPONSE ID 03
04	Financing (loan) other than the PG&E on-bill financing	RESPONSE ID 04
05	We would not have used any other payment methods besides on-bill financing	RESPONSE ID 05
-96	Other, please specify: [OPEN-END]	RESPONSE ID -96
-97	Don't know [MAKE ESCLUSIVE]	Q2
-98	Refused [MAKE EXCLUSIVE]	Q2

Q2. Besides the on-bill financing from Pacific Gas & Electric (PG&E), did your organization use any of the following to pay for the project? Please select all that apply. [MULTIPLE RESPONSES]

Response ID	Response	Action
01	Internal funds / organization's budget	RESPONSE ID 01
02	Credit card	RESPONSE ID 02
03	Grant	RESPONSE ID 03
04	Financing (loan) other than the PG&E on-bill financing	RESPONSE ID 04
05	Used nothing else besides on-bill financing	RESPONSE ID 05
-96	Other, please specify: [OPEN-END]	RESPONSE ID -96
-97	Don't know [MAKE EXCLUSIVE]	Q3
-98	Refused [MAKE EXCLUSIVE]	Q3

Q3. [DISPLAY IF Q2 IS NOT -97 OR -98] About what percentage of the total cost of the project was covered by each of the following. [MULTIPLE PERCENTAGE ENTRY, EACH RESPONSE 0-100, TOTAL SUM = 100]

Response ID	Response	Action

01	[DISPLAY IF Q2=01] Internal funds / organization's budget	RESPONSE ID 01
02	[DISPLAY IF Q2=02] Credit card	RESPONSE ID 02
03	[DISPLAY IF Q2=03] Grant	RESPONSE ID 03
04	[DISPLAY IF Q2=04] Financing (loan) other than PG&E on-bill financing	RESPONSE ID 04
05	PG&E on-bill financing	RESPONSE ID 05
-96	[DISPLAY IF Q2=-96] Other	RESPONSE ID -96
-97	Don't know [MAKE EXCLUSIVE]	Q4
-98	Refused [MAKE EXCLUSIVE]	Q4

Q4. [DISPLAY IF Q2 = <> 04] Did your organization seek any other financing (loan) options for the project? [SINGLE RESPONSE]

Response ID	**Response**	**Action**
01	Yes	Q5
02	No	Q5
-97	Don't know	Q5
-98	Refused	Q5

Q5. [DISPLAY IF Q4 = 02] If PG&E on-bill financing had not been available, would your organization have sought other financing (loan) options for the project? [SINGLE RESPONSE]

Response ID	**Response**	**Action**
01	Yes	Q6
02	No	Q6
-97	Don't know	Q6
-98	Refused	Q6

Q6. [DISPLAY IF Q5 = 01] On a scale of 0 to 10, with 0 being "not at all likely", and 10 being "extremely likely", how likely would your organization have been to qualify for other financing (loan) besides the on-bill financing you received from PG&E? [SINGLE RESPONSE]

Response ID	**Response**	**Action**
01	0 – Not at all likely	Q7
02	1	Q7
03	2	Q7
04	3	Q7
05	4	Q7
06	5 – Neither likely nor unlikely	Q7
07	6	Q7
08	7	Q7
09	8	Q7
10	9	Q7
11	10 – Extremely likely	Q7
-97	Don't know	Q7
-98	Refused	Q7

[IF THE ANSWERS TO **ALL** OF Q1, Q3, Q4, Q5, AND Q6 ARE "DON'T KNOW" OR "REFUSED", ASK TO SPEAK TO SOMEONE ELSE WHO CAN PROVIDE THE INFORMATION OR REQUEST THEIR NAME AND CONTACT DETAILS TO SCHEDULE A CALL BACK. THEN THANK THE RESPONDENT AND TERMINATE THE SURVEY.]

Q7. [DISPLAY IF Q1 = 05 AND Q2 = 05 AND Q4 = 02 AND Q5 = 02 OR (Q5 = 01 AND Q6 < 8)] You have indicated that PG&E on-bill financing was the only project financing your organization was willing or able to pursue. Could you briefly describe your organization's financial decision-making process and why this was the only option?

Response ID	Response	Action
01	[INSERT OPEN-ENDED RESPONSE]	Q8
-97	Don't know	Q8
-98	Refused	Q8

Project Awareness

[DISPLAY TEXT: Next, we would like to understand how your organization found out about the on-bill financing and/or rebates available through PG&E.]

Q8. How did your organization first learn about the on-bill financing? [SINGLE RESPONSE, RANDOMIZE 1-4]

Response ID	Response	Action
01	From a friend or colleague	Q9
02	Contractor, installer, or project developer	Q9
03	PG&E account or energy efficiency program representative	Q9
04	Utility bill insert	Q9
05	PG&E account website	Q9
06	Your own research	Q9
-96	Other, please specify: [INSERT OPEN-ENDED RESPONSE]	Q9
-97	Don't know	Q9
-98	Refused	Q9

Q9. Had your organization identified the need for part or all the project you completed prior to learning about the on-bill financing from PG&E? [SINGLE RESPONSE]

Response ID	Response	Action
01	Yes	Q10
02	No	Q10
-97	Don't know	Q10
-98	Refused	Q10

[DISPLAY IF Q10 = 01]

Q10. Prior to learning about the on-bill financing from PG&E, was the cost for part or all this project included in your organization's approved capital budget? [SINGLE RESPONSE]

Response ID	Response	Action
01	Yes, the full project cost	Q11
02	Yes, part of the project cost	Q11
03	No	Q11
-97	Don't know	Q11
-98	Refused	Q11

Q11. Which of the following reasons **best** describes why the energy efficiency project you received financing for was completed. [SINGLE RESPONSE. PROGRAMMER: PLEASE RANDOMIZE THESE OPTIONS]

Response ID	Response	Action
01	Wanted my organization to be more environmentally friendly The upgrade was part of the organization's sustainability policy tied to a carbon reduction goal	Q12
02	The upgrade was to reduce the organization's energy usage	Q12
03	The upgrade was required to meet a building code or compliance standard	Q12
04	The equipment broke and needed immediate repair or replacement	Q12

05	The equipment was upgraded to support the organization's expansion	Q12
-96	Other, please specify: [OPEN-END]	Q12
-97	Don't know	Q12
-98	Refused	Q12

General Financing Influence

[INSTRUCTION: This next set of questions is designed to understand how financing may have influenced the respondent's decision to participate in the program and complete the project. We are exploring the role that financing in general played in their decision, given that no other financing options were available to them. In other words, the counterfactual or what they could use are internal funds / organization budget.]
[DISPLAY TEXT: For the next set of questions, please think about a hypothetical situation where you **did not have any financing** to pay for your energy efficiency project. In other words, all you could use were internal funds or your organization's budget, a credit card, or grants.]

Q12. To start, please briefly describe why your organization chose to use financing to pay for the project.

Response ID	Response	Action
01	Please specify: [INSERT OPEN-ENDED RESPONSE]	Q13
02	Nothing would have changed about the project without the on-bill financing.	Q13
-97	Don't know	Q13
-98	Refused	Q13

Q13. On a scale of 0 to 10, with 0 being "not at all likely", and 10 being "extremely likely", how likely is it that your organization would have undertaken this project **without any financing**? [SINGLE RESPONSE]

Response ID	Response	Action
01	0 – Not at all likely	Q14
02	1	Q14
03	2	Q14
04	3	Q14
05	4	Q14
06	5 – Neither likely nor unlikely	Q14
07	6	Q14
08	7	Q14
09	8	Q14
10	9	Q14
11	10 – Extremely likely	Q14
-97	Don't know	Q14
-98	Refused	Q14

[DISPLAY QUESTION FOR UP TO 2 MEASURES [MEASURE_1/2]

Q14. **Without any financing**, when would your organization have installed the [MEASURE_1/2] equipment? [SINGLE RESPONSE]

Response ID	Response	Action
01	At the same time or sooner (e.g., within 6 months)	Q15
02	7 to 12 months later	Q15
03	13 to 24 months later	Q15
04	25 to 36 months later	Q15
05	More than 36 months later	Q15
06	Would not have completed this project at all	Q15
-96	Other, please specify: [OPEN-ENDED RESPONSE]	Q15
-97	Don't know	Q15

-98	Refused	Q15

[DISPLAY QUESTION FOR UP TO 2 MEASURES [MEASURE_1/2]

Q15. **Without any financing**, what level of efficiency would your organization have installed for the [MEASURE_1/2] equipment? [SINGLE RESPONSE]

Response ID	Response	Action
01	High efficiency (like the one you purchased)	Q16
02	Mid-level efficiency between standard and high efficiency	Q16
03	Standard efficiency as required by the building code	Q16
04	Below code efficiency	Q16
-97	Don't know	Q16
-98	Refused	Q16

[DISPLAY QUESTION FOR UP TO 2 MEASURES [MEASURE_1/2]

Q16. **Without any financing**, what percent of the [MEASURE_1]/2] equipment your organization installed using PG&E on-bill financing would you have installed? [SINGLE RESPONSE]

Response ID	Response	Action
01	Enter percent: [INSERT NUMERIC OPEN-ENDED RESPONSE, ACCEPT 0-100]	Q17
-97	Don't know	Q17
-98	Refused	Q17

On-Bill Financing Influence

[INSTRUCTION: This next set of questions is designed to understand how PG&E's On-Bill Financing may have influenced the respondent's decision to participate in the program and complete the project. We're exploring the specific role the program's On-Bill Financing played in their decision, compared to other financing options available to them. In other words, the counterfactual or what they could use are other financing options and internal funds / organization budget.]
[DISPLAY TEXT: Next, we would like to understand why your organization chose to use PG&E **on-bill financing** specifically.]

Q17. To start, please briefly describe why your organization chose to use PG&E **on-bill financing** instead of, or in addition to, other financing options. [MULTIPLE RESPONSE. RANDOMIZE OPTIONS]

Response ID	Response	Action
01	Removed upfront project cost	RESPONSE ID 01
02	Simple application and approval process	RESPONSE ID 02
03	Recommended by contractor	RESPONSE ID 03
04	Covered the full cost of the project	RESPONSE ID 04
05	No collateral or credit check	RESPONSE ID 05
06	Repayment via utility bill	RESPONSE ID 06
-96	Other, please specify: [OPEN-END]	RESPONSE ID -96
-97	Don't know	Q18
-98	Refused	Q18

Q18. On a scale of 0 to 10, with 0 being "not at all likely", and 10 being "extremely likely", how likely would your organization have been to complete the same project if the PG&E on-bill financing was **not** available? [SINGLE RESPONSE]

Response ID	Response	Action
01	0 – Not at all likely	Q19
02	1	Q19
03	2	Q19
04	3	Q19
05	4	Q19
06	5 – Neither likely nor unlikely	Q19

07	6	Q19
08	7	Q19
09	8	Q19
10	9	Q19
11	10 – Extremely likely	Q19
-97	Don't know	Q19
-98	Refused	Q19

[DISPLAY QUESTION FOR UP TO 2 MEASURES [MEASURE_1/2]

Q19. **Without the PG&E on-bill financing**, when would your organization have installed the [MEASURE_1/2] equipment? [SINGLE RESPONSE]

Response ID	Response	Action
01	At the same time or sooner (e.g., within 6 months)	Q20
02	7 to 12 months later	Q20
03	13 to 24 months later	Q20
04	25 to 36 months later	Q20
05	More than 36 months later	Q20
06	Would not have completed this project at all	Q20
-96	Other, please specify: [OPEN-ENDED RESPONSE]	Q20
-97	Don't know	Q20
-98	Refused	Q20

[DISPLAY QUESTION FOR UP TO 2 MEASURES [MEASURE_1/2]

Q20. **Without the PG&E on-bill financing**, what level of efficiency would your organization have installed for the [MEASURE_1/2] equipment? [SINGLE RESPONSE]

Response ID	Response	Action
01	High efficiency (like the one you purchased)	Q21
02	Mid-level efficiency between standard and high efficiency	Q21
03	Standard efficiency as required by the building code	Q21
04	Below code efficiency	Q21
-97	Don't know	Q21
-98	Refused	Q21

[DISPLAY QUESTION FOR UP TO 2 MEASURES [MEASURE_1/2]

Q21. **Without the PG&E on-bill financing**, what percent of the [MEASURE_1]/2] equipment your organization installed using PG&E on-bill financing would you have installed? [SINGLE RESPONSE]

Response ID	Response	Action
01	Enter percent: [INSERT NUMERIC OPEN-ENDED RESPONSE, ACCEPT 0-100]	Q22
-97	Don't know	Q22
-98	Refused	Q22

Q22. If the **PG&E on-bill financing was not available**, how would the project have changed, if at all? Please try to be as specific as possible. [SINGLE RESPONSE]

Response ID	Response	Action
01	Please specify: [INSERT OPEN-ENDED RESPONSE]	Q23
02	Nothing would have changed about the project without the on-bill financing.	Q23
-97	Don't know	Q23
-98	Refused	Q23

General Program and Non-Program Influence

Next, I'll ask you to rate how important various factors, including the program, were in your decision to move forward with this project. Please use a scale from 0 to 10, where 0 means "not at all important" and 10 means "extremely important." A rating of 8 reflects twice as much influence as a rating of 4. [PLEASE RANDOMIZE]

Q23. The availability of the PG&E PROGRAM on-bill financing. [SINGLE RESPONSE]

Response ID	Factor	Action
01	Record 0 to 10 score: [Record]	Q24
02	Not applicable	Q24
-97	Don't know	Q24
-98	Refused	Q24

Q24. The information or technical assistance provided by the PG&E OBF PROGRAM. [SINGLE RESPONSE]

Response ID	Factor	Action
01	Record 0 to 10 score: [Record]	Q25
02	Not applicable	Q25
-97	Don't know	Q25
-98	Refused	Q25

Q25. Organization practices such as internal policies or goals, maintenance or replacement schedules, and regulatory requirements. [SINGLE RESPONSE]

Response ID	Factor	Action
01	Record 0 to 10 score: [Record]	Q26
02	Not applicable	Q26
-97	Don't know	Q26
-98	Refused	Q26

Q26. [DISPLAY IF Q25=01 > 5] What specific organization policies or practices influenced your decision to do the project? [SINGLE RESPONSE]

Response ID	Response	Action
01	[RECORD RESPONSE]	Q27
-97	Don't know	Q27
-98	Refused	Q27

Q27. Payback or return on the PROJECT. [SINGLE RESPONSE]

Response ID	Factor	Action
01	Record 0 to 10 score: [Record]	Q28
02	Not applicable	Q28
-97	Don't know	Q28
-98	Refused	Q28

If you were given 10 points to award in total, how many points would give to the importance of the program and how many points would you give to these other factors.

Q28. How many of the 10 points would you give to the importance of the PROGRAM (financing and information / technical assistance) in your decision?

Response ID	Factor	Action
01	Record 0 to 10 score: [Record]	RESPONSE ID 01

Q29. And how many points would you give to these other non-program (organization policies and return on investment) factors?

Response ID	Factor	Action

01	Record 0 to 10 score: [Record]	RESPONSE ID 01

Q30. On a scale of 0 to 10, with 0 being "not at all likely", and 10 being "extremely likely", how likely is it that your organization would have undertaken this project at the same time that you did **without any financing** from the program? [SINGLE RESPONSE]

Response ID	Response	Action
01	0 – Not at all likely	Q31
02	1	Q31
03	2	Q31
04	3	Q31
05	4	Q31
06	5 – Neither likely nor unlikely	Q31
07	6	Q31
08	7	Q31
09	8	Q31
10	9	Q31
11	10 – Extremely likely	Q31
-97	Don't know	Q31
-98	Refused	Q31

Satisfaction Questions

Q31. I'm going to list a few elements that may have impacted your satisfaction with the project. For each, please indicate on a scale of 0 to 10, with 0 being "not at all satisfied", and 10 being "extremely satisfied", how satisfied were you with each of the following elements of your program experience [INSERT 0-10 SCALE WITH A DON'T KNOW -97, NOT APPLICABLE -99, AND REFUSED -98 OPTION, RANDOMIZE 1-7]

Response ID	Response	Action
01	The ease of completing the on-bill financing application	RESPONSE ID 02
02	Interaction with PG&E representatives	RESPONSE ID 03
03	Interaction with the contractor	RESPONSE ID 04
04	The amount of documentation required by the on-bill financing application	RESPONSE ID 05
05	The amount of time the program took to review and approve your on-bill financing application	RESPONSE ID 06
06	The amount of time it took to complete work after receiving project approval	RESPONSE ID 07
07	How and when then the on-bill financing charges show up on the bill	Q32

[DISPLAY FOR ANY RESPONSE ≤6 IN Q31]

Q32. You noted dissatisfaction with [INSERT ANY RESPONSE ≤6 IN Q9]. Why were you less than satisfied with this aspect of the program [OPEN-ENDED RESPONSES]

Response ID	Response	Action
01	[DISPLAY IF Q31_01 ≤6] The ease of completing the on-bill financing application	RESPONSE ID 01
02	[DISPLAY IF Q31_02 ≤6] Interaction with PG&E representatives	RESPONSE ID 02
03	[DISPLAY IF Q31_03 ≤6] Interaction with the contractor	RESPONSE ID 03
04	[DISPLAY IF Q31_04 ≤6] The amount of documentation required by the on-bill financing application	RESPONSE ID 04

05	[DISPLAY IF Q31_05 ≤6] The amount of time the program took to review and approve your on-bill financing application	RESPONSE ID 05
06	[DISPLAY IF Q31_06 ≤6] The amount of time it took to complete work after receiving project approval	RESPONSE ID 06
07	[DISPLAY IF Q31_07 ≤6] How and when then the on-bill financing charges show up on the bill	RESPONSE ID 07
-97	Don't know	Q33
-98	Refused	Q33

Q33. Aside from saving energy, which of the following benefits did your organization realize from completing this project? [MULTIPLE RESPONSES]

Response ID	Response	Action
01	Regulatory or corporate compliance	RESPONSE ID 01
02	Improved user comfort	RESPONSE ID 02
03	Greater competitiveness (e.g., brighter retail lighting)	RESPONSE ID 03
04	Increased productivity	RESPONSE ID 04
05	Lower maintenance costs	RESPONSE ID 05
-95	None [MAKE EXCLUSIVE]	RESPONSE ID 06
-96	Other, please specify: [OPEN-END]	RESPONSE ID -96
-97	Don't know [MAKE EXCLUSIVE]	Q34
-98	Refused [MAKE EXCLUSIVE]	Q34

Firmographics

[DISPLAY TEXT: We are almost done! We have a few final questions about your organization.]

Q34. In what year was this organization established at this location? [SINGLE RESPONSE]

Response ID	Response	Action
01	Enter year: [INSERT NUMERIC OPEN-ENDED RESPONSE]	Q35
-97	Don't know	Q35
-98	Refused	Q35

[DISPLAY IF Q31=97]

Q35. Would you say it was…? Your best estimate is fine. [SINGLE RESPONSE]

Response ID	Response	Action
01	Before the 1960s	Q36
02	In the 1960s	Q36
03	In the 1970s	Q36
04	In the 1980s	Q36
05	In the 1990s	Q36
06	Between 2000 and 2005	Q36
07	After 2005	Q36
-97	Don't know	Q36
-98	Refused	Q36

Q36. About what percentage of your overall operating costs does energy account for? [SINGLE RESPONSE]

Response ID	Response	Action
01	Less than 1 percent	Q37
02	1 to 2 percent	Q37
03	3 to 5 percent	Q37
04	6 to 10 percent	Q37
05	11 to 15 percent	Q37

06	16 to 20 percent	Q37
07	21 to 50 percent	Q37
08	51 percent or more	Q37
-97	Don’t know	Q37
-98	Refused	Q37

Q37. What is the main activity at this facility? [SINGLE RESPONSE]

Response ID	Response	Action
01	Please specify: [OPEN-END RESPONSE]	Q38
-97	Don’t know	Q38
-98	Refused	Q38

Q38. Approximately how many people are currently working at the facility where the project was completed, including both full and part time? [SINGLE RESPONSE]

Response ID	Response	Action
01	Fewer than 10	Q39
02	11 to 25	Q39
03	26 to 50	Q39
04	51 to 75	Q39
05	76 to 100	Q39
06	101 to 250	Q39
07	251 to 500	Q39
08	501 to 1,000	Q39
09	1,001 to 2,500	Q39
10	2,501 to 5,000	Q39
11	5,001 or more	Q39
-97	Don’t know	Q39
-98	Refused	Q39

Q39. What is the total enclosed square footage of the facility that you occupy at this location? Your best estimate is fine.

Response ID	Response	Action
01	[RECORD RESPONSE]	Q40
-97	Don't know	Q40
-98	Refused	Q40

Q40. Does your organization own, rent, or manage this facility? Please select the answer that best applies. [SINGLE RESPONSE]

Response ID	Response	Action
01	Own	Q41
02	Rent	Q41
03	Manage	Q41
-97	Don’t know	Q41
-98	Refused	Q41

Q41. Were the answers to the questions you provided in this survey representative of other sites for which you are responsible? [SINGLE RESPONSE]

Response ID	Response	Action
01	Yes	Q42
02	No	Q42
-97	Don’t know	Q42
-98	Refused	Q42
-99	Not applicable – only responsible for one site	Q42

Closing

Q42. Do you have any reflections you wish to share about what was asked in this survey? [SINGLE RESPONSE]

Response ID	Response	Action
01	Yes, please specify: [INSERT OPEN-ENDED RESPONSE]	Q43
02	No	Q43

[DISPLAY TEXT ON TERMINATION SCREEN: You have completed the survey and your responses have been submitted. Your contribution to this survey helps CPUC and PG&E to evaluate and improve program offerings. Thank you for your participation and time.]

PY2022 – PY2024 POPULATION NMEC PROGRAM ADMINISTRATOR INTERVIEW GUIDE

Background

The objective of this interview is to understand the operations and decision-making processes of [Program Name], including program design, implementation, customer acquisition, and how customers who are ineligible or unable to participate are handled. The interview will also explore the roles and responsibilities of organizations involved in program delivery, QA/QC processes, and how evaluation and M&V feedback influence program improvements, implementer compensation, and the accuracy of savings estimates. Finally, it will gather insights on Program Administrators' understanding of the Total System Benefit (TSB) framework, its relevance to claims, and its anticipated impact on incentive structures.

Interview Guide

Introductions

1. Thanks for talking to us today.
2. My name is __.
3. DNV is doing this work on behalf of CPUC to better understand the program.
4. We've reviewed the program PIP and other documentation, so in places where those cover these topics, we are looking for confirmation and updates to what's already documented.

Interviewee Background

5. What is your role in this program?
6. What, if any, other energy efficiency or DR roles do you have with [ORGANIZATION]?

Program design and implementation

The set of questions are to help us understand the program's design and how it is implemented.

7. Let's start with some program theory.
    a. What are the intended objectives or goals? What is it supposed to accomplish?
    b. What market barriers it is designed to cover

c. How is it designed to overcome those barriers?

d. What does the PA get out of the NMEC program(s) that it doesn't get out of the other programs? What is good about it?

    i. Is it reducing the cost to achieve savings? Gross savings or net savings or both?

e. What do aggregators and subcontractors get out of it? Why do they participate? What's in it for them?

    i. How does it compare to a traditional deemed EE program for them?

f. What do end-users get out of it? Why do they participate?

    i. How does it compare to a traditional deemed EE program for them?

g. What is bad about it?

    i. From PA perspective

    ii. From subcontractor perspective

    iii. From end-user perspective

8. Now let's talk about your specific program.

    a. Who are the different players? Who does what? – Get both organization and individual names here

        i. Does the program have an implementer?

        ii. Does the program have aggregators/contractors? What specifically does the aggregator do?

        iii. Does the program have other subcontractors?

        iv. [BAYREN ONLY: Are you able to tell us why you switched away from Recurve this year?]

Roles	Organizations
Program implementer	
Recruitment Implementer, aggregator, contractor, PA, other (please specify)	
Installation PA-delivered program with support from subcontractors, Implementer-delivered program with or without subcontractors, a third-party program, implementer/ aggregator-based program, others?	
QA/QC	
Statistical modeling	
[Other]	
[Other]	

9. The PIPs talked about implementers would have local presence. Which of these organizations is local to your service territory?
10. Do you have any written guidelines about the different roles and responsibilities that each organization in this framework are supposed to follow?
    a. How do you know if they follow those guidelines?
11. How does information flow through this implementation team?
    a. Who do you communicate directly with?
    b. What information do you give/get in those communications?
    c. Do you know who they communicate with and what they talk about?

Program recruitment and participation

The next few questions are to help us understand customer participation in this program.

12. Who in the organization framework has direct contact/communication with customers/end users?
    a. What does that contact involve?
    b. What do they tell the customers?
        i. What are customers told about program incentives or other possible benefits?
        ii. Are customers told they are participating in a ratepayer funded program?
        iii. [For programs with aggregators] What program information or details are aggregators required to share with customers?
    c. Do they keep records of that contact?
    d. Do they share those records with you or anyone else in the framework?
13. How does this program find potential customers?
    a. Through PA outreach activities? Through implementer recruitment efforts? Through contractor/aggregator referrals and marketing?
    b. What organizations in the framework do this?
    c. When a customer is recruited, how does the recruiting organization know whose service territory they are in?
14. How are customers vetted for eligibility?
    a. Who does the vetting? Does anyone review it?
    b. What happens to potential customers who do not meet eligibility criteria?
        i. Are they dropped from consideration? Are they directed to a site NMEC program? Are they directed to a Custom program?

c. What happens to customers who initially meet eligibility but are later dropped from program analysis? Do they still receive program benefits [financing, free equipment/installs, incentives, etc.]?

15. What measures does this program implement?
    a. How do they decide what to implement?
    b. How comprehensive is it?
    c. Who installs or implements those measures?
    d. How are records kept?
    e. Who has access to those records?
16. What QA/QC activities are there?
    a. Does anyone inspect any of the end-user sites? Who? How many?
    b. What do they look at?
    c. Do they keep records?
    d. Who has access to those records?
    e. Is there any other follow up with customers, such as to make sure they are satisfied with the project/program?
17. Many C&I customers are capable of implementing more than one projects. Is anyone looking out for the potential of repeat participation?
    a. Who?
    b. How?

Incentives

The next few questions are to help us understand how program incentives or performance payments are paid and managed.

18. What incentives and/or performance payments does this program provide?
19. Who gets those incentives or payments?
20. How do the implementers/aggregators/contractors make money by participating in this program?
21. What do customers get out of participating in the program?
22. What specific records or documentation are maintained to track the flow of incentive funds (e.g., invoices, payout reports, customer passthrough verification, aggregator payment logs)?
    a. Who maintains each type of record, and how long are they retained?
    b. Does your organization have full access to these records for audit or reporting purposes?
    c. Do you track whether incentives paid to aggregators are fully or partially passed through to customers? If so, how?

23. What specific records or documentation are maintained to track the flow of incentive funds (e.g., invoices, payout reports, customer passthrough verification, aggregator payment logs)?
    a. Who maintains each type of record, and how long are they retained?
    b. Does your organization have full access to these records for audit or reporting purposes?
    c. Do you track whether incentives paid to aggregators are fully or partially passed through to customers? If so, how?
24. The theory behind NMEC is pay for performance...
    a. How do you track performance?
    b. How do you manage payments relative to performance?
    c. Do you have provisions to recover overpayments or address underpayments to aggregators or customers when incentives are based on forecasted savings that later prove to be too high or too low?
25. Are there any incentive caps, minimum performance thresholds, or clawback provisions that affect how payments are issued or adjusted?

Savings traceability

The next few questions are to help us understand how the program handles savings claims and project data traceability.

26. Which organization tracks the list of customers participating in the program?
    a. How is the list maintained as sites are added or excluded?
    b. How is the list QC-ed when a cohort's claims are ready to be finalized?
    c. Do you have access to that list?
27. When cohort savings are claimed, do you have any requirements for maintaining the final data and models to ensure savings traceability?
    a. Which organization maintains the models?
28. How do you communicate with the reporting team when a program cohort is ready to make initial and true-up claims?
    a. How do you QC those claims?

Internal M&V usage

The next few questions are to help us understand how this program has incorporated internal M&V results.

29. Which individuals or groups at your organization review program M&V results?
    a. How frequently do they review M&V results?
    b. What is the granularity (e.g.: site level, cohort level, population level) of the M&V results?
    c. What's the "Truth" value used for the M&V?

30. What processes does your organization have for incorporating M&V results into program designs and operations?
    a. Is there a process in place to communicate these changes to the different implementation organizations?
    b. How does M&V know that changes were made so they can adjust their approach to factor in those changes?
31. Have you made changes to program design, delivery, or incentive structure based on M&V results?
    a. What changes were made?
    b. What were the specific reasons for the changes?

Evaluation feedback

The next few questions are to help us understand how this program has incorporated or will incorporate evaluation results.

32. Has this program previously been evaluated by an external third party? [NOTE TO INTERVIEWER: OBF and MCE have previously been evaluated.]
    a. Which individuals or groups at your organization review evaluation results?
    b. How do evaluation results influence program decision making?
    c. Have you made changes to program design, delivery, or incentive structure based on evaluation results?
        i. What changes were made?
        ii. What were the specific reasons for the changes?
33. What formal processes does your organization have for incorporating evaluation findings (past or future) into program designs and operations?
34. What information do you share with the other PAs?
    a. How frequently do you share?
    b. How useful has the information you've received from other PAs been for your program operations?

Total System Benefit

This set of questions are to help us understand the impact of TSB on claims and incentive structures.

35. What is your current understanding of the TSB framework as it relates to your population NMEC program?
36. How did you approach TSB for initial claims in PY2024, if at all?
37. How did you approach TSB for true-up claims in PY2024, if at all?
38. How do you plan to approach TSB for PY2025 claims?
39. Do you expect the shift to TSB to impact the incentive structures or experiences for program implementers or customers in population NMEC programs?

Closing

40. Do you have any additional thoughts about population NMEC programs you would like to share?

APPENDIX I. POPULATION NMEC EVALUATION REPORT COMPILED STAKEHOLDER COMMENTS

Table I-1 presents DNV's responses to the comments on the draft report that were received during the public review period.

Table I-1. Responses to comments on draft report

#	Commenter	Comment	DNV response
1	BayREN	[Recommendation 1: "PAs should strengthen participant eligibility criteria by enforcing existing M&V eligibility criteria—including the requirement for sufficient pre-installation data—and tightening upfront screening rules to exclude sites with known process-, event-, or production-driven loads. Eligibility screening should also include size limits, restricting participation to sites with annual energy consumption below a pre-defined threshold (for example, 1,000,000 kWh). These criteria will ensure enrollment is limited to sites suitable for standardized pop NMEC modeling, homogenous cohort formation, and compliance with NMEC rulebook precision requirements."] BayREN already applies and enforces rigorous eligibility screening in its BayREN Business program consistent with NMEC modeling requirements. However, BayREN agrees that increased consistency and transparency in upfront eligibility criteria would strengthen confidence in population-level savings claims. Thus, BayREN supports this recommendation to ensure modeling reliability and greater accuracy in savings, including: • Sufficient pre-installation data to support model stability, • Exclusion of sites with known process-, event-, or production-driven loads that cannot be adequately normalized, and • Clear documentation of eligibility decisions for auditability. BayREN agrees with the evaluator's concern that insufficient eligibility rigor has already led to material issues in practice, including enrollment of sites unsuitable for NMEC modeling and resulting impacts on savings accuracy. Greater flexibility must therefore be paired with clear, enforceable structure and accountability. Additionally, BayREN recommends the CPUC require data-driven threshold definitions within a clear, enforceable, and auditable eligibility screening plan. This approach allows PAs to establish program-specific size thresholds, such as those based on cohort baseline distributions or percentile cutoffs, while ensuring they are explicitly defined, justified, and consistently applied.	Noted. Eligibility screening should consider elements such as cohort homogeneity and the number of sites available for M&V, to meet pop NMEC precision requirements. Specific guidance of this nature--up to and including NMEC rulebook revisions--will be addressed separately, via a PCG-led or similar process.

#	Commenter	Comment	DNV response
		For example, learning from BayREN's experience in early 2023, the BayREN Business program implemented an energy threshold cap of 150,000 kWh and/or 50,000 therms as part of its eligibility criteria to target only small and medium hard-to-reach (HTR) businesses as stated in its Program Manual beginning in June 2023. By requiring transparent documentation of these thresholds and their rationale, the CPUC can strengthen eligibility rigor and auditability without imposing overly rigid limits that may exclude otherwise suitable participants. The BayREN Business program also includes in its Program Manual ineligible NAICS codes as part of its eligibility criteria in order to remove businesses that are not best-fit for population NMEC, such as agricultural and industrial businesses. However, the NAICS codes for some businesses may be incorrect in the underlying utility data or give the appearance of a NAICS code that is unsuitable. For example, small HTR businesses such as bakeries are often assigned manufacturing NAICS codes because they manufacture edible goods. However, they are unlikely to present the same process-load complexities as industrial manufacturing facilities that could pose challenges for a population-based NMEC program. As such, PAs should have consistent processes to allow case-by-case exceptions to NAICS codes to ensure businesses can be served.	
2	BayREN	["Recommendation 2. PAs should propose a solution to the reporting PCG that defines a customer project-level pop NMEC dataset that links with 100% accuracy to both initial and true-up aggregate CEDARS claims by the end of July 2026. The solution should reflect an official, date-stamped set of required project claim elements, including, but not be limited to, claim and parent claim IDs; project identifiers; participant (end-user) IDs and contact information; projected and true-up savings for all fuel types (kWh, kW, therms) and TSB; installed measures; EULs; realization rates; NTG values; and implementation partners (implementer, aggregator or contractor, and M&V vendor)."] BayREN agrees with the evaluator that current population NMEC reporting lacks sufficient standardization and transparency at the project level, which creates challenges for replicability, auditability, and efficient EM&V. The evaluation findings demonstrate that, in many cases, it was not possible to	Noted and thank you for the feedback. The idea of site-level reporting for non-residential pop NMEC sites will be discussed at PCG.

#	Commenter	Comment	DNV response
		directly reconcile project-level data with aggregate CEDARS claims without significant manual effort or reconstruction. BayREN acknowledges program-level QA/QC improvements are needed. To strengthen standardized reporting, BayREN recommends that the CPUC Rulebook for Programs and Projects Based on Normalized Metered Energy Consumption (NMEC Rulebook) be updated to add prescriptive reporting requirements for population-based claims. BayREN recommends that the CPUC update the NMEC Rulebook to define a standardized, auditable project-level dataset for population NMEC programs that: • Establishes a required, version-controlled set of claim elements (e.g., project ID, claim/parent claim ID, customer ID, projected and realized savings by fuel, NTG, realization rates, EUL, and implementer roles), • Ensures one-to-one traceability between project-level records and aggregated initial and true-up CEDARS claims, and • Is maintained in a consistent format across program administrators to support independent verification. BayREN supports the objective of achieving full linkage between project-level data and program-level claims. Importantly, BayREN recommends enabling site-level reporting for population-level NMEC. Current population NMEC reporting structures are often interpreted as requiring single-line or cohort-level reporting of savings which limits the ability to accurately attribute grid impacts using project-specific load shapes. This constrains the calculation of Total System Benefits (TSB), particularly when custom or measure/project-specific load shapes would more accurately reflect hourly savings impacts. Site-level claims would enable implementers and PAs to: • Apply custom or project-specific load shapes where appropriate, • Improve the accuracy and defensibility of TSB calculations, and • Better align reported savings with actual system value by capturing hourly variation in load reductions. • Improve traceability and replicability	

#	Commenter	Comment	DNV response
		As TSB plays an increasingly central role in program evaluation and cost-effectiveness, ensuring that reporting structures support accurate load shape attribution is critical. Moving toward site-level transparency would strengthen not only EM&V processes but also the integrity of grid-value calculations. In summary, BayREN supports the intent of the recommendation and emphasizes that the most effective path forward is NMEC Rulebook clarification and standardization of data structures, including enabling site-level attribution where necessary to support both traceability and accurate valuation of savings.	
3	BayREN	[" Recommendation 3: The PAs need to develop a plan to move toward simplified and consistent modeling approaches, such as time-of-use temperature (TOWT) models, which prioritize clear representation of weather and schedule effects while making it easier to identify NREs that influence eligibility decisions. The PAs should examine how alternative simplified modeling approaches affect hourly savings consumption profiles, ensuring these methods support accurate TSB estimates."] BayREN supports the evaluator's emphasis on improving the consistency and credibility of population NMEC savings estimates. As noted in Section 1.5.1.2, variability in modeling methods, treatment of non-program effects, also known as exogenous effects, and inconsistent characterization of uncertainty can create challenges for cross-program comparability and stakeholder confidence. However, BayREN believes the findings refer less to a need for prescribing specific modeling methodologies, but more to a need for clearer and more complete guidance on population-level uncertainty and savings validation in the NMEC Rulebook. The evaluation highlights that differences in modeling approaches and comparison group selection can lead to variability in savings estimates. This is an expected feature of a flexible, market-driven NMEC framework that accommodates diverse building types and operating conditions. Attempting to standardize model forms (e.g., promoting Time of Week and Temperature models) would not fully resolve this variability and may introduce unintended constraints and hamper innovation. At the same time, maintaining methodological flexibility does not imply an absence of standards. To ensure consistency and credibility, modeling approaches should meet minimum performance and transparency criteria, including reasonable goodness-of-fit, appropriate representation of key drivers such as weather and operational schedules, explicit identification and treatment	The NMEC rulebook has generally been interpreted as allowing methodological flexibility, which has been appropriate given the early stage of NMEC. However, this flexibility reflects the absence of explicit guidance rather than a clear intent to maintain methodological diversity across programs. At the same time, the rulebook is explicit that methods must be standardized within a given pop NMEC program. Our recommendation is intended to prompt consideration of whether greater consistency across programs could improve clarity, predictability, transparency, and stakeholder confidence, while still allowing methodological development where justified. Pop NMEC modeling approaches need minimum expectations for transparency, interpretability, replicability, treatment of NREs, uncertainty characterization, and clearly defined normalization approaches in program M&V plans. Broader rulebook considerations regarding method standardization, uncertainty, and normalization will be addressed separately through the PCG process.

#	Commenter	Comment	DNV response
		of non-routine events, and clear documentation sufficient for independent replication. Instead, the appropriate point of standardization is at the program-level savings validation framework, specifically through Fractional Savings Uncertainty (FSU). BayREN agrees with the evaluator that current uncertainty practices do not consistently provide sufficient evidence to demonstrate compliance with the NMEC Rulebook's precision expectations. However, this is likely due to a lack of clarity and completeness in the NMEC Rulebook's presentation on how FSU should be applied, in population-based programs. Thus, BayREN recommends that the CPUC clarify and strengthen FSU guidance in the next version of the NMEC Rulebook to address the following key gaps identified in the evaluation: 1. Application of FSU at the Appropriate Level. The evaluation notes challenges in demonstrating 90/25 precision and suggests relaxing this standard for smaller cohorts. However, the NMEC Rulebook defines precision expectations at the annual program savings level, not at the cohort or individual model level. Clarifying this distinction would help ensure consistent interpretation and avoid unnecessary segmentation or misapplication of precision criteria. 2. Inclusion of Cross-Site and Comparison Group Uncertainty. The evaluation correctly identifies that uncertainty introduced by comparison group selection and non-participant modeling approaches is not fully captured in current practice. This represents a material gap in FSU application. The Rulebook should explicitly define how to incorporate: • Variability introduced by comparison or control group applications, and • Cross-site heterogeneity in savings performance into program-level uncertainty estimates. 3. Limitations of Existing Referenced Methods (e.g., ASHRAE 14). Current references (such as ASHRAE 14) provide formulas applicable to single-model uncertainty, which are insufficient for population NMEC programs. Additional guidance is needed on aggregation of uncertainty across models and multiple sites to produce defensible program-level FSU values, such as FSU portfolio provided in the OpenDSM (formerly CalTRACK) Methods Section 4.3.2.5. 4. Consistency in Uncertainty Reporting and Documentation. The evaluation finds that uncertainty metrics were inconsistently calculated and reported across programs. Establishing standardized requirements for the following would improve transparency and comparability without constraining modeling approaches:	

#	Commenter	Comment	DNV response
		• FSU calculation methodology, • Documentation of assumptions, and • Reporting formats These validation requirements should be complemented by minimum expectations for model diagnostics and reporting to ensure that variability in modeling approaches does not translate into unbounded variability in results. In summary, BayREN supports the evaluation's focus on improving savings accuracy and stakeholder confidence. However, BayREN recommends that the CPUC: • Maintain methodological flexibility in modeling approaches, • Clarify and strengthen FSU as the governing population-level uncertainty metric, and • Provide explicit guidance on incorporating comparison group, cross-site, and aggregation uncertainty into program-level savings validation. This approach directly addresses the evaluators' concerns while preserving the flexibility needed for population NMEC programs to accurately represent diverse real-world conditions. BayREN also notes that the evaluation places emphasis on normalizing savings to "typical weather conditions" as a consistent modeling practice. While this is appropriate for site-level NMEC claims, BayREN does not agree that this requirement is explicitly defined or intended for population-level NMEC programs in the current NMEC Rulebook. Doing so for population-level programs would be computationally intensive and add program design restrictions. Using reporting year weather (Avoided Energy Use) is a valid savings claim approach for population-level programs under current rules.	
4	BayREN	[Recommendation 4: The PAs should apply program-level pop NMEC savings uncertainty approaches that reflect both modeling error and other sources of cross-site variability."] BayREN agrees with the evaluator that population-level NMEC programs require a more comprehensive and consistent approach to characterizing savings uncertainty. Accurately representing uncertainty is essential to demonstrating compliance with the NMEC Rulebook's precision expectations and building confidence in reported savings outcomes.	Specific guidance of this nature--up to and including NMEC rulebook revisions--will be addressed separately, via a PCG-led or similar process.

#	Commenter	Comment	DNV response
		However, the current NMEC Rulebook does not provide sufficient guidance on how to aggregate uncertainty across sites, models, and program structures, which leaves room for inconsistent application across programs. As such, BayREN recommends that the CPUC strengthen and clarify the treatment of FSU as the governing metric for population NMEC programs through a NMEC Rulebook update by defining the governing program-level uncertainty rules that include the following components: • Inclusion of uncertainty introduced by comparison group applications, • Aggregation methodology (portfolio-level FSU) with clear guidance on how to combine these uncertainty components across sites, to what level or extent, where to include comparison group uncertainty, and the equations to use for the single program-level metric (e.g., such as those outlined in OpenDSM).	
5	AF/SCE	[Recommendation 1: "Eligibility screening should also include size limits, restricting participation to sites with annual energy consumption below a pre-defined threshold (for example,1,000,000 kWh)."] SCE would caution against establishing rigid “size limit” requirements. Instead, we recommend adopting clear statistical approaches and methodologies to determine appropriate size limits based on the characteristics of the participating pool.	Noted. See response to Comment #1.
6	AF/SCE	[Section 1.5.1.2 "While all evaluated program implementers used modeling approaches to account for routine, recurring drivers of consumption, such as weather, and non-program-related factors, methodological choices did not always best serve savings accuracy. For example, only one program normalized savings to typical weather conditions, consistent with NMEC rulebook intent. Other programs used reporting-year weather, which increases savings errors in multiple ways."] SCE understand that PAs and Implementers would benefit from having clearer guidance on the specific approach/method for normalizing program savings using the reference TMY used in ACC vintage. SCE additionally understand that normalizing program savings to TMY introduces additional uncertainty to program claims.	TMY has clear advantages for estimating first year savings that will be combined with EUL for lifetime savings, as well as TSB. The recognized shortcomings of TMY do not, at this time, appear to change that finding, though further discussion on this topic is warranted.

#	Commenter	Comment	DNV response
7	AF/SCE	[Recommendation 4 "The PAs should apply program-level pop NMEC savings uncertainty approaches that reflect both modeling error and other sources of cross-site variability."] SCE agrees with recommendation; however, recommends for clear guidance on how (in addition to FSU) uncertainty at the cohort (and/or beyond the model error) should be evaluated and tracked. SCE additionally agrees that PAs would benefit having specific guidance for evaluating uncertainty for GPs and Comparison Groups.	Noted. See response to Comment #4.
8	kW Eng for SCE	[Section 1.5 "Savings measurability can be improved through a combination of stronger adherence to program rules regarding the appropriate type of load for pop NMEC meter-based methods and using technical methods that provide simple, transparent, and consistent results."] The PopNMEC program Rulebook rules can be enhanced by defining requirements in many areas: - modeling methods - goodness of fit requirements - accuracy requirements - approach to determining exogenous effects - population requirements - comparison group design -etc. We suggest on future enhancement to Rulebook guidance not to over-specify requirements which can lead to low participation.	Broader NMEC rulebook revisions and recommendations, including those related to modeling, screening, comparison group design, and balancing specificity with participation, will be addressed separately through the PCG process.
9	kW Eng for SCE	[Section 1.5.1 "Additionally, pop NMEC methods need to be streamlined so that they provide consistent final estimates of savings across PAs and implementers."] We suggest for program evaluator (DNV) to provide clear guidance as how methods need to be streamed in terms methods and requirements including those to control for exogenous effects. In short, more specific recommendations would be helpful.	Noted. Specific guidance of this nature--up to and including NMEC rulebook revisions--will be addressed separately, via a PCG-led or similar process.
10	kW Eng for SCE (and SCE)	[Section 1.5.1.1 "Nevertheless, approximately 15% of participants in the PY2022 – PY2023 programs came from sectors that commonly exhibit process- or event-driven energy consumption (manufacturing, agriculture,	Noted. We look forward to including some or all of these ideas in guidance going forward.

#	Commenter	Comment	DNV response
		infrastructure and utilities, etc.), which cannot be modeled accurately under pop NMEC. The participation of these ill-suited sites affects savings accuracy because their energy consumption patterns are often so variable they cannot provide a reliable baseline from which to estimate savings."] The Rulebook guidance can be improved to avoid sites with these issues: -drivers of energy use (weather, operation hours, etc.) must be the same for all sites in treatment and comparison groups. -loads must exhibit a regular and predictable pattern of energy use -there must be a large population of eligible sites to define significant populations within treatment and comparison groups. -Manufacturing, agriculture, infrastructure, and utility sites are not appropriate for population NMEC methods. - Energy consumption specification between the mean (kWh) consumption relative to percent consumption within the 3rd energy consumption quartile, e.g., assessment on the level of kWh consumption outside the "mean."	
11	kW Eng for SCE (and SCE)	[Section 1.5.1.1 "By addressing highly variable levels of energy consumption among their populations, all programs can better align with the NMEC rulebook and improve their savings precision, both of which are essential program goals."] We recommend for the evaluator (DNV) to clarify if recommendation is that programs may target multiple different cohorts of the same building type, climate zone, etc., but within each cohort the equipment, levels of use, drivers of energy use, etc. must be similar. Note that there are no specific requirements in Rulebook for participating building types within a cohort to be the same; rather, for similar drivers of energy consumption and/or energy savings attributions within sites.	The following lines from p.12 of the NMEC rulebook 2.0 clearly indicate to us that there should be similarity across sites within a cohort. This is also consistent with our understanding that this similarity is essential to address uncertainty in methods. This will all be part of upcoming NMEC methods discussions. 2. Population-level NMEC program sites must have building-type similarity such that: a. The sites can reasonably be expected to have similar types of equipment holdings, as well as drivers and levels of energy consumption. b. There should be a reasonable expectation that the factors that impact both 1) consumption over a 12-month period, as well as 2) energy savings from program interventions, will be similar across all sites in the population.
12	kW Eng for SCE	[Recommendations 1 & 2 "PAs should strengthen participant eligibility criteria by enforcing existing M&V eligibility criteria—including the requirement for sufficient pre installation data—and tightening upfront screening rules to exclude sites with known process ,event ,or production driven loads. Eligibility screening should also include size limits, restricting participation to sites with	See Comment #1. Accounting for exogenous effects is complex, and, as noted in the report, approaches such as synthetic controls warrant further investigation to assess their ability to control for non-program effects.

#	Commenter	Comment	DNV response
		annual energy consumption below a pre-defined threshold (for example, 1,000,000 kWh). These criteria will ensure enrollment is limited to sites suitable for standardized pop NMEC modeling, homogenous cohort formation, and compliance with NMEC rulebook precision requirements."] We understand that including a maximum annual consumption limit as a participation criterion may be unwise as some business types such as Hospitals may be excluded. A better criterion may be requiring large populations of comparison groups for eligibility – assuring that methods for accounting for non-program effects are based on significantly sized groups. The pop NMEC rulebook should recognize the different ways exogenous effects can be accounted for. It's a series of decisions and not a straightforward procedure the Rulebook can address. This is a big topic. Suggest reviewing the DSA/PG&E NMEC Accuracy report on PDA. Note that granular profiles used as synthetic controls may be developed from smaller consumption sites and may also explain non-program effects. This is an area that requires further research.	
13	kW Eng for SCE	[Section 1.5.1.2 "Approaches to modeling energy use at the site level varied in their alignment with NMEC rulebook expectations. While all evaluated program implementers used modeling approaches to account for routine, recurring drivers of consumption, such as weather, and non-program-related factors, methodological choices did not always best serve savings accuracy. For example, only one program normalized savings to typical weather conditions, consistent with NMEC rulebook intent."] Normalizing to typical weather conditions requires a model of reporting period energy use. Including another model adds additional savings uncertainty. Elsewhere the report suggests increasing the savings uncertainty limit to that of site-level NMEC (90/50) but also notes multiple programs evaluated savings uncertainty exceeded this limit as well.	Normalization to TMY is the default expectation and makes the most sense for TSB and lifetime savings. Application of a baseline period model to post (without the necessity of establishing model fit, etc. for that period) always seemed like the most practical interpretation of the intent of the NMEC rulebook on this issue. Savings uncertainty expectations need to be in line with best, most reasonable set of procedures.
14	kW Eng for SCE	[Section 1.5.1.2 "Most program implementers' uncertainty methods and reporting do not provide sufficient evidence to support the credibility of claimed pop NMEC savings under the NMEC rulebook's precision expectations. Characterizing uncertainty is essential for establishing claimed pop NMEC savings are valid. Despite this, only two programs reported their calculation of uncertainty metrics. In addition, common uncertainty practices (e.g., reliance on	Thank you for the comment. We agree on the need for further research into uncertainty contributions from the determination of non-program effects. While comparison group methods reduce bias, they also introduce uncertainty. Under current pop NMEC implementations, particularly those using granular profiles or embedded comparison adjustments, this uncertainty may not

#	Commenter	Comment	DNV response
		FSU) may not fully reflect sources of program-level uncertainty beyond model error, such as the uncertainty introduced by the selection of comparison sites to account for non-program effects."] Uncertainty contributions from the determination of non-program effects is an area needing further research. Note that granular profiles used as covariates in the regression model may embed this uncertainty directly in each model.	be fully reflected in reported precision metrics. Although granular profiles used as covariates may embed some of this uncertainty within model estimates, that uncertainty's magnitude and treatment are not well understood and warrant additional evaluation through sensitivity testing.
15	kW Eng for SCE	[Recommendations 5 & 6 "PAs should prioritize customers with demonstrated barriers (e.g., limited access to financing or constrained capital cycles), as NMEC uniquely credits “stranded savings” from to code upgrades that are more common in these segments."] This recommendation may severely limit eligibility considering that determination of non-program effects would also require populations of such customers. Are there enough such customers? Does this requirement place an undue burden on program implementation? A separate method for capturing stranded savings potential would be preferred.	Stranded savings is one way to make NMEC's existing conditions baseline, counted for the full EUL, consistent with the treatment of gross and net savings in other California EE programs. Our recommendation intentionally acknowledges this without putting forth an unrealistic hard requirement of stranded savings. Ultimately, this inconsistency will need to be addressed.
16	kW Eng for SCE	[Section 1.5.2 "Thus, most of the pop NMEC programs functioned more like lighting programs than comprehensive whole-building programs, which suggests that many pop NMEC programs are not fully realizing their intended role in capturing stranded savings through broader modernization."] Agree completely. This may be an outcome of using aggregators who are lighting contractors or similar trade professionals. Achieving EEM installs for multiple end-uses is the unaddressed issue that blocks deeper savings. Incentive structures with bonuses may be one way to encourage multiple trades on projects.	Noted. Our recommendation to increase the share of mixed-measure or more comprehensive projects is intended to improve alignment with NMEC policy objectives, not to prescribe changes within current program frameworks. This direction is also consistent with the CPUC’s focus on maximizing TSB, which is more likely to result from multi-measure projects. Achieving this shift would likely require changes to incentives, performance metrics, or broader program design.
17	kW Eng for SCE	[Section 1.5.2 "Pop NMEC should be reserved for customers where the approach adds unique value that addresses stranded savings, rather than absorbing high-risk savings from customers that have better-fit alternatives."] Agree. Pop NMEC seems most appropriate for the under-served small and medium businesses sector, where savings potential is great and populations are abundant - this directly addresses the hard-to-reach regulatory policy where other program methods fail. It may be worthwhile to add this in the NMEC Rulebook at a higher priority than achieving ‘stranded savings’ in what may be	As noted in our response to comment #15, stranded savings is the most obvious pathway toward consistency with the wider California EE savings framework. Under-served customers are certainly consistent with that emphasis.

#	Commenter	Comment	DNV response
		a limited population. Neither priority is also to achieve deep savings with a mix of EEMs.	
18	kW Eng for SCE	[Section 2.2 "How well are programs learning and adapting? Specifically, are they revising offering expectations and improving the accuracy of savings estimates based on implementation experience and available evaluation findings?"] Good objective. Pop NMEC is an emerging methodology and there are numerous areas of growth: -Granular profiles as synthetic controls -Pre-testing of methodologies and data from targeted building sectors and method of accounting for non-program effects. -how to get deeper multi-measure savings?	Noted and agreed.
19	kW Eng for SCE	[Section 2.2 "Are eligibility rules being applied in ways that are transparent and consistent with NMEC rulebook expectations? How and why are customers excluded or redirected from pop NMEC programs, and to what extent are these decisions justified? What guidance may be needed to ensure eligibility processes do not introduce risk or distort program outcomes?"] It's unclear what the eligibility rules are in the NMEC Rulebook. Perhaps, this may be an area where additional guidance and/or clarity is recommended.	Additional guidance may be useful to clarify pop NMEC eligibility expectations in the NMEC rulebook, and broader rulebook clarifications should be addressed through the appropriate PCG process.
20	kW Eng for SCE	[Section 3.2.1.1 "Validation. In the second step, we conducted a deeper assessment of the program M&V process itself, including the evaluation of savings precision. In particular, we assessed: • Compliance with NMEC rulebook requirements, including: – Customer eligibility and screening practices – Customer attrition and exclusion approaches – Modeling approaches and associated documentation – Normalization of pre-installation (baseline) and post-installation (performance) period results"] 'Normalization' is a word tossed about because it has many contexts. We don't believe the Rulebook spelling out that the context is what IPMVP calls 'adjustment' of energy use to other conditions and would suggest the Rulebook can clarify this.	Please see our response to comment #34. The use of "normalization" in the NMEC context would benefit from clarification through the appropriate PCG process.

#	Commenter	Comment	DNV response
21	kW Eng for SCE	[Section 4.1.1 Table 4.1 showing relative precision results.] Since each of the programs shown in the table do not meet the 90/25 criterion (and recognizing that the programs did not report their savings uncertainty), what does this say about evaluator's uncertainty results? Assuming implementers results are comparable or higher, did these programs fail to achieve their goals? Additional clarity and/or guidance is recommended herein.	To clarify, PY2023 OBF electric and demand savings estimates and PY2023 BayREN demand estimates do meet the NMEC rulebook precision requirements, as shown in Tables 4-1 and 4-3. The evaluation results indicate that many non-residential cohorts align more closely with a 90/50 benchmark, which is consistent with billing analysis practice, especially for smaller populations, highlighting the need for a consistent and broadly accepted uncertainty standard. At this point that standard is not clear enough to identify a failure to achieve program goals.
22	kW Eng for SCE	[Section 4.2.6.1 "On-Bill Financing (OBF) program (synthetic control approach). The OBF program used a synthetic control (SC) approach to model site-level hourly energy consumption as a function of piecewise-linear temperature terms, time-of-week indicators for each hour of the week, and a granular comparison-group load profile. Unlike the MAP-style and non-MAP-style approaches, which estimate participating customer site and comparison group models in parallel, this approach directly incorporates the comparison group as an additional explanatory variable within the site-level regression. This method of controlling for non-program effects is novel, and its statistical properties have not yet been widely vetted in the evaluation context. The SC approach is not flexible to month-over-month change in the same way as the other two specifications, but it will necessarily impose any month-over-month change that is present in the granular profile. This avoids the problem of incorporating NREs into the model itself but does still carry trends into counterfactual energy consumption."] On italicized portion of text: This sentence is unclear. Please clarify what is meant by 'SC approach is not flexible...' and 'will impose and month-over-month change that is present in the granular profile'. For the latter, GPs should be made up of a sufficiently large number of sites so that the presence of month-over-month change is justified.	Thank you for the comment. As noted, these are novel methods and we are attempting to understand their implications. What we consider to be an important finding, and one that we intend to discuss with the wider stakeholder group, is this: Change over time that is captured within the annual models is problematic in a number of ways. Monthly dummies are the most obvious example, potentially allowing a model to account for a month-long NRE though it is not "explained" by an informed independent variable, just a dummy. These are problematic because they obscure NREs in the eligibility phase and pass NREs through to impact shapes in the performance phase. The SC models do not have monthly dummies so do not have this specific problem. However, the inclusion of GPs does allow the models to reflect trend (or changing levels of consumption) within the year, which represent effectively the same problem. For the purpose of savings and TSB impact shapes, we want to represent change over time annually, between baseline and performance periods, but not within periods. Modeled GPs in a DID context would represent this year-over-year change (as long as no monthly variables were included) but it appears that SCs would also incorporate sub annual change over time in a way that appears problematic.
23	kW Eng for SCE	[Section 4.2.6.2 "The SC approach is a more recent extension of this framework and remains relatively untested in the broader energy efficiency evaluation industry, particularly in how it disentangles weather effects from non-program impacts. Additionally, while it avoids embedding NREs directly in the	We recognize the advantages of GPs from a data security perspective. The relevant question is, "What is the best applications of GPs in the context of pop NMEC?" As discussed in the report and comment above, it is the sub-annual change

#	Commenter	Comment	DNV response
		model, it can impose month-to-month trends from the comparison group profile, carrying those patterns into counterfactual energy consumption."] The SC method offers some advantages over comparison groups, including: -making them available on public websites -avoiding data security requirements for collecting non-participant data Isn't carrying month-to-month trends from the profiles actually a benefit of the method? If this isn't what is meant, please clarify statement.	over time that may undermine useful and meaningful impact 8760 shapes. We will make sure this is reflected in report verbiage, but reaching consensus on this topic will require a broader stakeholder process beyond the scope of this report and the present set of comments.
24	kW Eng for SCE	[Section 4.2.6.3 "The NMEC rulebook requires that pre-selected pop NMEC methods be applied uniformly across all sites within a cohort. Our review found that each PA employed a defined modeling framework that was applied consistently across participating sites, rather than tailoring approaches or inputs on a site-by-site basis."] Implied in this Rulebook rule is that the M&V methodology is fixed and applied uniformly to all participating sites and comparison group methods. Isn't it a fundamental condition that the M&V method be fixed so that all variability comes from the data? This can rule out methods where there is flexibility in the regressions, for example, as cited elsewhere in this report (e.g. elastic net and machine learning methods).	Yes, the M&V methodology should be pre-specified and applied consistently across sites so that variability arises from the data rather than changes in the modeling approach. The evaluation confirmed that a defined modeling framework was applied uniformly within each program.
25	kW Eng for SCE	[Section 4.2.8 Uncertainty Characterization section] This discussion does not mention anything about population size. As a rule, as population sizes grow, uncertainty diminishes. Were population sizes adequate to lower uncertainty? Due to their structure, pop NMEC programs often don't achieve large populations on their cohorts, so investigating population sizes is recommended.	The report discusses population size as one of the cohort characteristics that affects savings uncertainty and precision. As noted in Section 4.2.8, program-level uncertainty reflects modeling choices, comparison group construction, cohort characteristics, and program population size. Our evaluation also notes that smaller sample sizes can make it more difficult for non-residential pop NMEC cohorts to meet the 90/25 precision standard, with many results aligning more closely with the 90/50 benchmark commonly used in billing analysis. Section 4.9.2.2 further explains that pop NMEC relies on large sample sizes to manage the additional variance introduced by comparison-group adjustments and aggregate NREs. We will clarify this connection in the report so the role of population size in uncertainty and precision is more explicit.

#	Commenter	Comment	DNV response
26	kW Eng for SCE	[Section 4.2.8 "Given this, we characterized program-level uncertainty in a way that reflects these shared sources of uncertainty by using the standard error of the mean savings across sites (s/√n) to report relative precision."] Program uncertainty was based on the dispersion of a group of individual site and cohort-level savings? Please describe how this reflects 'shared sources of uncertainty' particularly the uncertainty in individual site savings. This seems to be incorrect. For individual sites, as in site-level NMEC, regression model and savings uncertainties are determined using ASHRAE's FSU (or other method) for each site which may then be propagated up to the program level. Assuming the models were re-created, why wouldn't this approach be used in the evaluation?	We recognize two ways of addressing uncertainty: (1) aggregating site-level uncertainty or (2) estimating uncertainty directly from the dispersion of savings across sites. The latter approach reflects the observed variability in realized savings across sites, which incorporates both within-site estimation error and between-site variability. It also captures comparison group uncertainty to the extent that it is reflected in the site-level "savings" estimates. This has also been the standard way of estimating uncertainty in billing analysis. We would argue this is most consistent with NMEC rulebook guidance but recognize that the rulebook is not explicit on this point. More explicit guidance on this will be considered in a stakeholder process.
27	SDG&E	[Overarching] Beyond the non-commercial building or customer types that account for the 15% of disqualified customers, are the remaining disqualified or reduced-savings projects primarily due to model disqualification or comparison group quality issues? If available, SDG&E would like to know what percentage of projects is attributed to each of the additional reasons for disqualifications or reduced-savings?	The 15% figure was used to inform future program design recommendations and did not determine whether sites received evaluated savings. Only sites with insufficient pre-installation data (about 3% across evaluated electric projects, with a higher share for gas) received no evaluated savings; other differences between claimed and verified savings were primarily due to replication, documentation, and validation differences, as summarized in the report's comparisons of claimed and evaluated savings (Figures 4-1 through 4-3 and Tables 4-1, 4-3, and 4-4). Thus, the 15% figure should be understood as evidence supporting stronger upfront screening, not as an exclusion rule applied in the evaluation.
28	SDG&E	[Section 1.4.2] The report observes that PAs used different avoided cost years for initial versus true-up claims, suggesting that this inconsistency may undermine the measurability of savings. SDG&E notes that reporting requirements require the use of the current Avoided Cost Calculator (ACC) version listed on CEDARS for the applicable year at the time of true-up, which can differ from the version used in initial claims. While applying a consistent avoided cost basis across NMEC projects could improve comparability, final reporting must still align with the ACC version in effect at the time of filing. Providing additional guidance on how to meet this	Additional guidance may be needed as PAs work to use the same avoided cost year for both the initial and true-up claims. Using different avoided cost calculator versions and CET quarters for the initial claim and the true up claim will result in inaccurate results because the delta will be calculated with different inputs than the initial claim. Stakeholders will need to collaborate to determine the best way forward.

#	Commenter	Comment	DNV response
		requirement, and how to interpret any resulting differences, would support more consistent and practical implementation.	
29	SDG&E	[Section 3.2.3] The NTG methodology relies on customer surveys and averages across multiple influence drivers to estimate NTGRs. SDG&E notes this approach may understate program influence for Population NMEC, as customers generally lack the capability to implement these projects without program support. Averaging across drivers may also dilute key sources of influence. Greater consideration of the level of technical support and effort required to deliver these projects may better reflect true program influence and produce more representative NTG values.	The NTG methodology does not assume that customers lack the capability to implement projects without program support. Rather, a core objective of the methodology is to assess, based on customer responses, whether projects would have been implemented absent the program. If customers lacked the capability to do so, the survey instrument is designed to capture that directly. Regarding the concern that averaging across influence drivers may dilute key sources of program influence, the averaging approach is intended to balance multiple dimensions of attribution. It reduces the impact of both high and low values relative to approaches that rely on a single driver or select the maximum or minimum value. The methodology applies equal weighting across three primary attribution indicators: decision-making factors, timing of the project relative to program involvement, and counterfactual timing and scope. To address potential dilution from a large number of influence factors, the methodology has already been refined. Specifically, the number of inputs within the decision-making indicator (PAI 1) was reduced from 14 to 4 to better focus on the most relevant drivers of program influence.
30	SDG&E	[Recommendation 7] The report recommends PAs should look beyond lighting-only projects and increase the share of projects toward a more mixed-measure approach. While this direction may have merit, it is unclear whether the current dominance of lighting reflects market conditions, cost effectiveness, or program design. Additional context would help determine the extent to which this trend can realistically be influenced. SDG&E notes that program design is developed by third-party implementers, and under the current model, SDG&E does not impose restrictions on how those programs are designed. The 3rd party program model intentionally allows for implementer flexibility, with measure selection left to their discretion as long	Noted. As indicated in response to comment #16, the recommendation to increase the share of mixed-measure or more comprehensive projects is intended to improve alignment with NMEC policy objectives, rather than to prescribe changes that can be readily implemented within existing program structures. This shift is also consistent with CPUC's broader focus on EE portfolios with greater TSB values, which are more likely to result from multi-measure than lighting-only installations.

#	Commenter	Comment	DNV response
		as program savings can be reliably verified. If there is a goal to shift toward more comprehensive projects, additional guidance and alignment may be necessary. Without such changes, it is uncertain whether programs can move away from a reliance on lighting-dominant projects within the existing framework.	
31	SDG&E	[Recommendation 8] The recommendation to establish clearer referral pathways for customers that are not a good fit for Population NMEC is reasonable in concept. However, current program structures are not designed to support this approach. The deemed pathway does not typically accommodate these project types, and custom or Site-Level NMEC pathways can be overly burdensome for smaller or less complex projects. Population NMEC remains attractive due to its flexibility across a wide range of sites. In cases where projects are not a fit for Population NMEC and are too small to justify Site-Level NMEC, they may not move forward at all. As a result, while the recommendation is noted, its practical feasibility is uncertain under existing program frameworks. SDG&E also notes potential inconsistency with broader NMEC guidance prioritizing NMEC-based program designs. Absent a more streamlined, “light-touch” alternative, there is a risk that viable opportunities are lost, not due to a lack of savings potential, but due to structural limitations in available pathways.	Since comments 31 and 32 raise the same issue, the response below addresses both. There are two distinct but related considerations. First, there is a need for a more complete accounting of potential participants that implementers or aggregators engage with but ultimately screen out of pop NMEC, along with the reasons for that determination. Establishing a consistent process to track these customers and the basis of their ineligibility would improve visibility into the types of projects not moving forward and allow PAs to better assess whether and how their savings potential could be captured through existing or modified pathways. Second, of those projects that were enrolled but later identified as unsuitable for pop NMEC, the evaluation found that approximately 15% were unsuitable due to load characteristics (for example, industrial, agricultural, or event-driven loads). These types of customers are generally better aligned with other pathways, such as Custom or SEM programs. Similarly, larger commercial or commercial-like customers with high energy consumption may be better suited to site-level NMEC. We acknowledge the concern that current program structures may not consistently provide a practical alternative pathway, particularly for smaller or less complex projects. However, our recommendation is not intended to assume that all screened-out projects can be immediately redirected within existing program designs. Rather, it is intended to ensure that PAs have visibility into these projects and their characteristics, so that any structural gaps in available pathways can be identified and addressed over time.

#	Commenter	Comment	DNV response
			The intent is therefore to improve transparency and program design alignment, not to impose a requirement to refer all projects to an existing alternative pathway where none may be feasible.
32	SDG&E	[Recommendation 8] The recommendation to implement a consistent tracking and referral process for projects screened out of Population NMEC is reasonable in concept. However, in practice, available pathways are limited. Transitions are generally within the same program from Population NMEC to Site Level NMEC, and there are not always viable deemed or custom alternatives for these project types. As a result, it is unclear how broadly this referral approach can be applied. Additional clarification on feasible pathways and program alignment would improve practicality.	See response to comment #31.
33	PCE	[Model Vintage Used in Replicability Analysis] Following review of the Draft Report, PCE identified what appears to be a discrepancy between the modeling methodologies described in the report and the methodologies used to generate savings claims for PCE's FLEXmarket Program during program years (PY) 2022-2023. Section 2.3 of the Draft Report states that BayREN, MCE, PCE, SCP, and SDG&E operated pay-for-performance programs with implementation structures similar to the CPUC's Market Access Programs (MAP). Section 4.2.6.1 subsequently states that "MAP-style programs used elastic net regression (OpenDSM) to model site-level energy consumption" and discusses the strengths and limitations of that modeling approach. [Draft Report, p. 39.] Based on this, Draft Report appears to assume that MAP-style program savings claims were generated using elastic net regression models during the evaluated period. However, elastic net regression analysis was released in OpenDSM in 2025. Earlier vintages of the codebase was a Time-of-Week and Temperature (TOWT) model. It was this TOWT model that PCE used to claim savings for PY 2022-2023. This distinction may be important because it is unclear whether the replicability analysis compared claimed savings against savings produced using the same	Thank you for this clarification regarding model vintages used in the PCE FLEXmarket Program. The replicability analysis compared (1) claimed savings to PA-provided (documented) savings generated using PA models and (2) DNV-replicated results to both documented and claimed values. Replication was conducted using the PA-provided links to code bases and documentation. Regarding model vintages and methodologies: (a) We will incorporate the clarification of vintage and model provided in the comment here. (b) The materials provided to DNV, including links to the OpenDSM codebase and associated resources, directed the evaluation team to implementations consistent with elastic net regression. The materials did not indicate that an earlier model vintage (for example, TOWT) was used to generate the original PY2022–2023 savings claims and did not provide direction to relevant sub-links for the earlier version. As a result, the replication relied on elastic net-based implementations reflected in the provided resources. (c) Given this, differences may exist between the models used to generate the original claims and the models used in

#	Commenter	Comment	DNV response
		model vintage and methodology that generated the initial and true up claims. If different model structures were used, that would suggest the analysis may be evaluating differences between modeling approaches in addition to evaluating the reproducibility of originally claimed savings. This question appears particularly relevant given Recommendation #3, which recommends greater use of TOWT-based approaches. The Draft Report suggests that certain MAP-style programs used elastic net regression models during the evaluated period. If those programs instead relied on earlier TOWT-based models, that context may be important for interpreting both the replicability findings and the discussion supporting Recommendation #3. PCE respectfully requests clarification in the Final Report regarding: (a) The specific model vintages and methodologies used in PCE's MAP-style program during PY2022-2023 to generate the original savings claims; (b) The model vintages used by the evaluator to replicate those claims; and (c) If there are any differences, how those differences between those methodologies were considered when interpreting realization rates, replicability findings, and related conclusions.	replication. To the extent that such differences exist, the replicability analysis may capture both (i) how well the analysis can be reproduced and (ii) differences in model structure. At the same time, replication results were closely aligned with documented values. Across MAP-style cohorts, replicated savings were within 5% of documented savings, and within 10% for PCE's cohort. This level of agreement suggests that, despite potential differences in model vintage, the overall findings are not highly sensitive to the specific model form used for replication. Additionally, elastic net regression represents an evolution of earlier TOWT-based approaches and is intended to provide improved model fit and robustness. In that context, use of the newer model form in replication would not be expected to introduce systematic degradation in results and may, in principle, improve modeled estimates. In response to this comment, we will revise the Final Report to: * Clarify that MAP-style programs during the evaluation period, including PCE's, relied on earlier TOWT-based models for savings claims * Acknowledge that replication relied on the model structures reflected in the provided OpenDSM materials, which did not specify model vintage * Clarify how potential differences in model structure are considered when interpreting replicability findings and Recommendation #3 These updates will ensure that the modeling context is more clearly reflected in both the replicability analysis and supporting discussion.
34	PCE	[Weather Normalization Requirements for Population-Level NMEC Programs] Section 4.2.6.2 states that certain M&V Plans approved for the evaluated population-level NMEC (Pop NMEC) programs normalize to reporting-year temperatures rather than typical meteorological year (TMY) conditions and therefore do not fully satisfy the NMEC Rulebook's normalization requirements.	In a document that explicitly discusses "avoided energy use" as an alternative way of describing savings, we think it is reasonable to understand the term "normalized" to refer to estimates that reflect either normal or typical weather as opposed to actual weather from the performance period. Furthermore, given the explicit instruction to that effect for site-

#	Commenter	Comment	DNV response
		PCE respectfully requests clarification regarding the basis for this conclusion and specifically asks the Final Report to identify the NMEC Rulebook provision or provisions on which this statement relies. The NMEC Rulebook contains explicit weather normalization requirements for site-level NMEC projects. Section III.1.E.2 states that site-level projects must normalize final savings claims using long-term weather conditions and specifies how those adjustments are to be applied to baseline and performance periods. [California Public Utility Commission, NMEC Rulebook Version 2.1 (released September 10, 2025) ("NMEC Rulebook"), p. 20, available at https://www.cpuc.ca.gov/-/media/cpuc-website/divisions/energydivision/documents/energy-efficiency/rolling-portfolio-program-guidance/nmec-rulebook-21-march-2025.pdf The provisions for Pop NMEC programs differ. Section II.2.A.1.b.i requires population-level M&V Plans to describe how analytical methods will address weather normalization and other factors. [NMEC Rulebook, p. 13.] Unlike the site-level NMEC provisions, the Pop NMEC provisions do not appear to contain an explicit requirement that savings be normalized to TMY conditions or long-term weather datasets. PCE requests clarification regarding whether the Draft Report reflects: a) An interpretation that the Rulebook currently requires TMY or long-term weather normalization for all Pop NMEC programs; or b) An evaluator judgment regarding preferred methodological practices. If the latter, PCE recommends that the Final Report clearly distinguish between explicit Rulebook requirements and evaluator recommendations regarding preferred normalization methodologies. Such clarification would improve transparency and help avoid confusion regarding existing Rulebook requirements for Pop NMEC programs.	level projects, and the reference back to site-level as the default in other aspects of the pop NMEC approach, we cannot conclude that pop NMEC follows an alternative approach that would be allowed through omission. Additionally, we see no reasonable argument for AEU as the basis of final savings. The purpose of a TMY-type approach is that the savings are put on terms most likely to reflect the average weather across the EUL. Actual weather could be extreme in either direction, making it less ideal for producing first year savings extended to EUL. Finally, with guardrails in mind, it makes little sense to have two options that will give slightly different savings levels from which an implementer would have to choose. We acknowledge that there are many places where the NMEC rulebook could be either more clear or more explicit. On the other hand, in the context of NMEC in general, and the wider context of billing analysis (UMP chapter 8, etc.), we are aware of no precedents for an AEU approach to savings that will be applied to multi-year EULs. Prior to applying novel thinking of this sort, we recommend a discussion with CPUC.
35	PCE	[Role of Stranded Savings in Pop NMEC Programs] PCE agrees that NMEC can provide unique value by capturing stranded savings opportunities and appreciates the Draft Report's discussion of this benefit. However, PCE is concerned that the report may conflate the ability of NMEC to measure stranded savings with the question of which customers should be eligible to participate in Pop NMEC programs.	The statements in the report are consistent with California policy. NMEC estimates savings relative to an existing conditions baseline and rewards those savings for the full EUL. In the wider context of California EE savings calculations, this is the most generous savings algorithm available, comparable to accelerated replacement savings only if the project was 100% accelerated. In that case, the project would be assumed not to

#	Commenter	Comment	DNV response
		The Draft Report states that, "Pop NMEC should be reserved for customers where the approach adds unique value that addresses stranded savings, rather than absorbing high-risk savings from customers that have better-fit alternatives."[Draft Report, p. 10.] These are related but distinct concepts. The ability to measure stranded savings is one benefit of NMEC, but it does not necessarily follow that stranded savings should be the primary criterion for participant targeting or eligibility. Recent Commission direction has emphasized broader use of meter-based savings methodologies. D.23-06-055 identifies multiple benefits associated with NMEC and other meter-based approaches beyond stranded savings, including improved measurement accuracy, shortened learning cycles, pay-for-performance opportunities, and support for market access programs (MAP). [D.23-06-055, pp. 41-42.] The Commission also directs broader use of NMEC, randomized control trials, strategic energy management (SEM), and other meter-based savings methodologies for qualifying retrofit programs. [D.23-06-055, pp. 41-42; OP 20.] PCE recommends that the Final Report clarify whether this discussion is intended as a conclusion drawn from the evaluated program results and findings, or a broader policy recommendation regarding the future role of Pop NMEC programs.	have happened during the full EUL of the measure, which is effectively the concept of stranded savings. Concern on this issue is evident in various places in Commission commentary, and considering stranded savings is the most direct way to make NMEC consistent with the longstanding California net savings framework. CPUC is open to logic that could support existing conditions savings for the duration of the EUL that does not overturn longstanding and foundational measure application type logic.
36	PCE	[Interpretation of Net-to-Gross Ratio Differences Across MAP-Style Programs] The Draft Report identifies BayREN's PY2023 MAP as having the highest net-to-gross ratio (NTGR) and PCE's PY 2023 as having the lowest NTGR among MAP-style programs. [Draft Report, p. 46.] The Draft Report further attributes observed differences, at least in part, to customer targeting and participant reported program influence. PCE appreciates the effort to identify factors that may contribute to NTGR variation across programs. However, additional context may be helpful when interpreting cross-program differences. Customer targeting may be one factor influencing participant attribution responses. Other factors may also contribute, including differences in participant characteristics, program objectives, incentive structures, eligible measures, budget constraints, and other program-specific design features.	Noted, regarding cross program observations. Regarding the implication of a comparison group for net savings, it is outside the scope of this evaluation to re-examine this issue. It is widely acknowledged that the inclusion of a comparison group, at the aggregate level, likely leads to a partial net estimate, that--with self-selection, timing, and applicability issues--is expected to be extremely partial as well as unmeasurable. The accepted approach is to apply a standard self-report NTG that, in this case, is consistent with all other programs. Also, please see our response to comment #35 for the explanation how NMEC projects likely receive over-estimated gross savings to the extent that they are not stranded savings. The potential magnitude of up-to-code savings multiplied by the

#	Commenter	Comment	DNV response
		Importantly, different Program Administrators operate under different Commission-authorized frameworks and objectives. RENs and CCAs are subject to different policy requirements and program design considerations, including different cost-effectiveness expectations and objectives. In addition, RENs are specifically intended to serve hard-to-reach (HTR) customers, which may result in participant populations that differ from those served by other Program Administrators. These differences may influence program design choices, customer eligibility and targeting, incentive levels, and measure offerings, which in turn may affect participant perceptions of program influence. For example, the Draft Report includes participant feedback expressing a desire for larger incentives and broader measure eligibility. Such feedback may reflect program design constraints rather than participant views regarding overall program effectiveness. Additional discussion of these differences may help provide context for interpreting participant attribution responses across programs. PCE also notes that FLEXmarket's approved M&V Plan incorporates a comparison group adjustment as part of its savings methodology. Specifically, the M&V Plan states that comparison groups are maintained for gross savings adjustments and that the program's incremental impact is calculated relative to a nonparticipant population. [PCE FLEXmarket Program Implementation Plan, pp. 35-36, (available on CEDARS).] This context may be relevant when interpreting NTGR results. The comparison group adjustment is intended to refine the program counterfactual by accounting for changes in energy consumption observed among comparable nonparticipants. These include exogenous factors that may influence metered consumption independent of program participation. Comparison group adjustments and participant self-report NTG surveys both seek to distinguish impacts attributable to the program from those that would have occurred absent the program. As a result, there is a possibility that some factors affecting attribution are reflected in both adjustments. PCE therefore recommends that the Final Report acknowledge this interaction when interpreting evaluated savings and comparing NTGR results across MAP-style programs. PCE also notes that Table 4-10 indicates that the PY2023 PCE NTGR estimate is based on one interviewed customer representing three projects, while the BayREN 2023 estimate is based on fourteen interviewed customers	years not actually accelerated will assuredly dwarf any partial net aspect brought on by the comparison group.

#	Commenter	Comment	DNV response
		representing fifteen projects. [Draft Report, p. 46.] PCE appreciates that the report acknowledges that small sample sizes and variance increase uncertainty in specific-year estimates. Given these differences in the underlying samples, additional context regarding the limitations of cross-program comparisons may be helpful.	
37	PG&E	[Section 4.8.3.2 "The dramatic overstatement of preliminary savings serves no one and reflects lax program management by PG&E....PG&E manages the OBF program and access to its financing, which gives it a strong position to address ongoing program issues. However, there is limited evidence that this position has been used to address persistent challenges."] PG&E respectfully disagrees that we have "lax program management" and that we have not done anything to address persistent challenges with the new and evolving Population NMEC programs. We hope the following comments help the evaluation team better understand the work PG&E is doing to improve the accuracy of claimed estimates and to address the balance between the programs being accessible to customers and being accurate in savings estimates. OBF project savings are estimated by the project developer, and reviewed by an independent engineering QA. Those savings are used as initial estimates. PG&E assesses the saving of the projects on a population basis. Non-routine events are not collected for the projects and therefore savings results are not adjusted for changes in customer activity. For example, a real estate customer premise results could be impacted by tenancy. Customers with high energy use can also have a material impact on portfolio results. PG&E has taken steps to reduce the size of customers who may participate in the Pop-NMEC portfolio so no individual customers results can materially impact the portfolio as a whole. However, there still remains the risk that NREs with customers, or changes in customer usage patterns are not accounted for using the PopNMEC approach. These risks to claimable savings are understood and accepted for projects supported with OBF loans given its primary goal of market support rather than a resource program with a primary goal of maximizing claimable TSB. Being a market support program, the OBF portfolio is not optimized for TSB claims. There is not a singular implementer coordinating a coherent portfolio of projects to deliver TSB, and projects are not selected for their ability to consistently deliver claimable TSB. The program is broadly open to the market	The full statement from our report is, "The dramatic overstatement of preliminary savings serves no one and reflects lax program management by PG&E." This criticism is limited to this specific area. Elsewhere we acknowledge the efforts being made by PG&E to improve the accuracy of many other aspects of the process. The overstatement of preliminary savings has been consistent for OBF and in direct contrast to the preliminary savings estimates for other programs. Explaining this as a result of the inaccuracy of the NMEC process in general (not controlling for NREs, etc.) does not seem to us like a fruitful line of reasoning. To our mind, the likely explanation is more obvious. Implementers for all other programs have a clear monetary incentive to not over-estimate savings. If they do so, there may be the necessity of clawing back incentives due to the initial over-estimation. OBF implementers do not face this negative incentive as high preliminary savings estimates have the sole effect of shortening the payback period. This occurs because there is no direct connection between trued up savings and the final loan terms, an essential part of NMEC. PG&E also benefits from the over-estimate of preliminary savings as is reduces the value of the free interest provided to the end customer by PG&E.

#	Commenter	Comment	DNV response
		with no specific customer or project types selected to maximize results. Some attrition in savings is therefore expected and accepted given the market support goals of the program and the incomplete information through not collecting NREs. PG&E, however, is seeking to improve realization of the portfolio and has successfully introduced a number of controls to improve realization. These controls include ensuring that customers are a good fit for NMEC analysis both in their usage profiles and the impact of on-site generation (e.g. Solar), and that their overall usage isn't individually too high to be included in the portfolio (those large customers are directed to use Site-NMEC). As a result, PG&E's realization rate of savings estimates to metered-savings true-ups are as follows; 29% for PY2022, 37% for PY2023, 45% for PY2024, and 52% for PY2025 measured to date (PY2025 M&V is in progress until the end of calendar year 2026). PG&E has also developed monitoring dashboards, whereby poor performing projects can be highlighted to Trade Professionals for investigation. These controls have improved the results, however, there is a trade-off of these controls resulting in the program becoming less accessible to customers, and no longer being as broadly available. PG&E has also addressed another issue that was impacting overall realization rate. The program year realization rates include projects with installation complete dates during that calendar year. Previously, PG&E's estimated claims in the evaluated years were based on loan check issued dates rather than installation complete dates during the metered savings true up. Comparing the estimated claim to metered savings true-up amount resulted in a larger adjustment because more projects were included in the estimate claim that could not be included on the installation date basis for the final results. PG&E has corrected this going forward, and will report savings estimates on an installation date basis rather than loan check issued basis, which is typically 1-2 months later. Finally, for 2025 PG&E is seeking to improve reporting accuracy by making adjustments to the initial estimated claim based on the early portfolio results. For example, in 2025 PG&E's early results for the portfolio showed a 52% realization rate in observed savings using the PopNMEC methodology (which doesn't account for NREs). Therefore, PG&E adjusted the initial claim by 52%. This should reduce the scale of the impact of true-ups in 2026.	

#	Commenter	Comment	DNV response
38	PG&E	[Section 4.8.3.2 "OBF could develop a two-stage process, updating the initial financing agreement with performance-based savings. Customers would then pay for their projects at a rate consistent with their meter-based savings. This change would force all parties to operate in the real world of meter-based savings, ensuring risk no longer lands solely on the customer. PG&E would have much greater motivation to manage the program carefully because poor performing projects would tie up funds for longer periods. Contractors would be motivated to identify good participants and serve them well because their activities will be tracked with respect to meter-based performance. Ultimately, the only way to fully justify OBF's presence in pop NMEC is to make it a performance-based program. The details of this transition will be complicated, but this improvement, if pursued thoughtfully and incrementally, should improve the program's service to both customer and ratepayer without meaningfully undermining the current program's stronger points."] OBF is a market support program, designed to provide customers access to capital for projects they are seeking to install but are not able to move forward on due to lack of capital. PG&E's focus of the program is to provide a fast and efficient process for providing capital for those projects. Delays in issuing checks following installation would create challenges for customers and project developers, as unlike incentive programs, OBF provides the full amount of funds for the project. Failure to fund a project soon after installation can cause cashflow challenges for contractors developing the projects. PG&E's OBF program is designed as a loan program, where PG&E provides loans to customers for projects developed by their contractor (PG&E does not normally have a contractual relationship with the contractor, unless it is a PG&E program implementer). This provides fewer guarantees that savings will materialize, but provides greater access to capital for customers in need. The OBF loan agreement is explicit to the customer that savings estimates are not guaranteed and that those savings estimates are provided by their contractor to PG&E. PG&E does require an independent engineering review of those savings estimates as a control to ensure that the savings estimates are reasonably prepared, as that would result in delays to project approval and increases in project cost. Adjusting the OBF program to a savings guarantee model would be a fundamental change to the program that could not be affordably managed as a	Noted. Better aligning OBF with NMEC principles, determining whether the program remains consistent with Advice Letter, PG&E 3697-G /4812-E, 3697-G-A/4812-E-A (https://www.pge.com/tariffs/assets/pdf/adviceletter/ELEC_4812-E-A.pdf), and/or undertaking a holistic review of the strengths and challenges of the OBF program are important and potentially worthwhile tasks, but they will require a more expansive process than is possible within this evaluation.

#	Commenter	Comment	DNV response
		program designed as a market support program, and would result in fewer customers being able to access capital for their projects. Small customers would be particularly impacted as requiring Site NMEC level analysis with NREs collected would increase project costs to a level that would not be feasible. For large customers there are financing products provided by ESCOs that provide this kind of savings guarantee model. They have high project developer involvement following installation, and require large margins to ensure that those overheads are covered. Individual site uncertainty is high in PopNMEC programs, generally over 100%. Savings uncertainty levels are only reasonable for determining measure impacts when aggregated to the program-year level. The 2024 OBF cohort installed 400 projects with metered savings realization rates ranging from -1000% to 1700%. 42 sites measured negative savings and 12 sites measured over realization rates greater than 200%. Renegotiating loan terms that speed up a customer's repayment schedule or extending repayment schedules beyond 10 years would not improve program realization rates. PG&E has been tracking individual trade pro performance and providing targeted feedback since 2024. PG&E launched a performance management OBF dashboard at the end of 2025 and has been providing trade-pros site level feedback periodically.	
39	PG&E	[Section 4.8.3.2 "Ultimately, the only way to fully justify OBF's presence in pop NMEC is to make it a performance-based program. The details of this transition will be complicated, but this improvement, if pursued thoughtfully and incrementally, should improve the program's service to both customer and ratepayer without meaningfully undermining the current program's stronger points."] PG&E chose PopNMEC to report savings for OBF projects as the reportable savings analysis is performed after installation, and does not slow down the project or the loan, and reduces customers' barriers to participation. PG&E accepts that PopNMEC does not provide sufficient information to properly analyze project performance at any individual site, as to do that requires non-routine event information. However, this approach best supported the objectives of the program, despite not optimizing claimable savings.	Noted. See response to comment #38.

#	Commenter	Comment	DNV response
		The rulebook states "Ideally, 100% of total PA payments for each population-level program should be made based on payable savings determinations made using NMEC methods. At a minimum, 50% of total PA program payments for each Population-level NMEC Program (not including PA administrative or PA measurement and verification costs) must be based on payable savings determinations made using Population-level NMEC methods." PG&E agrees with this approach for incentive based programs where performance is incentivized. OBF is a loan program where the customer repays the loan in full. The OBF program is entirely PA administrative and M&V cost. Historically loans have not been viewed as program payments since they are repaid by the customer. Renegotiating loan terms based on a metered savings is not a practicable to operate a loan program. Additionally, the large majority of loans issued in OBF are from Trade Allies which do not hold a contract with PG&E and not subject to pay for performance compensation. Their agreement and relationship is with the customer receiving the loan.	
40	PG&E	[Section 5.2 "PG&E should offer a plan that achieves program-level forecast realization rates (initial claimed savings relative to verified savings) greater than 90% by the end of July 2026. This approach will better align program incentives with realized performance, address the current misalignment between initial claimed savings and actual outcomes, and align payment schedules with actual bill-level savings."] While this may make sense for site-level NMEC projects, it is not feasible for population-level NMEC programs. There are persistent challenges with site NREs and customer load growth that limit the programs ability to achieve 90% overall. 10% of sites in the 2024 cohort measured negative savings overall which points to customer changes rather than estimate inaccuracy being a large driver of the low realization rate. We have made program improvements that resulted in an increase in realization rate from 29% in 2022 to 52% in 2025. Additionally, we note that the recommended 90% forecast realization rate threshold does not appear to be grounded in existing CPUC policy or the NMEC Rulebook. While PG&E remains committed to continuously improving our savings estimation accuracy—as demonstrated by our progress from 29% to 52% realization rates—we would welcome a collaborative discussion on appropriate benchmarks that account for the inherent characteristics of population-level NMEC programs. We believe that program design decisions,	Noted. See response to comment #38.

#	Commenter	Comment	DNV response
		including forecast accuracy targets, should remain within PG&E's purview as the program administrator, provided that final true-up claims accurately reflect verified savings. We look forward to working with the evaluation team to establish realistic and mutually agreed-upon expectations that support both program accessibility and accountability.	
41	SoCalgas	[Section 1.5 – Key Findings and Recommendations] SoCalGas requests DNV to provide specific guidance on how savings measurability can be improved, with respect to gas projects, to help establish a framework that helps the design and scalability of gas population NMEC projects. Some areas of interest include modeling methods, goodness of fit requirements, accuracy requirements, approach to determining exogenous effects, population requirements, and comparison group design.	This is outside of the scope of this evaluation report but definitely an important topic to be pursued.
42	SoCalgas	[Recommendation 7 - Broader Technologies for Greater Whole Building Upgrades] SoCalGas requests further insight into gas eligibility requirements as they relate to under-served, small, and medium businesses sectors, where the savings potential and the population is likely abundant. Specific recommendation on increasing gas project eligibility can help scale future gas Population NMEC projects.	Please see response to comment #41.
43	Mendota	[Overarching] We appreciate DNV's in-depth analysis of the PY 2022-2024 Population-Based NMEC programs and thoughtful consideration of ways that these programs can be improved. Although we disagree with some observations, we very much appreciate the methodical approach DNV took to assess program performance and approaches. This will help to further improve application of this novel approach in helping California achieve its energy efficiency, carbon reduction, and bill savings goals.	Thank you.
44	Mendota	[Overarching] In several parts of the study, DNV implies that population-level NMEC programs should play a limited role in the portfolio of resource acquisition energy efficiency and fuel substitution programs. This implication runs counter to CPUC and legislative intent. D.23-06-055, OP 20 states, "For programs that meet all of the below characteristics, portfolio administrators shall now be	The highlighted decision reflects an extreme position not consistent with subsequent decisions. It also does not recognize that a default existing conditions baseline undermines much of the existing net savings framework in California. These issues should be addressed in a separate process to solidify the definition of NMEC.

#	Commenter	Comment	DNV response
		required to use, as described in Section 6.2 of this decision, normalized metered energy consumption (NMEC), randomized control trials, strategic energy management, or another meter-based savings evaluation method, or otherwise the justify in the implementation plan why meter-based methods are not used for feasibility or cost-effectiveness reasons." Further to this point, D. 17-11-006 emphasizes in its Conclusions of Law 2, "AB 802, approved soon after the Commission approved the IOUs' To-Code Pilots, effectively obviated the need to address the first research objective of the To-Code Pilots, i.e., to understand the degree to which there is below-code equipment that is not getting replaced quickly enough through natural turnover or existing programs." The Decision terminated the to-code pilots because AB 802 "required the Commission to make the following changes to energy efficiency program policies starting September 1, 2016: · use 'existing conditions' as the default baseline for determining energy efficiency savings, and · provide incentives for measures that bring buildings into compliance with code, but do not necessarily exceed code (D. 17-11-006, p. 7). We believe strongly that population-level NMEC should be used as a fundamental tool in California's efforts to build a more energy efficient, low carbon future and only be limited by cases where statistical validity does not support its use. California energy efficiency resource impact programs face considerable challenges, due to minimal financial incentives, advances in code, and burdensome participation requirements. Because pop-NMEC programs can provide meaningful incentives, provide opportunities for to and above-code projects, and can and should have streamlined participation requirements because they rely on metered savings, they should not be characterized as niche solutions.	
45	Mendota	[Recommendation 1; Section 5.1. Program savings accuracy can be improved with more homogenous populations.] The study cites the NMEC Rulebook's references to Permissible Project Types for Population-Level sites; we disagree with this provision and believe the Rulebook's wording leads to misinterpretation. It is both unreasonable to expect and we do not see in practice that Aggregators will identify, recruit, and install a large number of similar projects for similar customers, within the same quarter. In our view, participating sites need not be homogeneous. Rather, our	The NMEC rulebook verbiage on homogenous populations is consistent with decades of experience with consumption data analysis. The interpretation is quite straightforward. We will continue to work with stakeholders to figure out whether these methods can be adapted to support a program that appropriately addresses all risks, including those to the ratepayer.

#	Commenter	Comment	DNV response
		interpretation of this rule is that energy use patterns at non-participating sites that comprise comparison groups must be sufficiently similar (as determined using reasonable methods of statistical validation) to those of participating sites to conclude that the non-participating sites have similar drivers and levels of energy consumption as the participating sites. One of the core concepts of pop-level NMEC is not that it is necessary to group similar projects/participating sites in the analysis but, rather, that it is key to develop and apply a uniform methodology to all participating sites (as individuals) that meet eligibility criteria.	
46	Mendota	[Recommendation 2; Section 5.1.1 Clear, consistent, and transparent documentation.] The study states that evaluators had a difficult time collecting project level documentation and that PAs were inconsistent in their approaches to documentation. We agree that project level documentation should be maintained and easily accessible by evaluators. We ask, however, that DNV specify what information is and what information is not important to the evaluation process and that DNV recommend that PAs treat project review and required documentation consistently. For instance, if maintaining a record of CET inputs is important, then that requirement should be explicitly stated. In addition, if requirements generally imposed on Custom projects (e.g., detailed savings estimation calcs, photos, spec sheets, etc.) are not necessary for pop-NMEC evaluations, then these items should be explicitly listed as Not Necessary so that PAs do not require program implementers and participants to collect and archive such data. We note that some PAs, out of an abundance of caution, have very heavy documentation requirements and that these requirements both increase implementation costs and are significant deterrents to robust participation in pop-NMEC programs.	Our point regarding inconsistent documentation primarily relates to the availability of project-level records that clearly and completely map to aggregate program-cohort claims in CEDARS. This includes consistent identifiers (for example, project, customer, and claim IDs), installed measure information, EULs, and project-level savings values that can be fully reconciled to claimed totals. As noted in the report, pop NMEC savings are claimed at the aggregate cohort level rather than at the individual project level. As a result, verification of claimed savings depends on PAs maintaining project-level datasets that support a clear, traceable linkage to the aggregate claims. The key requirement is that evaluators can reliably connect individual projects to the final claimed savings without ambiguity or extensive follow-ups with PAs. With respect to the types of documentation needed: * Important for evaluation are complete and well-organized datasets that identify the specific projects and associated project-level information underlying each aggregate claimed and documented savings value and that allow those relationships to be verified with 100% accuracy. * Not generally required for pop NMEC evaluation purposes are the types of detailed engineering documentation typical of custom projects (for example, detailed calculation workpapers, extensive photo documentation, or equipment specification sheets).

#	Commenter	Comment	DNV response
			Our intent is not to require extensive documentation, but to ensure that records clearly connect project-level data to total program savings. Consistent with our recommendation, PAs should propose a solution to the reporting PCG that defines a customer project-level pop NMEC dataset that links with 100% accuracy to both initial and true-up aggregate CEDARS claims.
47	Mendota	[Section 5.1.1 Eligibility rigor and auditable records.] Although we agree that upfront screening rules should exclude sites with known process-, event-, or production-driven loads (which should be evident from the site's CV(RMSE)), we do not believe that sites above a certain size should be prohibited from participating. Rather, rules should indicate that sites that exceed certain size thresholds (we use 1 MW of peak demand) require additional screening. Additional screening would include determining whether there is a valid comparison group with which to normalize savings. An arbitrary size limit, in and of itself, should not be a disqualifier if the site can, otherwise, meet statistical requirements that allow for accurate savings measurement.	Noted. Please see responses to comments #1, #5, and #12.
48	Mendota	[Section 5.1.2 The use of different methods to calculate savings increased variability and risk in savings results.] Although we agree that implementers employ different methods of calculating savings and that final claimed savings may reflect methodological choices, we disagree with the conclusion that this dictates need for a single method of calculating savings. NMEC is still relatively new and limiting the methods by which savings are calculated stifles innovation. In particular, methods by which comparison groups are formed are likely to continue to evolve, which is beneficial to advancing the concept and improving measurement accuracy. The Rulebook already requires that the savings calculation methodology is documented in the M&V plan, which PAs are expected to review and approve. PAs may choose to engage working groups or subject matter experts to assist in the methods review, ask for clarification, and ensure that the proposed methods are reasonable before approving. This process should be sufficient to ensure that implementers are following industry best practices, while leaving room for innovation via the PAs' discretion. We also take issue with extrapolating this observation to conclude that this "reduces the confidence of stakeholders, including participating customers and	The concern is noted. With respect to standardization, the NMEC rulebook has generally been interpreted as allowing methodological flexibility. This has been appropriate given the early stage of NMEC, as it has allowed methods to be applied and demonstrate their value. However, this interpretation reflects the absence of explicit guidance rather than a clear intent to maintain methodological diversity. At the same time, the rulebook is explicit that methods must be standardized within a given pop NMEC program. This raises the question of whether greater consistency across programs could similarly provide benefits, such as improved clarity and predictability for customers. Setting aside concerns about innovation, it is not clear what inherent value methodological variability provides to NMEC customers or to the program track as a whole. The intent of this finding is to prompt consideration of these questions and whether a more standardized approach could strengthen the program.

#	Commenter	Comment	DNV response
		contractors, in cross-program comparisons" and that this "runs counter to pop NMEC's standardization intent". In our experience, stakeholders, contractors and customers are less concerned about cross-program comparisons and more concerned about the approaches different Program Administrators take to administering (regulating) pop-NMEC programs. We also disagree that there is a "pop NMEC standardization intent" and respectfully request that DNV provide a citation for this observation.	
49	Mendota	[Recommendation 7; Section 5.2 Broader technologies for greater whole-building upgrades.] We agree with this recommendation; however, the studies implies that the PAs or Implementers are pursuing primarily lighting projects and not delivering comprehensive projects. This implication ignores a primary design element of existing pop-NMEC programs which is that they are market driven through aggregators. More specifically, the market, not PAs or Implementers are determining which projects to pursue. If the CPUC wishes to see more comprehensive projects, the sole success metric should be TSB and the use of TRC or perhaps any cost effective metric should be removed.	Noted. As specified in response to comments #16 and #30 and in our report, the recommendation to increase the share of mixed-measure or more comprehensive projects is intended to improve alignment with NMEC policy objectives, rather than as a prescriptive requirement. This shift is also consistent with CPUC's broader focus on EE portfolios with greater TSB values, which are more likely to result from multi-measure than lighting-only installations.
50	Mendota	[Section 5.1 Approaches to modeling energy use at the site level varied in their alignment with NMEC rulebook expectations.] The study states that, "only one program normalized savings to typical weather conditions, consistent with NMEC rulebook intent. Other programs used reporting-year weather, which increases savings errors in multiple ways." Please replace "rulebook expectations" with an actual citation from the Rulebook. The Rulebook requires typical weather conditions for site-level NMEC but is flexible on pop-level, only requiring that the weather normalization method be specified in the M&V plan. If DNV is recommending a change to pop-level NMEC rules, please state this explicitly.	Please see response to Comment #34.
51	Mendota	[Section 1.4.2; Section 4.6 Total System Benefit.] The study recommends that, "all projects … should use the same avoided cost year for both the initial and true-up claims". We agree with this recommendation and suggest that guidance go further in specifically stating that projects shall use the applicable ACC and CET Quarter corresponding to the installation completion date for TSB calculations. Subsequent true-up	Additional guidance may be needed as PAs work to use the same avoided cost year for both the initial and true-up claims. Given that this is a new finding as we evaluated trued-up TSB claims for the first time, we did not want to be overly prescriptive.

#	Commenter	Comment	DNV response
		claims would also use claim savings in the quarter applicable to the period of measurement. We also have found that some PAs require that implementers use different avoided cost years for initial and true-up claims. Although we have not yet achieved a full year of performance and, therefore, this approach can change, it points up an issue that we view as more critical than consistency in savings calculation approaches (our comment 4) and that is the different way that PAs are administering population-level NMEC programs. As we understand it, based on input from participating aggregators, the PAs differ significantly in: the amount of rigor they apply to review of initial project savings estimates and the required project documentation from the implementer and program participants. Our experience working with the Summer Reliability Market Access programs and current market access programs has demonstrated that several PAs view the new population-level NMEC programs similar to custom programs, greatly hindering customer/aggregator participation and substantially extending project timelines. We recommend that DNV's future studies review the different ways in which individual PAs administer population-level NMEC programs with an aim to develop consistency in administration of these programs. We believe this type of consistency, especially if it reduced participation barriers, would be heartily welcomed by stakeholders, contractors and customers.	
52	Mendota	[Section 1.4.2; Section 4.6 Consistent use of the initial claim impact profile.] The study states that it is important for initial and true-up claims to use consistent energy savings profiles and indicates that, "Pop NMEC programs will transition to empirical cohort-specific, meter-based hourly impact shapes." We support transitioning to "empirical cohort-specific, meter-based hourly impact shapes" (custom impact profiles), but it would be helpful for DNV to cite the source for the statement. In addition, since DNV points out the importance of maintaining consistent energy savings profiles for initial and true-up claims, it would be helpful if DNV described how implementers (and PAs) would have the necessary information to use consistent custom impact profiles for initial and true-up claims. Initial claims would, presumably, have to use "best guess" impact shapes while the final true-up claims would use actual hourly impact shapes.	This comment addresses two distinct issues: transitioning to empirical shapes and using consistent profiles. The transition to empirical cohort-specific, meter-based shapes. The statement regarding a transition to empirical cohort-specific, meter-based hourly impact shapes reflects an expectation based on ongoing methodological and infrastructure developments, rather than a formal adopted requirement. This expectation is tied to the development of the CPUC impact profile library and its integration with the Cost-Effectiveness Tool (CET), which enable the use of site- and population-level impact shapes with associated avoided cost values in place of legacy DEER load shapes. As this infrastructure becomes operational for program years 2026–2027, it creates a pathway for implementers and program

#	Commenter	Comment	DNV response
			administrators to develop, store, and reuse cohort-specific impact profiles for both initial and true-up claims. Using consistent energy savings profiles. This point is related to the load shapes used in PY2022 and PY2023 and the way in which true-ups are made. Currently, true-ups are specific to the energy savings, meaning that if a cohort saves more or less than expected a positive or negative kWh, kW, and therm value is entered into the CET. This produces a positive or negative TSB delta indirectly, meaning that the TSB delta is calculated solely based on the true-up information entered. Consequently, if elements beyond energy savings that impact TSB are also changed, such as the energy savings profiles, this will produce a different outcome than if TSB were calculated directly as the difference between the planned TSB (initial claim) and the actual TSB for the project's total savings. The issues with changing impact shapes when transitioning to empirical shapes will need to be addressed so that the final TSB results are accurate.
53	Mendota	[Overarching Meter data.] Our programs have experienced issues with poor quality meter data (e.g., high numbers of estimated reads), which has forced us to reject viable projects. We recommend that future studies assess the degree to which the availability of high quality meter data may be negatively impacting projects.	Noted. That was not something we observed in this study.

About DNV