Demand Side Analytics
DATA DRIVEN RESEARCH AND INSIGHTS

FINAL EVALUATION PLAN

PY2025 Statewide Critical Peak Pricing Impact Evaluation

Prepared for: **Pacific Gas & Electric**
Southern California Edison
San Diego Gas & Electric

Prepared by: **Alana Lemarchand**
Tim Larsen
Demand Side Analytics

November 17, 2025

TABLE OF CONTENTS

1 INTRODUCTION

This evaluation plan lays out the requirements and analysis approach to evaluate load impacts for the PY2025 California Statewide Critical Peak Pricing (CPP) rates. The plan outlines out a common framework to evaluate CPP impacts for each of PG&E, SCE, and SDG&E (hereafter the "Joint Utilities").

The aim of the evaluation is to measure CPP event-day impacts for each IOU by customer size:

- Small (under 20kW maximum demand—PG&E and SCE only)
- Medium (20 kW to 200 kW maximum demand)
- Large (200 kW and above)

This consists of estimating hourly ex post load impacts for PY 2025 and ex ante load impacts through 2036. SDG&E's Small CPP rates are evaluated separately and as such not included here. SDG&E also had no CPP events in 2025, so its customers will only be included in the ex ante portion of the evaluation.

There are two main objectives for this evaluation plan. The primary objective is to engage in science and avoid after-the-fact analysis, where there is a temptation to modify models to find the desired results. This requires:

- Specifying the intervention and documenting the hypothesis
- Establishing the sample size and the ability to detect a meaningful effect
- Identifying the data that will be collected and analyzed
- Identifying the outcomes that will be analyzed and segments of interest, and
- Documenting in advance the statistical techniques and models that will be used to estimate energy savings and demand reductions.

The goal is to leave little to no ambiguity regarding what data will be collected or how the data will be analyzed. The secondary objective is to comply with the California Load Impact Evaluation Planning Protocols.[1]

While the Joint Utilities' CPP rates have many common features—essentially adders on event days with lower pricing on non-event days—the structure and related program provisions vary by utility. This plan also seeks to outline any different treatment of the rates and questions of interest by IOU.

[1] The full set of load impact protocols can be found here: https://www.cpuc.ca.gov/industries-and-topics/electrical-energy/electric-costs/demand-response-dr/demand-response-load-impact-protocols, with additional updates here: https://docs.cpuc.ca.gov/PublishedDocs/Published/G000/M549/K296/549296803.PDF

1.1 SUMMARY OF CPP RATES BY IOU

Each utility offers a relatively similar CPP rate program, with the largest difference being the events, which are called separately by each. As such the evaluations will follow a very similar structure, but with unique considerations for each based on the events and any nuances in program rules. A brief summary of the CPP rate programs by utility is shown below:

Table 1: Summary of CPP Rate Programs by Utility

Utility/ Program	PG&E	SCE	SDG&E
Marketed as	Peak Day Pricing (PDP)	CPP	CPP
Peak Window	4-9 p.m. year round	4-9 p.m. year round	4-9 p.m. year round
Resource Adequacy (RA) Window	5-10 p.m. Nov. – May 4-9 p.m. Jun. – Oct.	5-10 p.m. Nov. – May 4-9 p.m. Jun. – Oct.	5-10 p.m. Nov. – May 4-9 p.m. Jun. – Oct.
Number of Events (2025)	9	12	0
Min/max possible Events	Min. 9, Max. 15	Min. 12, Max. 15	Max. 18 (no Min.)
Event Triggers	Day ahead with high temps, high demand, or short supply	Forecasted system emergencies or extreme weather conditions, day-ahead prices, or CAISO Energy Emergency Alerts	Day-ahead system load forecast > 4,000 MW (Can also be triggered for high temp.'s, extreme conditions, emergencies)
Default rate for C&I customers (bundled)?	Yes	Yes	Yes
Opt out available?	Yes	Yes	Yes
CCAs included?	No	No	No
Ag. Included?	Yes	Yes	Yes
Customers eligible for AutoDR programs?	Yes	Yes	Yes
Other ineligible categories	Other energy incentives, energy reduction, peak hour or direct bidding programs	Direct Access (DA) customers	Direct Access (DA) customers
Incentive	Lower energy rates (per kWh) during other summer peak hours	Summer bill credits (fixed amount)	Lower energy rates (per kWh) during other summer peak hours (demand charges vary)
Bill Protection	Yes, for first year	Yes, for first year	Yes, for first year

Utility/ Program	PG&E	SCE	SDG&E
Loads for Impact Evaluation	Delivered loads	Net loads	Net loads

1.2 CALIFORNIA LOAD IMPACT PROTOCOLS

The California Load Impact Protocols require that for every demand response program and dynamic evaluation, an evaluation plan be produced that establishes a budget, a schedule, and a preliminary approach to meeting the evaluation and reporting requirements.[2] The evaluation plan should also develop an approach to determine what additional requirements, if any, will be met in order to address needs that may arise for long-term resource planning or other applications, such as customer settlement or CAISO operations.

At a high level, the requirements for a load impact evaluation are to provide:

- Impact estimates for each of the 24 hours on various event-day types for event-based resource options (and other day types for non-event based resources)
- Uncertainty-adjusted impacts, reported for the 5th, 50th, and 95th percentiles, reflecting the uncertainty associated with the precision of the model parameters and potentially reflecting uncertainty in key drivers of demand response, such as weather
- Outputs that utilize a common format for ex post evaluation (a slightly different reporting format is required for ex ante estimation)
- Ex ante estimates for each day type
- Various statistical measures so that reviewers can assess the accuracy, precision and other relevant characteristics of the impact estimates
- Ex ante estimates that utilize all relevant information from ex post evaluations whenever possible, even if it means relying on studies from other utilities or jurisdictions
- Detailed reports that document the evaluation objectives, impact estimates, methodology, and recommendations for future evaluations

1.3 SUMMARY OF STUDY DESIGN AND EVALUATION CRITERIA

Table 2 lists the study design question in the California Load Impact Protocols and details how the evaluation plan addresses each study design issue for Small, Medium and Large CPP participants:

[2] TOU rates are considered to be non-event-based resource. If the TOU rates are already embedded in the customer class, it may not be possible to evaluate load impacts.

Table 2: Study Design Questions

#	Study Design Question	Small Commercial	Medium Commercial	Large Commercial
1	Will the evaluation rely on a control group? If so, how will it be developed and what comparisons between the treatment and control group will be made?	Yes. We will identify matched controls for all sites. A nonparticipant with similar pre-treatment usage patterns will be matched to each participant for comparison. For participants without strong matches, we will fit an individual regression model with control group usage included on the right-hand side.		
2	Will the evaluation rely on data from non-event days to establish a baseline?	Yes. 2025 event days will be compared to 2025 non-event days in all models.		
3	Will the study rely on a sample or include the full population of participants? If a sample is used, does it meet 90/10 precision requirements?	Sample Yes	Likely full population, sample if necessary	Full Population
4	Is the study designed to detect a specific effect size? And, if so, how was statistical power assessed?	No. We anticipate sufficient precision from sample sizes at least as large as in previous evaluations.		N/A – We will analyze full population.
5	What is the study's threshold for statistical significance?	90% confidence using a two-tailed test		
6	What is the size of the control and treatment groups, if applicable?	Treatment groups = at least as large as in previous Small CPP evaluations at SDG&E Control = same size		Treatment group = all participants Control = same size
7	How will the evaluation address outliers?	Individual customer regressions will be used for sites with outlier loads not meeting the criteria in #1.		
8	How will the evaluation address attrition?	Ex post impacts are estimated for all customers on CPP rates as of the event day. Ex ante will incorporate any information about changes in enrollments over time.		
9	How will standard errors be calculated?	Matched-Control Diff-in-Diff: Standard errors produced by difference-in-differences Individual site regressions: Robust standard errors from regressions		
11	Will energy savings be estimated?	No		
12	Will overlap with energy efficiency programs be estimated?	No		

2 GENERAL APPROACH AND METHODS

The primary goal of any load impact evaluation is to answer two key questions:

1. What were the ex post load impacts in the current evaluation period?
2. What are the program's estimated load impacts going forward?

This second question is of particular importance as it can be leveraged for long term resource planning and DR capacity for resource adequacy.

In this document, we focus on developing a plan to produce unbiased ex post estimates, with these estimates then fed into a robust ex ante estimation process. Key issues to be addressed in developing the ex post and ex ante impacts are summarized in Table 3:

Table 3: General Considerations for CPP Load Impact Evaluations

Evaluation Consideration	Framework
Will both ex post and ex ante impacts be produced?	Yes
What, if any, changes are expected over the forecast horizon to either the program or participant characteristics? Should these be incorporated into ex ante estimates?	Each utility's program staff will provide a summary of expected program changes, which will be incorporated into the analysis. Each utility's program staff are responsible for developing an ex ante enrollment forecast, including assumptions that account for any such changes.
Will impact persistence be explicitly incorporated into the analysis?	Program impacts can be compared to impacts from previous years and assessed for changes, but they have been generally stable over time so a formal persistence analysis is not planned.
Is M&V activity needed to address the issue of persistence or of program changes?	As impact evaluations are conducted annually, no additional M&V activities are expected to be leveraged to monitor persistence.
Will impacts be developed for geographic sub-regions? If so, what are these sub-regions?	Yes, impacts will be reported by LCA, SubLAP, and climate zone for each IOU.
Will impacts be developed for participant sub-segments? If so, what are these sub-segments?	Yes, industry, dual enrollments, AutoDR, customers receiving notifications, NEM, and large generators.
Will impacts be developed for sub-hourly intervals?	No. Impacts will be reported at an hourly level.
Will impact estimates be developed for additional day types beyond what the protocol specifies?	Impacts will be estimated for each event day and an average event day (likely for weekdays only).

Evaluation Consideration	Framework
Will any additional investigations be conducted to determine why the impacts are what they are, rather than simply reporting the estimates?	Ongoing involvement with each utility's program staff should provide expert context to program performance, but no additional metering or analysis will be performed.
Will common methodologies or data be used across multiple utilities' CPP programs.	A common methodology will be used to estimate each utility's program impacts. However, no data will be pooled for modeling and no participant information will be shared across IOU's.

2.1 KEY RESEARCH QUESTIONS

Different evaluation methods will be applied to different sites, depending on the number of potential control group sites in their respective subgroups. However, the overall goals for each subgroup's evaluation are the same–to answer these key research questions:

- What were the demand reductions due to program operations and interventions in PY 2025 – for each event day and hour and for the average event? How do these results compare to the ex post results from the prior year?
- How do load impacts differ for customers who are dually enrolled in other programs?
- How do weather and event conditions influence the magnitude of demand response?
- How do load impacts vary for different customer sizes, locations, and customer segments?
- What is the ex-ante load reduction capability for 1-in-2 and 1-in-10 weather conditions? And how well do these reductions align with ex-post results and prior ex-ante forecasts?
- What concrete steps can be undertaken to improve program performance?

2.2 DEMAND RESPONSE EVALUATION METHODS

The primary challenge of impact evaluation is the need to accurately detect changes in energy consumption while systematically eliminating plausible alternative explanations for those changes, including random chance. To estimate demand reductions, it is necessary to estimate what demand patterns would have been in the absence of dispatch – called the *counterfactual* or *reference load*.

The ability to measure demand reductions accurately essentially depends on four key components:

- The effect size – The effect or "signal" size is most easily understood as the percent change. It is easier to detect large changes than it is to detect small ones. With CPP rates, as well as other behavioral programs, percent impacts are relatively small since loads are not directly controlled.
- Data volatility – The more volatile the load, the more "noise" is present in the data, making it difficult detect small changes. Energy use patterns for many businesses follow regular patterns by hour-of-week and temperature, though some have more idiosyncratic usage.

- The ability to filter out noise – At a fundamental level, statistical models, baseline techniques, and control groups – no matter how simple or complex – are tools to explain variation in the data, which filters out noise, allowing impacts to be more easily detected. Matched control groups paired with a difference-in-differences model can filter out much of this statistical noise for most non-residential customers.
- Sample/population size – Since percent impacts are small, it will be important to have large sample sizes for each subgroup of the analysis. It is easier to precisely estimate average impacts for a large population than for a small population because individual customer behavior patterns smooth out and offset across large populations.

A key factor for many, but not all, demand response resources is the ability to dispatch the resource. The primary intervention – demand response dispatch – is introduced on some days and not on others, making it possible to observe energy use patterns with and without demand reductions. This, in turn, enables us to assess whether the outcome – electricity use – rises or falls with the presence or absence of demand response dispatch instructions.

In general, there are seven main methods for estimating demand reductions, as summarized in Table 4. The first four only make use of use patterns during days when DR is not dispatched to calculate the baseline. The latter three methods incorporate non-event data but also use an external control group to establish the baseline. The control group consists of customers who are similar to participants and experienced the same event day conditions but are not dispatched during events. Control and participant groups should have similar energy usage patterns when the intervention is not in place and diverge when the intervention is in effect. The only systematic difference between the two groups should be that one is dispatched for events while the other group is not.

Our general approach will be Method #5, Differences-in-Differences with a matched control group, since an RCT is not feasible for this evaluation. We will additionally use regressions models for small groups with few potential control group sites and for larger sites with unique loadshapes. These are generally an application of Method #3 in the table, though we will test for the best model specifications using models including an synthetic control groups (Method #6). For sites with erratic loads from day-to-day, we will also test models with a day-of adjustment based on morning loads from 6 to 10 a.m. Since CPP incentivizes within-day load shifting, we will restrict the use of these terms to sites where models are otherwise unable to accurately predict the baseline level of operations at a site on a given event day (e.g. sites that do not run at full capacity each day or on a set schedule).

Table 4: Methods for Demand Response Evaluation

General Approach		Method	Method Description
Use non-event days only to establish	1	Day matching baseline	This approach relies on electricity use in the days leading up to the event to establish the baseline. A subset of non-event days in close proximity to the event day are identified (e.g., Top 3 of 10 prior days). The electricity use in each hour of the identified days is averaged to produce a baseline. Day matching baselines are often supplemented

General Approach		Method	Method Description
the baseline			with corrections to calibrate the baseline to usage patterns in the hours preceding an event – usually referred to as in-day or same-day adjustments.
	2	Weather matching baseline	The process for weather matching baselines is similar to day-matching except that the baseline load profile is selected from non-event days with similar temperature conditions and then calibrated with an in-day adjustment.
	3	Regression models (interrupted time series)	Regression models quantify how different observable factors such as weather, hour of day, day of week, and location influence energy use patterns. Regression models can be informed by electricity use patterns in the day prior (day lags) and in the hours before or after an event (lags or leads) and can replicate many of the elements of day and weather matching baselines.
	4	Machine learning (w/o external controls)	Most machine learning approaches (e.g., random forest, neural networks, etc.) rely exclusively on non-event day data to establish the baselines. The algorithms test different model specifications and rely on a training and testing datasets (out-of-sample testing) to identify the best model and avoid overfitting.
Use non-event days plus a control group to establish the baseline	5	Matched control groups	Matching is a method used to create a control group out of a pool of nonparticipant customers. This approach relies on choosing customers who have very similar energy use patterns on non-event days and a similar demographic and geographic footprint. The non-event day data is incorporated by either analyzing the data using a regression model, a difference-in-differences model, or both.
	6	Synthetic control groups	This approach is similar to matching except that multiple controls are used and weighted according to their predictive power during a training period. A key advantage of this approach is that it can be used to produce results for individual customers.
	7	Randomized control trials	Participants are randomly assigned to different groups, and one group (the "control" group) is withheld from dispatch to establish the baseline. The control group provides information about what electricity use would have been in the absence of DR dispatch – the baseline. The estimate is refined by netting out any differences between the two groups on hot non-event days (difference-in-differences).

Approaches that use an external control group typically provide more accurate and precise results on an aggregate level when there are many customers (i.e., several hundred). They also make use of non-event days to establish the baseline but have the advantage of also being informed by the behavior of the external control group during both event and non-event days. Except for synthetic controls, the two fundamental limitations to control groups have been: the limited ability to disaggregate results, and the inability to use control groups for large, unique customers. The precision of results for control group methods rapidly decreases when results are disaggregated, and a control group cannot be used to estimate outcomes for individual customers (except for synthetic controls).

Methods that rely only on non-event days to establish the baseline – such as individual customer regressions – are typically more useful for more granular segmentation. Individual customer regressions have the benefit of easily producing impact estimates for any number of customer segments. Because they are aggregated from the bottom up, the results from segments add up to the totals. However, the success of individual customer regression hinges on having non-event days comparable to event days. When most of the hottest days are event days, as has been the case historically, estimating the counterfactual requires extrapolating trends to temperature ranges that were not experienced during non-event days. This produces less accurate and less reliable demand reduction estimates for the hottest days when resources are needed most.

2.3 CONTROL GROUP SELECTION

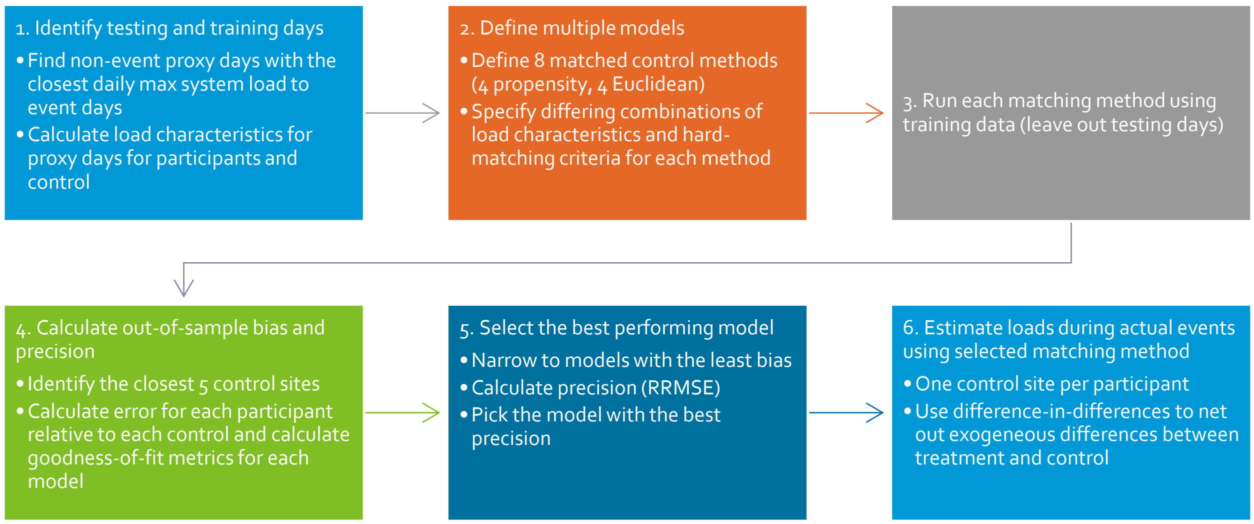
Where possible, we will construct control groups for the analysis. There are three main methods of control group construction:

1. **Randomized:** Where a group of customers that are statistically identical to the participants are used as a proxy for what the participants would have done had they not been dispatched. These control customers could either be a subset of program participants that were withheld from program dispatch, or a random subset of eligible customers group of customers who were not offered the treatment. This approach is ideal as it results in the least impact bias and clearest causal link between treatment and impact.
2. **Matched:** If a randomized control group is not possible, a matched control group can be constructed, typically using a method such as propensity score matching. This method identifies a subset of non-participants that are statistically similar to the participants based on a variety of characteristics like annual consumption, load profiles, geographic location, or NEM status. This approach, when properly executed, can approximate randomized control group performance, however it relies on a large pool of non-participants from which to construct the control group.
3. **Synthetic:** In cases where there may not be a suitable matched control group, due to very unique participant load shapes or a small pool of non-participants, a synthetic control group may be appropriate. This method constructs a reference load for any given customer from the weighted-average of several non-participants – the synthetic control group.

Both matched and synthetic control groups should be scrutinized to ensure that they produce accurate and precise counterfactuals. Out of sample testing – comparing the constructed control group to the participant population on non-event days - should be performed to select the best matching model and to confirm that the bias and precision of these methods are within acceptable ranges (generally less than 1% absolute bias). Control groups comprised of randomly selected customers are generally not subject to as much scrutiny once the randomization has been confirmed to have been performed correctly, as the only difference between participants and the control group is random chance.

Figure 1 summarizes the process that will be used to select matched controls for the difference-in-difference analyses. To identify the control pool sites that best match each participant site's energy use patterns on event-like, proxy days (similar in weather and system conditions to event days), eight matching methods will be tested. These methods include different matching algorithms (e.g. Euclidean and propensity matching) and different site characteristics. Matching methods include different combinations of proxy day load characteristics such as load factor, load shape, and weather sensitivity. Control candidates will also be "hard-matched" on subLAP, net metering status, and size bin.[3]

Figure 1: Out of Sample Process for Control Group Selection



2.4 MODEL SELECTION

CPP Impacts will be estimated using either:

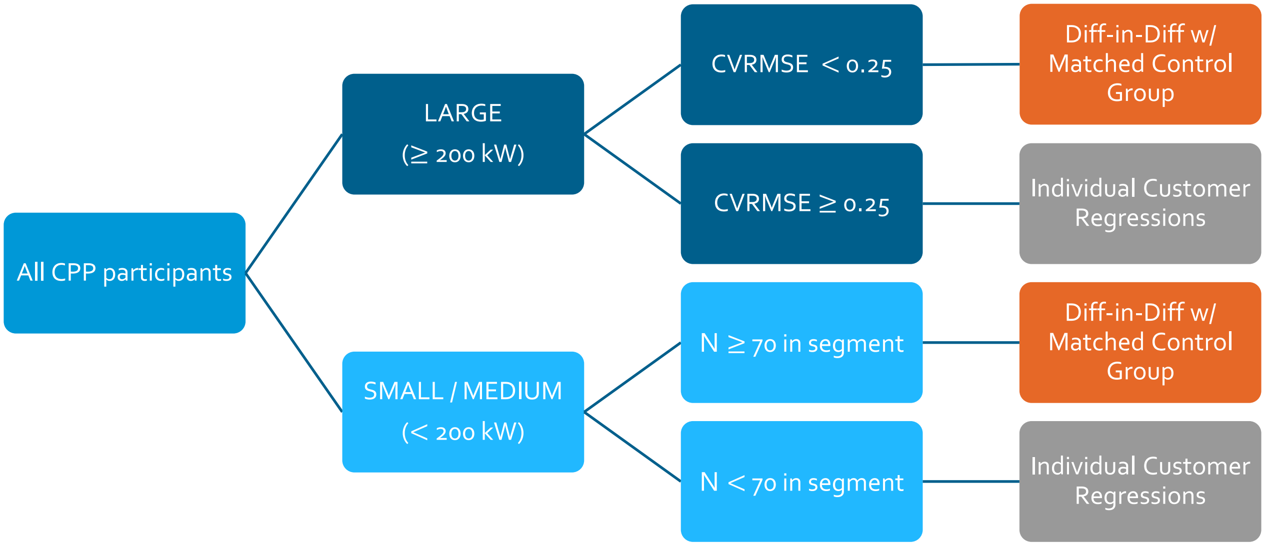
1. Difference-in-differences with matched controls, or
2. Individual customer regressions.

Figure 2 summarizes the selection framework that will be used to determine the appropriate method for each site. Most sites will utilize a difference-in-differences model, except for in cases where there are not enough sites in a given segment in terms of size (Small, Medium, or Large) and geography (SubLAP for PG&E, climate zone for SCE and SDG&E). We will also use individual customer regressions for Large sites whose daily usage patterns which exhibit substantial statistical noise (CVRMSE ≥ 0.25).

---

[3] Bins will be constructed using average usage on event-like, proxy days. For solar customers, bins were constructed based on system size.

Figure 2: Methodology Selection Framework for Ex Post Estimates



Site-specific models for individual customer regressions will be selected among dozens of potential specifications, which will include synthetic controls using one or more matched control sites to help control for factors outside of the CPP events.[4] Similarly, the difference-in-differences approach will use a matched control group to net out changes in energy usage patterns not due to the CPP events. As such, regardless of evaluation methodology, each participant site will be matched to one or more non-participant using a matching tournament where match quality is compared across eight different matching models to identify the best performing model.

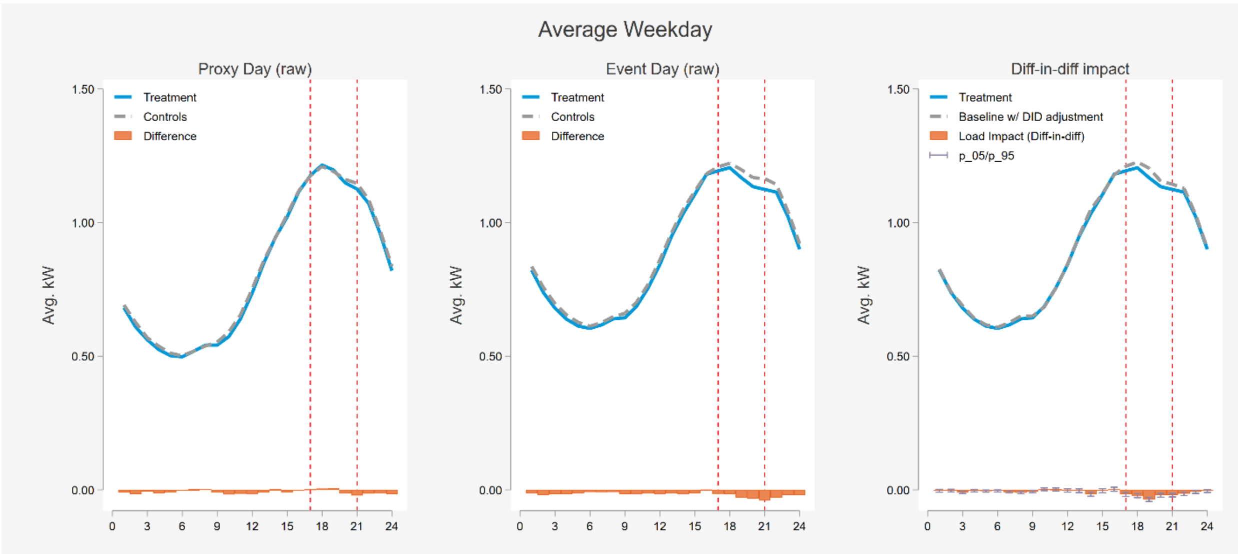
As described above, difference-in-differences with matched controls will be the primary evaluation methodology used, except in cases where there were few sites or large sites with noisy load patterns.[5] Figure 3 below demonstrates the mechanics of a difference-in-difference calculation. In the first panel, average observed loads on proxy days are shown for participants and for their matched controls. The difference between these two is the first "difference" and quantifies underlying differences between participants and their controls not attributable to event participation. Note that this first difference is very small, indicative of a high-quality match and sufficient sample size to neutralize the noise inherent

[4] The functional form of a regression with synthetic controls differs from a panel difference-in-difference regression in that usage for the controls is specified as a right-hand-side predictor variables.

[5] Out of sample testing was used to calculate RRMSE and other bias and fit metrics to compare across multiple pooled methods (average customer regressions and panel regressions). Based on this testing, difference-in-differences was determined to outperform or at least be comparable in robustness to the other methods. In contrast to the pooled regression-based methods, difference-in-difference has the advantage of enabling segmentation of results (by size, subLAP, industry, solar status, etc.) without the need to run additional regressions while ensuring that segment results add up to group totals.

in individual customer loads. The second panel shows the average observed participant and matched control loads on event days. The gap between these two is the second "difference" which includes both the difference due to event participation and the underlying first difference observable on non-event days. The third panel shows the average event day loads after netting out the proxy day difference from the event day control load. The result is the difference-in-differences impact.

Figure 3: Difference-in-Differences Calculation Example



In cases where a difference-in-differences approach is not deemed appropriate due to insufficient sample size or for large sites with noisy loads, site-specific individual customer regression models will be selected using another out of sample tournament to select the most accurate regression model specification for each participant site. To implement out of sample testing, the top 50 system load days, excluding event days, will be randomly divided into testing and training datasets. Bias and fit metrics will be calculated using the testing dataset and the model with the best fit (lowest Root Mean Squared Error) will be selected among models with the least bias (Mean Absolute Error).[6] Site specific load impacts will be estimated using the winning model for each site.

Table 5 summarizes the metrics for bias and precision we employ. Bias metrics measure the tendency of different approaches to over or under predict and are measured over multiple days. The mean percent error describes the relative magnitude and direction of the bias. A negative value indicates a tendency to under predict, and a positive value indicates a tendency to over predict. This tendency is best measured using multiple days and hours. The precision metrics describe the magnitude of errors for individual events days and are always positive. The closer they are to zero, the more precise the results. The mean percentage error is used to narrow down to the three models with the least bias. The

[6] MAE was used rather that Mean Average Percent Error (MAPE) to ensure robustness for sites with loads very close to zero, common for sites with solar or other generation.

Relative RMSE metric is used to identify the most precise and final model among the remaining candidates.

Table 5: Definition of Bias and Precision Metrics

Type of Metric	Metric	Description	Mathematical Expression
Bias	Average Error	Absolute error, on average	$AE = \frac{1}{n}\sum_{i=1}^{n}(\hat{y}_i - y_i)$
	% Bias	Indicates the percentage by which the measurement, on average, over or underestimates the true demand reduction.	$\% \, Bias = \frac{\frac{1}{n}\sum_{i=1}^{n}(\hat{y}_i - y_i)}{\bar{y}}$
Precision	Root mean squared error (RMSE)	Measures how close the results are to the actual answer in absolute terms, penalizes large errors more heavily	$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(\hat{y}_i - y_i)^2}$
	Relative RMSE	Measures the relative magnitude of errors across event days, regardless of positive or negative direction. It can be thought of as the typical percent error, but with heavy penalties for large errors.	$CV(RMSE) = \frac{RMSE}{\bar{y}}$

2.5 EX POST IMPACTS

Once the counterfactual event day load has been developed, the difference between that reference load and the observed load is the program impact. Impacts will be reported:

- For each hour on each event day
- For the average event hour on the average event day

Ex post impacts will also be reported out for particular sub-segments of enrolled participants. While the exact segments will vary depending on the subgroup, the typical set of segments include the following:

- **Size:** peak demand less than 20 kW (Small) , between 20-200 kW (Medium), and greater than 200 kW (Large)
- **Region:** Local Capacity Area, SubLAP, and climate zone
- **Industry Segment:** for non-residential customers, identified by customer's NAICS code
- **Dual Enrollment:** either dually enrolled with another program or not
- **NEM/Solar Status:** included if participants have a high penetration of rooftop solar
- **AutoDR enrollment**

- **Large generators:** sites that predominantly export power during most hours, such as solar farms
- **Sites receiving day-ahead notifications**

2.6 EX-ANTE IMPACTS

A key objective of the DR evaluations is to quantify the relationship between demand reductions, temperature, hour-of-the-day, and dispatch strategy. The purpose of doing so is to establish the demand reduction capability under 1-in-2 and 1-in-10 weather conditions for planning purposes and, increasingly, for operations. When possible, we rely on the historical event performance to forecast ex-ante impacts for future years for different operating conditions.

The process of estimating ex-ante impacts essentially involves:

1. Use at least two years of historical performance data
2. Decide on an adequate segmentation to reflect how the customer mix evolves over time
3. Estimate the relationship between reference loads and weather
4. Use the models to predict reference loads for 1-in-2weather conditions)
5. Estimate the relationship between weather and percent impacts
6. Predict percent reductions for different weather conditions (and/or dispatch hours)
7. Combine the reference loads (#4) and percent reductions (#6) to produce per-customer impacts
8. Multiply per-customer impacts by the enrollment forecast

The process can be used to develop ex-ante estimates of demand reduction as a function of temperature, event start time, and event duration. It can be used to develop estimates for 1-in-2 and 1-in-10 weather year planning conditions, and it can be used to develop time-temperature matrices useful for estimating reduction capability for operations or a wider range of planning conditions.

The conversion of ex post impacts to an ex ante forecast should be transparent and understandable to outside stakeholders. In general, the differences between the two are due to several key distinctions:

1. **Customer Mix:** Difference in participant population mix or forecasted enrollment
2. **Weather:** Ex post observed weather may be hotter or colder than ex ante planning conditions
3. **Event Time:** Ex post events may not occur during the RA window for which ex ante impacts are developed
4. **Historical Data:** Ex ante data should explicitly incorporate multiple years of impacts, so average impacts may change when additional years of ex post data are included
5. **Program Design:** If dispatch strategy, eligible months, or program participation options change, ex post impacts may not represent the future capability of the program

As part of the reporting process, we will capture the impact each of these changes has on the difference between ex post and ex ante impact estimates.

Finally, as the results of demand response impact evaluations are increasingly used to support operational concerns, the evaluation team will also provide time-temperature matrices for all subgroups. These matrices will rely on the ex ante impact estimates to predict, for different event start times, durations, and weather conditions, what the average customer hourly impact could be. This will be provided to each utility's program staff separately from the ex ante load impact tables.

At each utility's discretion, the evaluation team may produce, in parallel to the current ex ante reporting requirements, ex ante estimates to PY2025 programs in the upcoming reporting format as described in the ongoing Resource Adequacy proceedings. As proposals in this proceeding are not yet finalized or approved, the evaluation team makes no attempt to summarize the specific reporting requirements here. The new proposal is expected to be finalized in the next six months and will apply to the 2025 compliance year.

For each subgroup, a slice-of-day table will be provided in addition to the standard weather year ex-ante impact tables. A slice-of-day table shows the hourly impacts for the worst day of each month based on the year selected.

PROGRAM-SPECIFIC VERSUS PORTFOLIO-ADJUSTED IMPACTS

Attribution rules for dual program enrollment vary by CPP subgroup, as summarized in Table 6 and Table 7. Evaluation analyses will be conducted in alignment with these rules and therefore portfolio adjusted ex ante estimates for CPP will be incremental to the dual dispatched programs. Note that while dual enrollment is allowed between CPP or PDP and ELRP, adjustments for these dual enrollments will be addressed in the ELRP evaluation. Therefore, CPP or PDP impacts specific to sites dually enrolled in CPP or PDP and ELRP will not be reported on or adjusted for this evaluation. Also, no utility had CPP participants that were also enrolled in CBP in 2025, and in future years dual participation in these two programs will not be allowed, so CBP will not factor into any ex ante estimates. There were no other dual enrolled groups for SDG&E.

Table 6: CPP/Peak-Day Pricing (PDP) Subgroup Dual Enrollment Rules for Impact Estimates – PG&E

Dual Group	Study	Ex-Ante Program Specific	Ex-Ante Portfolio Adjusted
PDP + BIP	PDP	PDP and overlapping events, single and dual customers	Any impacts beyond FSL
	BIP	BIP and overlapping events, single and dual customers	Impacts are capped at FSL

Table 7: CPP Subgroup Dual Enrollment Rules for Impact Estimates - SCE

Dual Group	Study	Ex-Ante Program Specific	Ex-Ante Portfolio Adjusted
CPP + BIP	CPP	CPP and overlapping events, single and dual customers	Any impacts beyond FSL
	BIP	BIP and overlapping events, single and dual customers	Impacts are capped at FSL
CPP + SDP	CPP	CPP and overlapping events, single and dual customers	SDP event average removed from impacts
	SDP	CBP and overlapping events, single and dual customers	Any impacts

2.7 EXECUTIVE SUMMARY AND CPUC ENERGY DIVISION REQUESTS

A requirement over the last several years has been to provide supplemental reporting to the Energy Division for long term planning. For all programs in Utilities' PY2025 portfolio, including the statewide programs, several additional reporting features are due to the CPUC on or before November of 2025. Demand Side Analytics will provide these per the requirements below, with both a public and confidential version enclosed:

1. Ex Ante Load Impacts in plain Excel format, due on or before April 1st of each year:
    a. Portfolio aggregate ex-ante load impacts for 1-in-2 weather year monthly system peaks for each of the 11 ex-ante forecast years, for both the IOU's service area and each LCA within the service area
    b. Portfolio aggregate ex-ante load impacts for 1-in-10 weather year monthly system peaks for each of the 11 ex-ante forecast years, for both the IOU's service area and each LCA within the service area
2. Portfolio aggregate ex-ante load impacts by program for 1-in-2 year August system peak for each of the full ex-ante forecast period years, disaggregated by WECC busbar. Due by November 1, 2026
3. Portfolio aggregate ex-ante load impact by program for the 1-in-2 weather year monthly system peak in the final year of the forecast, for all program operating hours (not just RA window). Document the methods used to estimate non-RA hour impacts. Due by November 1, 2026

3 QUALITY CONTROL PROCEDURES

The Demand Side Analytics team takes analysis accuracy seriously. We have several processes in place to ensure all data management, analysis, and reporting are delivered with the highest quality. A summary of our philosophy, however, is enumerated below:

1. There is clear oversight in each project by an expert in Demand Response evaluation. Our senior staff are familiar with the types of programs being evaluated, the preferred methods and their respective strengths and weaknesses, and the California demand response landscape. We understand these programs and their evaluation challenges.
2. Whenever possible, we rely on automated reporting and tabulation. This allows us to go from data validation to reports quickly and efficiently, without errors caused by version control, manual data entry, or copy and paste errors.
3. We understand the reporting requirements to conform to the California Load Impact protocols. Because of our background, we don't anticipate surprises in the format, content, or timeline of the key project deliverables, which means that utilities will get the right information at the right time in a clear, accessible format.

3.1 DATA CHECKS

The first step for quality control is to make sure that all data that had been requested is both accounted for and does not contain spurious values. To that end, we have implemented a detailed checklist for our demand response evaluations that investigates common data pitfalls for each type of data typically used in a demand response evaluation. A summary of these questions typically includes:

1. **Interval Data:** Is the data in the right units? Adjusted for Daylight Savings and any grid export/net demand?[7] Is there a full panel of data for all customers? Are there outliers in terms of customer size? Did we receive all the interval data for the customers we requested?
2. **Customer Characteristics:** Do we have all the relevant participant and control groups? Do we have DR enrollment data for all customers and were they affected by other interventions during the analysis period? Do we have all the characteristics that are needed for reporting?
3. **Treatment and Event Data:** Do we have the correct event days identified? Are the event days and hours properly coded? Can we visually see when customers are reducing loads during events?
4. **Weather Data:** Is the DST adjustment in the weather data consistent with that of the interval and event data? Is it in the right time zone and units?

[7] PG&E impacts will use delivered load, SCE and SDG&E impacts will use net loads.

Because incorrect data will lead to incorrect results, any issues that are identified to be significant to the evaluation will be addressed with each utility's team to ensure quick resolution.

3.2 ANALYSIS CHECKS

Analysis checks are critical to a successful evaluation, and where our expertise in DR evaluations will provide value. Because of our familiarity with these demand response programs and the California load impact protocols, we are able to quickly identify results that do not make sense and either correct the issue or identify the reason why results differ from our initial assumption. While analysis checks tend to be program specific, the general considerations are:

1. **Analysis Dataset Construction:** Is the control group constructed appropriately? Is it statistically indistinguishable from the treatment group on days when no customer was dispatched? What are the results of out of sample testing? Given model precision and bias, will we be able to detect the expected effect?
2. **Ex post results:** Are the results generally in line with prior years, given no substantial program changes? Are all customers dispatched as expected? Do weather sensitive programs see greater impacts on hotter days? Do reference load patterns follow the same trend as the raw data with regards to temperature? What are the distributions of impacts - are there large customers that are driving the majority of impacts? Are there particular customer segments that respond differently?
3. **Ex ante results:** Given the differences between ex post and ex ante weather and participation, do reference loads look appropriate for each day type and weather year? What about percent impacts? Have we captured the effects of dual enrollment for program and portfolio impacts appropriately? Have changes to program design or enrollment been captured in the ex ante forecasts?

The focus of these questions is to ensure that there are no surprises in the evaluation report and that all results are situated in their full context. In collaboration with each utility's team, we will work to frequently share draft findings and raise any issues as they arise.

3.3 REPORTING CHECKS

Many iterations are expected in the process of producing draft and final evaluation reports, load impact tables, and other results memos. In those cases, opportunities arise for omissions, copy/paste errors, and gaps in reporting updates. To the extent possible, the evaluation team relies on automated reporting and table generation, where the latest version of the analysis is automatically written into a report. This ensures that reports and load impact tables are consistent in their results, and that all values are updated whenever an updated version of the analysis is implemented.

3.4 PROJECT MANAGEMENT CHECKS

As discussed in the kickoff meeting, Alana Lemarchand will be the key contact for all project management topics. They will both be responsible for ensuring that the project remains on time and on budget and will identify bottlenecks or issues likely to affect the project timeline as soon as possible to the Statewide CPP team. As part of this process, monthly reporting on budget, key tasks completed, upcoming deliverables, and any changes to the schedule will be provided to the Statewide CPP team.

4 DATA NEEDED

Demand Side Analytics is delivering initial data requests along with the draft version of this evaluation plan. At a high level, the data requests include nine items:

1. Customer characteristics file for participants
2. Hourly interval data for participants
3. Event data
4. Outage data (included in ELRP data requests for PY 2025)
5. Weather data (included in ELRP data requests for PY 2025)
6. Customer characteristics file for control sites
7. Hourly interval data for control sites (to be requested after a sample draw)
8. Dual program enrollments
9. Event notifications

5 TIMELINE

Table 8 below shows the next steps for the evaluation of the Statewide CPP programs:

Table 8: Timeline of Key Deliverables

	Task	Deliverables	Timing
1	Project Management	Regular Meetings	October 2025-March 2025
		Kick-Off Meeting	9/16/2025
		Kick-Off Memo	9/23/2025
2	Evaluation Plan	Draft Evaluation Plan	10 business days after kick-off meeting: 9/30/25
		Final Evaluation Plan	5 business days after comments received; SDG&E to submit to CPUC by December
3	Data Collection and Validation	Data Request	9/30/2025 Secondary request for AMI data for potential control pool to follow
4	Ex-Post Results	Draft and Final Result Spreadsheets	Present draft results: 12/5/2025 Comments on draft load impacts: 12/15/2025 Draft ex post table generators: 12/22/2025 Final ex post table generators: 1/10/2026
5	Ex-Ante Results	Draft and Final Result Spreadsheets	Draft Time-Temperature Matrices: 1/15/2026 Present draft results: 1/25/2026 Comments on draft load impacts: 2/5/2026 Draft ex ante table generators: 2/10/2026 Final ex ante table generators: 2/28/2026
6	Documentation & Reporting	Draft Evaluation Report	Draft to IOUs: 2/10/2026 Comments due: 2/23/2026
		Final Evaluation Report	2/28/2026
		Executive Summary Tables	3/10/2026
		CALMAC Abstract	3/10/2026
7	Presentation of Results	Internal Presentations DRMEC Workshop	April 2026 May 2026
8	Database Documentation	Produce database files	3/1/2026