On their best performance: Are NMEC programs the future of demand side programs?

Christie Torok, California Public Utility Commission
Coby Rudolph, California Public Utility Commission
Peng Gong, California Public Utility Commission
Ken Agnew, DNV
Lullit Getachew, DNV
Stephanie Whalley, DNV

ABSTRACT

NMEC (normalized meter energy consumption) and other performance-based program designs are increasingly seen as an opportunity to carry demand-side programs into the future. These programs measure savings at the meter and use those estimates to reward implementers and/or customers. NMEC designs offer several key advantages for all parties involved.

- Customer – A broader and flexible menu of offerings that are attuned to the market and provide more reliable savings.
- Implementer – Performance-based payment that directly rewards the quality of delivery.
- Program Administrator (PA) – Third-party program implementation and payment based on measured savings.
- Ratepayer – Lower risk as programs get paid for demonstrated savings.

Each of these areas of advantage carries downside risks that could limit or even undermine the potential of NMEC. In particular, establishing NMEC programs that empower implementers while overcoming methodological constraints to address non-routine events presents a significant challenge.

This paper will introduce the reader to the concept of NMEC in practice and illustrate the current state of NMEC programs in California. We consider whether NMEC methods, as currently conceived and implemented, can provide accurate claims of savings and, if this is the case, what would be the possible implications for evaluation of NMEC programs. A recently completed round of evaluations for population NMEC (residential and small commercial) and site-level NMEC (commercial) programs provide evidence for this summary. The evaluations also provide preliminary evidence regarding the relative success of NMEC. In total, the paper will add to the long process of answering the question: Are NMEC programs the future of demand side programs?

Introduction

NMEC is an emerging offering within the California energy efficiency portfolio with pay-for-performance elements. While, at times, the name NMEC refers to meter-based methods in general (e.g., strategic energy management, etc.), this paper uses NMEC as the name applied to a new set of programs sanctioned and developed to highlight these methods. NMEC as a focus, developed out of California's specific policy environment. This focus grew out of the aspirations of SB350's[1] mandate to double energy efficiency (EE) in CA and AB802's[2] offerings

[1] SB350. Stats. 2015, Ch. 547, authored by Senator DeLeon, signed by Governor Brown October 7, 2015.
[2] AB802. Stats. 2015, Ch. 590, authored by Assembly member Williams, signed by Governor Brown October 8, 2015.

as part of the solution to that massive challenge. NMEC also stepped into the rigorous CA EE evaluation environment. There is an ongoing discussion regarding whether this novel endeavor of NMEC is being cautiously nurtured or hindered as it develops.

The new NMEC programs' touted potential stems from their ability to estimate savings directly from meter data and the greater freedom afforded implementers given the performance-based measurement and risk sharing across all parties involved. For this potential to be realized, performance assessments by implementers must be able to accurately quantify program-related savings. Ultimately, if implementers can accurately quantify savings, then there is the further hope that NMEC might reduce the need for ex post evaluation. Driving these aspirations is a perceived treasure trove of below code savings that are unavoidably included as NMEC methods reward all savings measured from the existing conditions baseline.

This paper cannot definitively assess each of these aspects of the NMEC project nor conclusively answer the question in the title but it seeks to offer a preliminary perspective now that programs have completed multiple years of implementation and a round of evaluation. As regulators and evaluators, the authors of this paper offer a view from our place overseeing the policy guardrails in which NMEC operates as well as a careful examination of the first fruits of NMEC labor. We approach this with no preconceived position regarding the potential future success of NMEC. The paper will attempt to answer the following questions: Can NMEC methods adequately address non-routine events (NREs) to give confidence in the calculated savings? How do program and evaluation operations have to change in an NMEC environment? Does NMEC induce expanded savings beyond what's likely to be achieved through more traditional programs?

This paper will start with a quick overview of the relevant policy, place the policy in the wider context of performance-based programs, and then introduce the specific NMEC offerings in California, i.e., population NMEC and site-level NMEC programs. We will provide results from the first impact evaluations of these NMEC programs and then summarize findings both specific to the evaluations and to NMEC in general.

Background

In 2015, Assembly Bill 802 introduced the world to NMEC. Section 381.2(b) in Assembly Bill 802 states that "the commission, in a separate or existing proceeding, shall, by September 1, 2016, authorize electrical corporations or gas corporations to provide financial incentives, rebates, technical assistance, and support to their customers to increase the energy efficiency of existing buildings based on all estimated energy savings and energy usage reductions, taking into consideration the overall reduction in normalized metered energy consumption as a measure of energy savings."[3] The separate proceeding, to which the bill referred, was Rulemaking 13-11-005, *Assigned Commissioner and Administrative Law Judge's Ruling Regarding High Opportunity Energy Efficiency Programs or Projects,* generally referred to as the HOPPs ruling. For these early, experimental expressions of NMEC programs, the High Opportunity Programs or Projects (HOPPs) ruling clarified AB802 with the following guidance.

Those new NMEC programs "shall include:

- Energy usage reductions resulting from the adoption of a measure or installation of equipment required for modifications to existing buildings to bring them into conformity

[3] AB802. Ibid.

with, or exceed, the requirements of Title 24 of the California Code of Regulations, as well as
- Operational, behavioral, and retrocommissioning activities reasonably expected to produce multi-year savings."[4]

The HOPPs ruling offered points of clarity for these early examples of NMEC projects that remain relevant for NMEC programs that have developed subsequently. NMEC projects should occur in existing buildings and either bring those buildings up to or beyond code (Title 24) or apply behavioral, operational, and retrocommissioning activities, otherwise referred to as BRO measures. The guidance points to addressing "stranded potential," below code energy savings that are not currently sufficiently targeted or rewarded in the context of program offerings that primarily target savings above code. It states that "PAs should draw from the wealth of saturation studies, pilots, EM&V, and other analysis, to find and propose clear winner projects/programs." The guidance underlines this focus by italicizing what is "newly permissible" which is this focus on reaching stranded potential.

The most important "newly permissible" aspect of NMEC is the opportunity to measure the savings of energy efficient improvements from an existing conditions (EC) baseline for the full expected useful life (EUL) of project measures rather than only crediting the savings above existing code or Industry Standard Practice (ISP) for all or part of the EUL, as is the case with most existing, relevant EE programs. An NMEC approach to measuring savings, the difference between baseline and metered energy consumption, can only provide a savings estimate that includes all of the savings from the EC baseline;[5] both the savings that would be gained by replacing with the least efficient measure available on the market (code, by definition) and any marginal efficiency improvements above that. This is why NMEC projects must be "stranded potential" opportunities. With the allowance of projects that just return an existing building to code, there is a clear danger that these NMEC projects will be indistinguishable from the routine updating of a building's systems. The only thing that distinguishes NMEC from being a universal PA-funded opportunity for routine building upgrades is the clear intention to limit the focus to upgrades from existing conditions that would not happen in the absence of the PA support.

Decision 16-08-019, Decision Providing Guidance for Initial Energy Efficiency rolling portfolio Business Plan Filings, further clarified the question of baselines. "(T)he purpose of the changes directed by AB802 was to unlock further project opportunities that the utilities and industry were observing were not being captured by the code baseline default framework. This is perhaps the converse of our historic concern that using an existing conditions baseline will result in paying for projects that would have happened anyway. . . The reality is that both of these concerns are valid – they are truly two sides of the same coin."[6] This line highlights that stranded potential exists because, for customers with repair-capable systems, the default code baseline may represent too high a threshold from which to start awarding incentives for them to be motivating. These customers require some form of incentive just to replace the existing

---

[4] https://docs.cpuc.ca.gov/PublishedDocs/Efile/G000/M157/K362/157362236.PDF
[5] These details are primarily relevant to commercial and/or site-level scenarios. For residential savings, the assumption that savings are accelerated is accepted up front and, where not the case, addressed in net-to-gross ratios. Residential savings do not face a shortened ECB period for accelerated replacement.
[6] Decision 16-08-019 August 18, 2016. https://ccag.ca.gov/wp-content/uploads/2016/10/6.5-Attachment-ALJ-Decision-16-08-019-081816.pdf

equipment and move to code let alone to greater efficiency options. The decision clearly articulates the concern about the similarities with a free ridership scenario and says that NMEC, with the EC baseline, must serve the projects that need unlocking. This understanding of NMEC as targeting stranded potential provides the logic for how NMEC can induce savings beyond what's likely to be achieved through more traditional programs. This is particularly important in the site-level NMEC context where estimating savings relative to EC baseline for the full EUL is a substantial increase in savings over the typical accelerated replacement dual baseline that only uses ECB for the accelerated period.

NMEC as an example of P4P

NMEC is the name assigned to a specific California-based instance of a performance-based program (P4P) design. The details of how NMEC savings are derived is essential to understanding how accurate implementer performance assessments may be and whether the evaluation of those savings can and should change as a result. The key attribute of a P4P scheme is that savings are claimed and paid based on demonstrated reductions in consumption using meter-based data rather than, for example, the assignment of pre-established deemed values based on installed measures. Measuring performance in this way requires that consumption from before and after the installation of EE measures is characterized on common terms across the baseline and performance periods with respect to drivers such as weather. With all drivers of consumption set to equal across baseline and performance periods, the remaining difference is defined as savings. If either baseline or performance consumption is poorly characterized, the estimate of savings may not reflect the true underlying savings.

The process of characterizing consumption and controlling for differences across periods leverages statistical regression techniques. In addition to now readily available site-level interval consumption data, weather and schedule variables are common to most P4P models. HVAC systems and their weather-sensitive consumption frequently represent a large portion of site-level consumption and will vary from baseline to performance period as the weather varies. Schedule variables may be as simple as the day of the week or more site-specific, such as school activity calendars. The regressions capture consumption dynamics with respect to these drivers and are then observed at common weather (TMY, CZ2021) and schedule conditions to set up the baseline versus performance period comparison that yields an estimate of savings.

P4P regressions need to explain enough of the natural variation in the consumption data that the resulting estimate of savings can be reasonably believed to not be just the result of natural variation itself. The success with which regressions characterize a site's consumption is captured in various goodness-of-fit metrics that measure the variation not explained by the regression in general and, importantly, relative to the expected savings. Regressions control for "routine events." That leaves variation not controlled for by the regression to be referred to as non-routine events (NREs), though practically, the relevant NREs are bigger, "unrelated facility-changes that impact savings and are not accounted for in the calculations."[7] Unplanned load spikes and short-term site shutdowns are examples of NREs.

As the name suggests, NREs represent consumption that should not be considered part of typical site consumption for estimating savings. NREs represent a significant risk to P4P programs because "NREs that go unaccounted for can introduce unacceptable levels of error and

[7] IPMVP Application Guide on Non-Routine Events and Adjustments. October 2020. EVO 10400-1:2020. Page 1.

skew reported savings making them less meaningful."[8] NMEC programs may lower rate-payer risk related to poorly determined or inappropriately assigned deemed savings but the potential error in savings estimates leads to other risks that could affect the ratepayer as well as implementers and participating end customers. How NMEC methods address NREs will determine the level of confidence in the calculated savings.

Consumption modeling in general and, more specifically, the identification and treatment of NREs in particular are important because P4P approaches have key structural differences from traditional EE programs. P4P generally moves the calculation of performance estimates to the implementer to determine claims and payments. This effectively moves what has been traditionally an EM&V function to the implementer. This change raises practical challenges for Program Administrators (PAs) or implementers for whom this is a novel function. This shift also raises a conceptual challenge as what has heretofore been a role played by independent evaluators, by definition, without a horse in the race, is now performed by a party that will either claim the savings or be paid for having facilitated those savings. This necessitates a reformulation of roles and requirements. From the regulatory perspective, guardrails are required to prevent strategic or inadvertent skewing of reported savings. If guardrails can fully address these kinds of issues, it may be possible to re-conceive the role of evaluation since there would be limited added value to a fully redundant quantitative assessment of estimated savings.

Population NMEC

General description

Population NMEC is currently applicable for residential customer populations and some commercial populations. A key eligibility requirement in the NMEC Rulebook states that "The sites can reasonably be expected to have similar types of equipment holdings, as well as drivers and levels of energy consumption."[9] Similarly, factors that impact both consumption and energy savings should be similar. Population NMEC is a P4P program. While end customers may receive incentives that are not performance-based, implementer payments are at least 50% performance-based with a stated ideal of 100%.

Population NMEC programs assess performance using similar methods evaluators have traditionally applied ex post, with the PA applying them on the implementation side to estimate implementer payable savings and final claimed savings. For population NMEC, the same consumption model is applied to all customers in a program. Routine events and weather are addressed with a single model structure and modelling approach. NREs are not addressed individually, at the site-level, but at the aggregate level. Whereas site-level approaches address NREs directly to remove their effect from estimated savings, population-level approaches assume that average comparison group change over time is a reasonable estimate of aggregate participant NREs and mitigates the effect of NREs on estimated savings. In the context of an aggregate population, NREs are synonymous with non-program, or exogenous, change.

The details of the methods are relevant because the specific implementation can have substantial implications for results. While randomized controlled trial designs, the best design for aggregate performance estimates, are specifically mentioned in early guidance as ideal for

[8] IPMVP ibid.

[9] Rulebook for Programs and Projects Based on Normalized Metered Energy Consumption. Version 2.0. CPUC. Page 12.

NMEC, they are not feasible for measure-based program designs where control group customers would need to be denied access to the program benefits. For the alternative quasi-experimental methods to successfully achieve an unbiased estimate of program savings, several assumptions must prove true or, if not, their potential implications accepted. The comparison group should be representative of the participant group (apart from the NMEC measures) both in the baseline period and in its consumption trends over time. In combination, this allows the comparison group to be a good counterfactual for participants in the performance period. That is, the comparison group represents, and controls for, how participants would have changed without the program. Traditionally, evaluators have applied matched comparison group methods that approximate representativeness in the baseline period by matching pre-period consumption and other characteristics. The potential for trend differentials between the participant and comparison group is not so easily addressed and will lead to biased results if present.[10] It is important to note that these issues also plague the ex post evaluation.

The NMEC Rulebook is intentionally not prescriptive regarding specific methods, and a detailed consideration of the specific methods in use for population NMEC is out of scope for this paper. However, the method currently in use in California does represent the introduction of a novel method to energy M&V, which needs to be acknowledged. The method departs from the typical approach discussed above that uses baseline and performance period data from participants and a group of selected non-participants combined in a difference-in-difference. It uses granular profiles that replace the comparison group with an average hourly consumption series aggregated across non-participants from the same category (solar status and climate zone for residential and industry, climate zone, and size for non-residential) as participants. The approach also integrates the non-participant data in a novel way. The method uses these granular profiles as additional explanatory variables (synthetic controls) in each participant's baseline and performance period models in place of estimating separate pre- and post-installation models for a comparison group. This approach replaces the traditional selection of matched non-participants with an average non-participant shape incorporated in a different model structure. While the PA's M&V provider has provided some vetting of the method, a complete understanding of the method in the consumption modeling context will take time. This approach has the useful attribute of avoiding the need to share individual non-participant data with implementers, with the associated data-sharing concerns. While there is no evidence these new methods have greater shortcomings than traditional methods, nonetheless the full range of implications of the methods has not been explored.

A second area of concern within population NMEC relates to the eligibility requirements discussed above regarding similarity of participants. Commercial programs, like on-bill financing (OBF), generally have smaller numbers of participants, with the potential for extreme differences in consumption magnitude across customers and more variable consumption patterns. The population NMEC commercial program in place limited participation based on size relative to overall program consumption to address this issue, but the evaluated savings of this program raise the question of whether those limitations were sufficient. Only time and additional research will produce a final recommendation of, for example, size and variability limitations for a commercial population NMEC program.

---

[10] A related challenge for these kinds of models comes in the estimation of peak effects. The comparison group is matched on more general consumption patterns but expected to work for consumption during extreme conditions. Rather than a trend differential across time, a trend differential moving into more extreme temperature conditions causes the estimation of biased peak effects.

Current programs

The evaluation of four PG&E administered population NMEC programs from 2019 to 2021, each managed by different implementers, form the backdrop of this section. An external provider conducted performance-based M&V services for these programs. While the M&V provider and methods evolved over the three-year evaluation period, the final M&V provider recalculated results retrospectively for all three years.

Figure 1 provides information on the population NMEC programs evaluated, which offers insights into the current status of NMEC programs in California. All of these programs started as HOPPs programs and preceded the CPUC NMEC Rulebook governing such programs. The evaluated population NMEC programs were diverse with distinct designs and offerings. The three residential programs included behavioral interventions based on audits of the energy use profiles of homes, deep residential retrofits or HVAC upkeep, residential energy-saving kits, and online discounts for small energy-efficient products. The commercial programs provided on-bill financing for commercial sector energy-efficient upgrades.

Pay for Performance Comfortable Home Rebates Program

PG&E residential downstream program that helps customers reduce energy use and bills and improve their comfort.

Offers a retrofit of the home through an improvement pathway or upkeep of the HVAC system through a maintenance pathway.

Pay for Performance HomeIntel Program

PG&E residential downstream program that audits the energy use profile of homes via a web app to educate participants about energy use and energy saving tips.

Provides a support staff to help participants implement energy saving practices and solutions.

Pay for Performance Home Energy Rewards Program

PG&E residential direct install and downstream that offers customers energy audits, saving tips, education, a free energy-saving kit, online discounts on energy efficient products, and rebates for smart thermostats.

On-Bill Financing Alternative Pathway Program

PG&E downstream non-residential program that offers customers utility financing for all energy efficient upgrades including those outside of utility rebate and incentive programs.

Figure 1. Population-based NMEC programs, PY2019 - PY2021

Impact evaluation high-level findings

The population NMEC programs required two sets of evaluation results to reflect an evolution in program savings results. [11] NMEC claims are finalized and claimed after the 12-month performance period has passed. Official savings claims were filed for these programs during that timeframe for each implementation year. Subsequently, the PA had their M&V contractor updated data and methodology, including a complete overhaul of M&V methods. The original, "official" claimed savings for the population NMEC programs were quite different from the updated M&V results done by the PA. Our validation efforts focused on the updated results (and methods) because these new methods will guide future efforts. However, the initial claimed savings are still the official savings.

As Table 1 illustrates, we present two realization rates (RR), one that adjusts the original "official" claim to the updated PA result and a second that represents the evaluation assessment

[11] https://www.calmac.org/publications/Final_Population-Based_NMEC_Program_Impact_Evaluation_Report_-_Program_Years_2019_to_2021_toCALMAC.pdf

of the updated PA result. The PA's ratios of updated M&V to claimed savings (PA M&V RR) for the residential P4P programs were significantly lower than 100% and varied widely across programs, from 16% to 59%. The commercial OBF program, originally serviced by a different M&V contractor produced a PA M&V RR of 135%. For the residential programs, the evaluation savings estimates largely concurred with the updated PA M&V results resulting in RRs close to 100%. The OBF evaluation result was well below the PA M&V result producing an RR of 62%. The low OBF RR reflected the inclusion of just two additional sites that the PA had excluded and highlights the variability of site-level results for the commercial program.

Table 1. PY 2021 Population NMEC Impact Results

Program	Electric sites	Final claimed savings	PA M&V results	PA M&V RR	Evaluated savings	Evaluation RR
P4P - HomeIntel	2,648	1,994,086	1,185,439	59%	1,185,409	100%
P4P - Comfortable Home Rebates	2,276	3,201,543	1,386,773	43%	1,383,443	100%
P4P- Home Energy Rewards	1,779	673,671	108,948	16%	107,230	98%
On-Bill Finance	392	10,206,131	13,774,580	135%	8,596,574	62%

Note: RR in the table refers to realization rate.

The high residential RRs reflects the ongoing efforts by the PA to establish a final M&V process. The OBF program followed the same general process but with the smaller population size and more variable savings estimates, different exclusion rules had substantial effects. The divergent estimates of savings illustrate why the CPUC NMEC Rulebook requires PAs to establish M&V plans before programs start and apply them consistently once programs are underway. In this case, the deviation likely reflected the inception of programs predating the Rulebook and the general novelty of NMEC methods for claiming savings.

Table 2 indicates that two of the residential programs (HomeIntel and Comfortable Home Rebates) delivered notable energy savings, with average household electric savings of about 7%. These two programs also produced program savings with high precision, each by chance attaining the goal of 10% error bands at 90% confidence. The third residential program (Home Energy Rewards program) did not produce statistically meaningful savings. It was originally intended to feature contractor-driven EE interventions but had to change its offering to small energy-saving measures due to the COVID pandemic. The non-residential on-bill finance program provided an average 7% reduction in electric consumption but with low precision.

Table 2. PY 2021 Population NMEC Evaluated Electric Savings and Savings Percentages

Program	Gross realization rate	Savings per customer	Percent savings	Relative precision at 90% confidence
P4P - HomeIntel	59%	448	6.7%	10%
P4P - Comfortable Home Rebates	43%	608	7.2%	10%
P4P- Home Energy Rewards	16%	60	1.0%	112%

On-Bill Finance	84%	21,930	7.0%	86%

Though comparing across programs is difficult, the two residential programs that performed well motivated more savings than typically provided by previous EE programs providing similar interventions.[12] The OBF program was more difficult to benchmark, and precision indicates that the result was quite variable.

Site-level NMEC

General description

Site-level NMEC savings efforts are unique to each site, in contrast to a population NMEC approach that applies a consistent method to many sites. Models may include site-specific variables and adjustments to address specific site characteristics. Additionally, identified NREs are addressed uniquely for each site. The site-specific accuracy of the savings estimate is essential as payments for site-level participants are at least 50% performance-based.

Organizationally, site-level NMEC projects occur within the custom projects' ecosystem in California and face possible pre-implementation project review, as do custom projects. While custom projects must meet review requirements to move forward, the site-level NMEC review process is advisory and focused on project feasibility documentation, the baseline model, and the measurement and verification plan for estimating performance-based savings after the 12-month performance period ends.

Another aspect of the project review is the expected useful life (EUL). While NMEC provides a meter-based estimate of first-year savings, it does not offer an alternative path to an EUL. Instead, implementers must develop preliminary estimates of measure-level savings in their pre-implementation documentation to support a savings-weighted average EUL across measures essential to deriving lifecycle savings.

The COVID pandemic has affected almost all the NMEC projects evaluated, both site-level and population NMEC. The pandemic was a particularly challenging confounding effect to address in the site-level NMEC context. The initial reviews occurred during the pandemic, prompting subsequent updates to many projects to better address those issues. Projects shifted baseline periods, and incorporated occupancy variables, among other efforts in an attempt to address the impacts of shutdowns and reduced building occupancy.

The novelty of the NMEC projects has also been a challenge for the PAs beyond the responsibility of conducting the performance assessments. NMEC savings claims are more complicated because they are claimed initially using forecasted savings and then trued up after the performance period ends. The CEDARS database, where PAs report claimed savings, was not designed for NMEC's multiple project-level claims, made a year apart but necessarily linked. In addition, the project documentation supporting claims was uneven and inconsistent. This

[12] Residential direct install programs that DNV evaluated in program years 2020 and 2021 are comparable to population NMEC residential programs, particularly Comfortable Home Rebates, which offered measures. The evaluated savings for these programs were considerably lower than claimed levels and provided savings that were less than 1% of average whole-home energy (kWh) consumption. https://www.calmac.org/publications/CPUC_Group_A_PY2021_Residential_Install_Program_Impact_Evaluation_-_Final_Report_CALMAC.pdf https://www.calmac.org/publications/Group_A_Residential_PY2020_RES_HVAC_Final_Report_CALMAC.pdf

ultimately led to an evaluability study to sort out the available projects and determine if the ex post evaluation was even feasible.

Current programs

Six programs across four PAs contributed site-level projects included in the impact evaluation, which also form the backdrop for this paper. PG&E ran three programs and SCE, SCG, and SoCalREN each ran one. The projects focused on commercial, public sector, and refrigerated warehouse (commercial-like) customers.

Industrial Systems Optimization Program (ISOP) PG&E's ISOP program targets industrial manufacturing and food process customers and focuses on mechanical systems and behavioral, retro-commissioning, and operational (BRO) measures.	**Commercial Calculated Incentives** PG&E's Commercial Calculated Incentives program provides technical assistance and incentive support for commercial projects requiring custom calculations or whole-building NMEC methodologies.	**University of California/ California State University (UC/CSU)** PG&E's UC/CSU program offers incentives for retrofit projects, monitoring-based commissioning, and training for campus energy managers.
Public Sector Performance- Based Retrofit HOPPs SCE's Public Sector HOPPs program targets public sector buildings with stranded savings due to improvement delays or indefinite equipment repairs.	**Commercial Energy Management Technology for Lodging (CEMTL)** SCG's CEMTL program targets small and medium commercial lodging buildings and seeks to provide savings opportunities that encompass the whole building rather than individual rooms.	**Public Agency Metered Savings Program** SoCalREN's Public Agency Metered Savings Program targets public sector stranded savings.

Figure 2. Site-level NMEC programs, PY2020 - PY2021

Evaluability Study findings

As indicated above, the Evaluability Study[13] was motivated because the initial process of identifying a sample for the impact evaluation was not possible, given the nature of the tracking data. The objective of the evaluability study was to identify and characterize site-level NMEC projects, determine whether the projects were ready to be evaluated, and make recommendations for how to improve project documentation and timeliness for future years to improve program evaluability. The site-level NMEC evaluability report provided the following overarching findings.

The site-level NMEC projects were dominated by two customers who in combination represented 65% of projects and 50% of claimed electric savings. This finding motivated efforts to develop as diverse an evaluation sample as possible beyond those two customers by including projects from PY2022. It also meant that final impact estimates necessarily reflected the substantial contributions to savings of those customers.

PY2020 and 2021 NMEC projects were almost universally impacted by the COVID pandemic. The pandemic increased installation times from 3-10 months to more than two years for some complicated projects. The pandemic also affected the modeling process, as COVID is a long and variable NRE. Implementers responded with the inclusion of occupancy variables and the changing of both baseline and performance periods to address this.

[13] https://www.calmac.org/publications/Site-Specific_NMEC_Evaluability_Study_Report_-_Final.pdf

Documentation was varied and inconsistent at every level, from CEDARS reporting practices to variation across reports from a single implementer. The CEDARS reporting guidance was clear enough that at least one PA followed the required steps. However, the challenge of fitting NMEC reporting into a measure-level tracking database not built to accommodate true-ups across multiple years was evident in each PA's reporting. The PAs interpreted guidance regarding "trueing up" savings, updating the initial savings estimates reported at the time of implementation with final estimates based on the full performance period, in varying ways. The evaluators held multiple meetings with individual PAs to identify NMEC projects and link them with their true-up entries. The project documentation, particularly around changes made to the measures installed or the models from plans, was often lacking or limited in detail.

NMEC project savings claims were primarily electric (kWh). Seventy percent, or 36 NMEC projects, made savings claims for energy (kWh) savings only and did not claim demand (kW) savings or natural gas savings. While nearly half of the projects forecasted demand or gas savings, most of these savings were never claimed with only a quarter of projects claiming demand savings and less than 10% claiming gas savings. Both gas and demand (kW) models met basic model eligibility requirements much less frequently than the kWh models leading to foregone claims in gas and kW.

Impact evaluation findings

The impact evaluation included project file reviews, participant interviews, and a final policy, engineering, and model review. While site-level NMEC projects have embedded M&V and have the potential to have realization rates near 100%, different kinds of issues led to lower-than expected-savings. Similar to population NMEC it was necessary to provide two different sets of realization rates because the official total gross realization rate (GRR) did not fully represent program performance. The overall GRR of 70.9% in Table 3 was calculated relative to official claimed savings that were incorrectly trued up in many cases as discussed earlier. Despite this, the overall site-level NMEC first-year realization rates are favorable compared to recent custom evaluation results. The Documentation Realization Rate (DRR) was calculated relative to the savings calculated in project documentation. If reporting had been completed correctly, the official savings would be the same as the savings in the documentation. The DRR would have been the realization rate had the tracking data been reported correctly and is higher than the GRR at 81.2%. By comparison, the PY2020-2021 Custom and Savings by Design electric first-year GRR was 59%.[14] The low SCE GRR is due to mistaken double counting during the ex post true up process, a relatively visible example of documentation issues discussed above that affected all PAs. The DRR shows that SCE's projects over-performed relative to their documented savings expectations. Site-level net-to-gross ratios also compared favorably to the other custom programs.

Table 3. Site-level NMEC program Impact Results

PA	Projects	PA Claimed	GRR	RP%*	DRR	RP%*
PG&E	25	7,994,128	79.7%	±9.0%	79.6%	±10.0%
SCE	14	1,398,331	27.7%	±0.0%	105.3%	±0.0%

[14] https://pda.energydataweb.com/#!/documents/2816/view

SoCalREN	8	1,011,894	78.6%	±19.0%	81.8%	±18.0%
Statewide	47	10,404,353	70.9%	±8.0%	81.2%	±9.0%

Many of the site-level NMEC impact evaluation's findings echoed those in the Evaluability Study. The largest source of savings discrepancies was the incorrect entry of savings claims in the tracking database system. Documentation of project implementation details (installed measures, project timing, program attribution, etc.) and the site's regression-based M&V process were rarely complete or transparent and did not reflect all project changes. In both respects, the complications caused by COVID effects exacerbated challenges already present due to a novel reporting process. In addition, maintenance plans for BRO measures, the requirements for which were well documented, varied substantially in terms of detail and completeness. Perhaps, most surprisingly, some PAs ignored some of the recommendations made during the project review phase. Issues that were not addressed despite being highlighted in project review recommendations included an overlooked cogeneration system and mis-specified EULs. The lack of attention to these highlighted issues led to artificially increased and extended claimed savings. Despite the challenges, most participants expressed high levels of satisfaction with the programs, driven by the programs' technical support and incentives.

Overarching takeaways

The overarching takeaways answer the following questions: Does NMEC induce expanded savings beyond what's likely to be achieved through more traditional programs? Can NMEC methods address non-routine events (NREs) adequately to give confidence in the calculated savings? How do program and evaluation operations have to change in an NMEC environment?

Population and site-level NMEC show potential to expand the scope of EE savings

One of the potential benefits of P4P programs, especially when paired with third-party (3P) implementation, lies in the alignment of incentives with benefits. Tying incentives or technical assistance of programs directly to actual savings may motivate implementers to aim for higher savings throughout the performance period. The population NMEC evaluation results provide an early piece of evidence that this hypothesis may hold. Per-customer savings appeared to compare favorably with similar non-NMEC programs evaluated using equivalent methods. The evaluation identified some possible improvements made by the 3P implementer that may have driven these results– better targeting, improved communication, etc.

Assessing the relative promise of NMEC in a site-level context is more complicated. Whether projects appear to be addressing stranded potential is one potential metric for site-level NMEC success. The evaluation found roughly half of the projects appeared to be addressing stranded potential while there was less evidence of stranded potential for the other half. There were site-level projects with all the characteristics one would expect when addressing stranded potential: old but repairable existing systems or poorly commissioned and scheduled systems that produced dramatic savings when replaced. In contrast, multiple participants explored different program options, suggesting that their focus might not have been solely on addressing stranded potential. There are situations where this could be valid if, for example, meter-based savings were more likely to reflect higher savings than pre-determined deemed savings levels given site-level conditions. In the context of true stranded savings, NMEC should have clear advantages over other programs.

NREs pose significant challenges for NMEC programs

NREs, generously construed as any unaddressed, non-program effects conflated with savings, always present a challenge for NMEC programs. An inevitable asymmetry of information limits the implementer or evaluator's ability to know and account for all of them.

In the population NMEC context, NREs, as the customer-level drivers of aggregate exogenous trends, cause potentially biased results from quasi-experimental methods. Population NMEC inherits the quantification challenges of the ex post evaluation. All parties are at the mercy of the extent to which NREs that the incorporated comparison population cannot suitably address occur. Commercial population NMEC programs face this challenge at a higher level. Commercial customer consumption is more variable both within and across sites than residential customers. Furthermore, commercial programs tend to have smaller populations, so less opportunity to address variability with large numbers. This combination makes it challenging to develop reliable savings estimates with population NMEC methods, even without NREs. NREs occurring among large sites in a population could undermine a whole program's savings.

Site-level NMEC project NREs can be addressed if they are visually or statistically obvious or reported to the implementer. This leaves a substantial gray area of unverifiable changes in consumption that would likely be conflated with the savings estimate. Unreported non-program-related activity occurring during the implementation phase is undetectable from the meter data and is unavoidably conflated with savings. It's effectively impossible to know all meaningful NREs have been addressed under the best, most open conditions. The potential for strategic behavior to affect savings estimates has to be taken seriously.

The COVID pandemic was a significant NRE

The COVID pandemic functioned as an extreme and variably ongoing NRE. For population NMEC programs, the NRE was addressed with the incorporation of non-participant comparison data of similar customers experiencing the same upheavals. As is always the case, if the participants were substantially different than selected nonparticipants, savings would likely be affected, but generally, the ups and downs of COVID were addressed.

Site-level NMEC projects do not have the advantage of matched comparison sites. Despite this, implementers in the site-level programs made substantial efforts to address COVID impacts. The most common approach was adding an occupancy variable based on entry swipes or computer connections, introducing a time-varying variable capable of capturing a portion of the COVID-related occupancy fluctuations. Other sites delayed the performance period until normal conditions re-started. These approaches produced what appeared to be reasonable estimates of savings. Given the challenges posed by the COVID pandemic it was difficult to dispute the results. However, this level of uncertainty would likely be problematic for site-level NMEC in the long run.

Transparent, well-documented, and easily replicable savings estimates are critical for establishing the validity of the process and facilitating evaluation

For both site-level and population NMEC, the performance calculations are designed and produced by the PA or implementer. While the methods applied are not necessarily new, in many cases, the role of modeling, estimating savings, and maintaining files to share with evaluators is new. In the NMEC context, the series of choices that undergird any statistical analysis, from data preparation to model development to results production, requires a level of transparency not historically required of PAs or implementers. Population NMEC methods, in particular, represent data transfer and statistical complexity that is a challenge to pull off, let alone provide ready access to an evaluator to replicate. The variety of coding languages in use adds a potential additional layer of complexity.

NMEC/P4P tracking is different

NMEC is a new addition to California's long-established complex process of documenting, reviewing, and reporting program activities. However, NMEC requires different documentation at every step in the process, is limited by explicit guidance to a different level of review (site-level NMEC vs Custom) and only fits into the reporting structure if produced a certain way (marginal true-ups) and linked across years. These issues will improve over time.

The role for NMEC project evaluations is evolving

Given the embedded performance assessment already attached to each site or program, NMEC carries with it a degree of promise of reducing the role of the independent ex-post evaluation. While this may be possible in the long run, NMEC program evaluation would be essential for the foreseeable future. However, as the role of implementer evolves under NMEC, so might the role of evaluator. Our evaluations took the first steps to explore whether the evaluation role could be relaxed, what that would require from the PA/implementer side, and, if that occurred, what an NMEC evaluation would look like. The preliminary analysis indicates more work is needed to come to a final answer on all three considerations.

NMEC participants indicated high levels of satisfaction with most of the programs

For site-level NMEC, program satisfaction was driven by the programs' technical support and incentives. For population NMEC, satisfaction was driven by energy and cost reductions, convenience, and improved comfort and safety (residential programs) from program installations. High participation satisfaction is a positive sign for the future of NMEC, particularly given the challenges of a relatively new program pathway.

References

CPUC (California Public Utilities Commission). 2020. Rulebook for Programs and Projects Based on Normalized Metered Energy Consumption. Version 2.0, 7 January 2020. https://www.cpuc.ca.gov/-/media/cpuc-website/files/legacyfiles/n/6442463694-nmec-rulebook2-0.pdf

California State Legislature. 2015. *Assembly Bill 802: An act to amend Section 25402.10 of the California Civil Code, relating to energy conservation*. https://leginfo.legislature.ca.gov/faces/billNavClient.xhtml?bill_id=201520160AB802

California State Legislature. *2015. Senate Bill 350: Clean Energy and Pollution Reduction Act of 2015*. https://leginfo.legislature.ca.gov/faces/billNavClient.xhtml?bill_id=201520160SB350

DNV. 2023. *Custom Industrial, Agricultural, and Commercial (CIAC) 2020-21 Impact Evaluation*. EM&V CPUC Group D - Final Impact Evaluation Report. https://pda.energydataweb.com/#!/documents/2816/view

DNV. 2023. *Population-Based NMEC – Program Years 2019 – 2021*. EM&V CPUC Group A - Final Impact Evaluation Report. https://www.calmac.org/publications/Final_Population-Based_NMEC_Program_Impact_Evaluation_Report_-_Program_Years_2019_to_2021_toCALMAC.pdf

DNV. 2023. *Residential Direct Install Program Impact Evaluation - Program Year 2021*. EM&V CPUC Group A - Final Impact Evaluation Report. https://www.calmac.org/publications/CPUC_Group_A_PY2021_Residential_Install_Program_Impact_Evaluation_-_Final_Report_CALMAC.pdf

DNV. 2023. *Site-Level NMEC Evaluability Study, Program Years 2020-2021*. EM&V CPUC Group D - Final Impact Evaluation Report. https://www.calmac.org/publications/Site-Specific_NMEC_Evaluability_Study_Report_-_Final.pdf

DNV. 2022. *Impact Evaluation of Residential HVAC Measures Residential Sector - Program Year 2020*. EM&V CPUC Group A - Final Impact Evaluation Report. https://www.calmac.org/publications/CPUC_Group_A_PY2021_Residential_Install_Program_Impact_Evaluation_-_Final_Report_CALMAC.pdf

IPMVP (International Performance Measurement and Verification Protocol). 2020. *IPMVP Application Guide On Non-Routine Events & Adjustments.* October 2020: EVO 10400 1:2020. https://evo-world.org/images/denisdocuments/NRE_TOC.pdf