BEFORE THE PUBLIC UTILITIES COMMISSION

OF THE STATE OF CALIFORNIA

FILED
05-18-10
04:59 PM

Order Instituting Rulemaking to Examine the Commission's Energy Efficiency Risk/Reward Incentive Mechanism.	Rulemaking 09-01-019 (Filed January 29, 2009)

COMMENTS OF SAN DIEGO GAS & ELECTRIC COMPANY (U 902 M) AND SOUTHERN CALIFORNIA GAS COMPANY (U 904 G) ON THE ASSIGNED COMMISSIONER'S RULING PROVIDING ENERGY DIVISION REPORT AND SOLICITING COMMENTS ON SCENARIO RUNS

Steven D. Patrick
Attorney for
SAN DIEGO GAS & ELECTRIC COMPANY and
SOUTHERN CALIFORNIA GAS COMPANY
555 W. Fifth Street, Suite 1400
Los Angeles, CA 90013-1046
Phone: (213) 244-2954
Fax: (213 629-9620
E-Mail: sdpatrick@semprautilities.com

Dated: May 18, 2010

BEFORE THE PUBLIC UTILITIES COMMISSION

OF THE STATE OF CALIFORNIA

Order Instituting Rulemaking to Examine the Commission's Energy Efficiency Risk/Reward Incentive Mechanism.	Rulemaking 09-01-019 (Filed January 29, 2009)

COMMENTS OF SAN DIEGO GAS & ELECTRIC COMPANY (U 902 M) AND SOUTHERN CALIFORNIA GAS COMPANY (U 904 G) ON THE ASSIGNED COMMISSIONER'S RULING PROVIDING ENERGY DIVISION REPORT AND SOLICITING COMMENTS ON SCENARIO RUNS

I.
INTRODUCTION

On May 4, 2010, Assigned Commissioner Bohn issued an "Assigned Commissioner's Ruling Providing Energy Division Report and Soliciting Comments on Scenario Runs" ("Ruling"). In this Ruling, Commissioner Bohn directed parties to comment on the merits of the various scenarios presented in the 2006-2008 Energy Division Scenario Analysis Report ("Scenario Report"). San Diego Gas & Electric Company and Southern California Gas Company ("Sempra Utilities") hereby respectfully submit their comments on the Ruling and urge the Commission to accept the Sempra Utilities' scenario (provided by San Diego Gas & Electric Company, Southern California Gas Company, Pacific Gas & Electric Company and Southern California Electric Company on April 20, 2010 and incorporated herein by reference) into the record as an appropriate basis for calculating the 2010 true-up claim.

II.
THE ENERGY DIVISION DID NOT INCLUDE THE SCENARIO REQUESTED BY THE SEMPRA UTILITIES

A. The Sempra Utilities Scenario Is Consistent With Commission Policy

The Scenario Report contains multiple earnings scenarios that are derived from the Energy Division's Evaluation Reporting Tool ("ERT"). However, the Scenario Report does not contain the scenario proposed by the Sempra Utilities in their April 20 comments[1], despite the

[1] The Sempra Utilities submitted joint comments with PG&E and SCE.

fact that the Commission requested parties to offer proposals in the *Assigned Commissioner's Ruling on Process for True-Up of Incentive Earnings* (April 8 ACR).[2] The Scenario Report simply ignores the Sempra Utilities scenario and offers only a casual reason for its exclusion.

> ….parties proposed scenarios that would change the parameter values embedded in the ERT application tools. Accordingly, such proposed scenarios are outside the scope of variations contemplated under the April 8, 2010 ACR… (p.2)

The Sempra Utilities' proposed scenario is consistent with Commission policy and can and should be entered into the record for the 2010 final true-up process, notwithstanding the ERT's errant embedded "parameter values" that can be modified to run the Sempra Utilities' proposed scenario. In addition, the Scenario Report must be adjusted to comply with the Commission policy as clearly set forth in D.09-12-045.

The Scenario Report erroneously endorses its Scenario 7 as "the only outcome consistent with current Commission Policy."[3] This is blatantly false. Commission policy, as articulated in the April 8 ACR, requires the Energy Division to develop a range of scenarios to broaden the record. Furthermore Scenario 7 is in complete contradiction with current Commission policy as it endorses a 9 percent shared savings rate, despite the fact that in D.09-12-045, the Commission determined that the 2010 final true-up claim will be calculated determining a 12 percent shared savings rate (see Section III).[4]

The Commission has clearly stated its intention to consider policy assumptions and scenarios other than those presented in Scenario 7 in order to finalize the 2010 true-up process. The April 8 ACR, specifically notes Decision 09-12-045 ordered the parties to engage in settlement discussions and work toward a resolution that was consistent with the policies adopted in D.09-012-045.[5] The Commission specifically set forth "the goal of finalizing the true-up of incentive earnings based upon 'simplified assumptions or metrics not necessarily tied to the

---

[2] Assigned Commissioner's Ruling on Process for True-Up of Incentive Earnings, April 8 2010, O.P.1, p.11
[3] 2006-2008 Energy Division Scenario Analysis Report, May 4 2010, p.1
[4] Decision 09-12-045, p.67
[5] April 8 ACR, p.2

detailed and minute level of calculations embodied in the Final Performance Basis Report for the 2006-2008 cycle."[6] The Sempra Utilities' proposed scenario follows this direction precisely.

Furthermore, the Scenario Report errs in the calculations of SoCalGas' PEB. Specifically, SoCalGas' PEB should not change depending on the inclusion or exclusion of interactive effects. This is consistent with the results shown in the 2006-2008 Verification Report issued through Resolution E-4272.[7]

The Sempra Utilities' proposal is not only consistent with the policy in D.09-12-045 and the April 8 ACR, both of which called for consideration of alternatives to the embedded ERT scenarios, but it also clearly sets forth the basis for the recommendations in the following proposal,[8] each of which is supported directly by Commission decision or otherwise furthers established Commission policy.

In their proposal, the Sempra Utilities first identify errors in the embedded scenarios that must be corrected in order to align with current Commission policy, and then provide the corrections for those errors, including, for example:

- All scenarios should apply a 12% shared savings rate, in accordance with D.09-12-045;

All scenarios should compare 2006-2008 energy savings to 2006-2008 savings goals, as opposed to 2004-2008 cumulative goals, in accordance with D.09-12-045;

- All scenarios should include the net benefits associated with 2006-2008 Codes and Standards activity, in accordance with the Energy Efficiency Policy Manual.

In addition, the Sempra Utilities recommended that each ERT scenario should apply *ex ante* values for Net-to-Gross (NTG), Expected Useful Life (EUL), In-Service Rates (ISR) for upstream-delivered Compact Fluorescent Lightbulbs (CFLs), and Interactive Effects as found in the 2005 DEER. The basis for this proposal includes the following reasons: (1) the Commission has acknowledged that NTG will be lowered as market transformation is achieved and now the

---

6 April 8 ACR, p.3

7 See 2006-2008 Verification Report Tables 30, 31 and 32.

8 The ACR states that "[the April 20] comments should present any supporting basis for parties' positions as to the appropriateness of these scenarios and policy assumptions used to calculate the incentive earnings true-up figure." (ACR, p.9)

IOUs are being penalized for their efforts[9];; (2) updated values for NTG and EUL were not released early enough for the IOUs to make meaningful mid-course corrections during the 2006-2008 program cycle;[10] and (3) credit should not be ignored for CFL installations resulting from 2006-2008 purchases since customers will install these measures in the near future. Even the Draft 2006-2008 Energy Efficiency Evaluation Report acknowledges that "feedback provided at the conclusion of a program cycle is less than desirable, as it may limit timely adaptation of programs based on findings in the field."[11] As such, the Sempra Utilities' proposal is clearly consistent with Commission policy.

Finally, the Sempra Utilities make policy recommendations regarding the incorporation of updated avoided costs adopted in the Commission's April 8, 2010 Decision on Evaluation, Measurement and Verification, D.10-04-029.

In addition to the Energy Division's Report, the Commission stated it has the opportunity to consider "other relevant evidence, as part of the record within this proceeding."[12] Furthermore, the April 8 ACR specified that the Scenario Report was to present a series of scenario runs including "any additional assumptions that may be added based on review of the parties' April 20, 2010 comments"[13]

Despite the fact that the Sempra Utilities' proposal called for correction of errors necessary to align the scenarios with current Commission direction and also set forth other recommendations that are consistent with Commission policy, the Energy Division did not include the Sempra Utilities' proposed scenario. The Commission should enter the scenario proposed by the Sempra Utilities in the record and consider the scenario as an option for finalizing the 2010 true-up process as contemplated in the April 8 ACR. The Commission should also disregard scenarios that do not comply with D.09-12-045.

---

9 D.09-09-047 (at page 38), "This reflects our expectation that our energy efficiency program efforts are in fact resulting in market transformation changing consumption habits and preferences, while acknowledging that measure update in the absence of program support may not be universal."

10 The Commission assumed at that time that the utilities would have "ample opportunity to adjust their portfolios in response to available data." D.07-09-043, p. 171

11 Draft 2006-2008 Energy Efficiency Evaluation Report, p. 124 - Recommendation 7.

12 April 8 ACR, p.7

13 April 8 ACR, p.9

B. The Sempra Utilities' Scenario Produces An Appropriate Earnings Level

1. Utility Estimated Earnings

The Sempra Utilities' scenario represents an appropriate and reasonable assessment of the utilities 2006-2008 program accomplishments for which the Commission can rely upon to determine the 2010 final true-up claim. Because the Scenario Report did not include the Sempra Utilities' proposed scenario, the Sempra Utilities utilized the ERT model to estimate the earnings associated with the Sempra Utilities' scenario. The results, per Commission policy in D.09-12-045, for the Sempra Utilities are presented in Table 1 below:

Table 1: Sempra Utilities' Scenario Results

Utility	Performance Earnings Basis	Earnings Rate	Total 2006-2008 Earnings	Amount Collected From D.08-12-059 And D.09-12-045	2010 Final True-Up Payment
SDG&E	$127.6M	12%	$15.3M	$11.1M	$4.2M
SCG	$105.5M	12%	$12.7M	$7.3M	$5.4M

2. The Sempra Utilities Utilized The Following Process To Estimate Earnings Associated With the Sempra Utilities' Scenario

The Sempra Utilities used the existing framework provided by the Energy Division in the ERT to run the Sempra Utilities' scenario. Because the Sempra Scenario was not a pre-defined scenario contained in the ERT, the Sempra Utilities had to customize the ERT framework to run the scenario. The ERT allows users to easily run some aspects of the Sempra Utilities' scenario, including ex ante net-to-gross ratios, ex ante effective useful lives, and ex post unit energy savings. However, to include ex ante in-service rates for upstream delivered CFLs, the Sempra Utilities had to modify the ERT Input Sheets to reflect the ex ante values, while retaining the ex post installation rate values for all other measures.

In addition to running the above Sempra Utilities' scenario through the ERT, the Sempra Utilities applied an average factor to the resource benefits to estimate the effect of increasing the GHG adder to $30 a tonne. The Commission should direct E3 to conduct such an update and

inform the record of its methodology and results in time for use in the Commission directed settlement conference.

Finally, the Sempra Utilities ran the aforementioned methodology through the Energy Division's ERT developer to receive confirmation that the methodology would produce a credible result.

III.
THE SCENARIO REPORT SHOULD BE CONSISTENT WITH COMMISSION POLICY

A. The Commission Has Adopted A 12% Shared Savings Rate For The 2010 True-Up Claim

In Decision 09-12-045, the Commission directed the 2010 true-up claim be calculated using a 12% shared savings rate. In response to the controversies surrounding the implementation of the RRIM, the Commission correctly determined that the Minimum Performance Standard threshold should be calculated using the ex ante values used to set the goals, and not the updated assumptions contained in the Verification Report. Specifically, the Commission stated:

> "We adjust the shared savings rate to 12% based on the use of the utilities' proposed ex ante assumptions in comparing the utilities' results with the Commission goals."[14] and further clarified this policy would extend to the 2010 true-up claim stating: "…it is reasonable, ***for purposes of both this interim claim and the 2010 final true-up***, to compare those goals with results that reflect the same underlying assumptions used in establishing those goals."[15]

However, despite this clear and unequivocal Commission direction, each of the eight scenarios Energy Division presents in the Scenario Report is calculated using a 9% and a 12% shared savings rate. This is entirely inappropriate as current Commission policy directs the 2010 true-up payment to be determined using only the 12% shared savings rate. As such, each of the sub-scenarios contained in the Scenario Report that utilize a 9% shared savings rate should be disregarded as they violate current Commission policy. This narrows the range of scenarios for

---

[14] Decision 09-12-045, p.3
[15] Decision 09-12-045, p.67 *(emphasis added)*

Commission consideration to only those that actually comply with D. 09-12-045and rely upon a 12 percent shared savings rate.

B. The Commission Should Direct Energy Division To Update The GHG Value To Appropriately Reflect The Utility Portfolios To The State's GHG Reduction Goal.

In the EM&V Decision, the Commission directed Energy Division to update the GHG adder to $30 per tonne.[16] The Draft 2006-2008 Energy Efficiency Section 6—Estimated Reductions touts the contribution of Energy Efficiency

> "One key benefit of the IOU energy efficiency programs implemented in 2006-2008 was the reduction in CO2, NO and particulate emissions that would otherwise occurred in California."

Therefore, it is appropriate to use $30 per tonne to evaluate the benefits associated with the utilities' 2006-2008 energy efficiency portfolios. This would represent the most appropriate estimation of emissions reduction program benefits and should be utilized by the Commission.

C. The Commission Should Direct Energy Division To Update The PEB To Account For 2006-2008 Codes & Standards Activity, Consistent With Commission Direction

The Commission's policy rules for energy efficiency state that: "One hundred (100) percent of verified savings from post-2005 Codes and Standards Advocacy Programs shall count towards the energy savings goals, minimum performance standards and performance earnings basis for the 2006-2008 and 2009-2011 program cycles."[17] The ERT, however, does not currently reflect any net benefits associated with any C&S activity initiated within the 2006-2008 program cycle. For example, the net benefits associated with the 2008 Title 24 Tier II lighting should be included in the PEB calculation according to the Commission's policy rules. In D.09-12-045, the Commission accepted ED's explanation for their non-inclusion of such C&S benefits in the record of that Decision. The Decision concluded that such information was not yet available for incorporation into the Verification Report and that "since the requisite data will be incorporated for purposes of the 2010 true-up, the utilities will be made whole for the effects of any updated data that may change the incentive earnings amount."[18]

---

[16] Decision 04-10-029, O.P. 5, p.56
[17] Energy Efficiency Policy Manual, Version 4.0 (August 2008), p.21
[18] Decision 09-12-045, p.64-65

Such information is currently available and yet, inexplicably, neither ED's Draft Report, nor ED's May 4 Scenario Analysis Report include this information. Omission of this information in derogation of Commission direction misleads the Commission by failing to inform and thereby systematically undercounts the benefits associated with the utility 2006-08 programs. In accordance with the Commission's directive, the Draft Report and embedded ERT scenarios must be modified to include 100 percent of the efficiency savings and net benefits from the aforementioned C&S.

Furthermore, in D.10-04-029 the Commission determined that it is appropriate to count 100% of C&S savings toward the 2010-12 cumulative goals based on its finding that:"…better technical data about savings is now available as compared to when the original 50% determination was made in D.05-09-043, including Evaluation Protocols and elimination of concerns about double-counting and base case forecasts."[19]

While this recent Commission decision was issued in the context of the 2010-12 program cycle, the rationale expressed therein nevertheless supports counting 100% of all C&S savings toward 2006-08 goals as well. This is especially true since the 2006-08 savings will likely be used to measure progress towards the 2010-12 cumulative savings goals.

IV.
ALL CURRENT SCENARIOS ARE FLAWED SINCE THEY RELY ON THE DRAFT EVALUATION REPORT

A. The Draft Evaluation Report's Reliance On Flawed Measurement Studies Is Inappropriate and Unprofessional

The Sempra Utilities' scenarios that were run by the Energy Division fail completely to recognize or address the fundamental, inherent problem with the current ERT process: that is; the ERT scenarios and embedded "parameter values" as currently completed rely solely on the flawed and inaccurate measurement evaluation performed on the utility 2006-2008 Energy Efficiency programs. The Sempra Utilities, along with other parties have, on numerous occasions, illustrated the significant flaws riddled throughout these studies and urged the Energy Division to correct these flaws. The Sempra Utilities have provided detailed comments on both the load impact evaluation studies and draft 2006-2008 Energy Efficiency Evaluation Report.

---

[19] Decision 10-04-029, p.46

The comments are available at, http://www.energydataweb.com/cpuc/home.aspx, and incorporated herein as Attachment and Attachment A.

The most critical of these persistently uncorrected errors are summarized below:

- Sample sizes: In many cases, the change in study focus from evaluating the program in its totality, and to instead focus on High Impact Measures (HIM), made the original sample design unworkable without revising the sample. This resulted in sample sizes that are so small (often below that required in the protocols) that it is difficult or even impossible to make inferences related to the entire population. These small sample sizes lead to inappropriately large error bounds that often even include the ex ante values themselves! As a result, the ex ante values in cases where sample sizes are inappropriately low should be retained, and not rejected as in the ERT.
- Metering/Monitoring Issues: Site specific measurements are subject to significant measurement error depending upon when and how the measurements were taken. Due to EM&V schedules, the Sempra Utilities do not believe that the EM&V consultants sufficiently mitigated for the bias created by this issue. Most measurements are undertaken over very short time periods and then extrapolated to the entire three-year program cycle. This in itself is problematic because all of the seasonal variations cannot be measured, but it is especially problematic during an economic downturn which was being experienced during the measurement time frame.
- Net-To-Gross Issues: The problems with NTGs are particularly disturbing. The 'evaluated' results clearly do not reflect the program's true accomplishments. Energy Markets are not stagnant; they are dynamic and naturally change over time. Net-to-gross studies only take a snapshot (and a blurry picture at that) of one instant in time. As the market changes due to program execution and delivery, the NTG element of the market will change. Program plans and delivery mechanisms are based on how a market looks at the beginning of the cycle. The NTG studies used in the ERT use data collected at the end of the three-year cycle. There is an obvious disconnect with this approach. In fact, a lower NTG could be taken as a sign of program success, not failure. Further compounding the measurement problem is when customer decision makers often take sole credit for the decision made and fail to give sufficient credit to the utility program influencing their decision. No one wants to admit that they were unaware or unfamiliar with something, particularly an action that is environmentally friendly, they should be aware of for their job. Decision makers, particularly several years after the fact are very likely to take credit for knowledge that they actually received directly from the EE program. Again, it is not clear that the studies sufficiently mitigated these problems.
- Non-Compliance with Protocols: In many of the measurement studies used in the ERT, Energy Division failed to follow established and vetted protocols. Examples of this are contained in the billing analysis in the Small Commercial study and also the econometric analysis provided in the Residential Retrofit study. This problem is then compounded when evaluators create ad hoc measures of program success that are clearly outside the Commission-adopted EM&V protocols. For instance, in the upstream lighting evaluation, words like "vulnerability and "leakage" are used to help justify the study's findings. Clearly this is inappropriate.

Except for a scenario that excludes the use of NTG, all of the problems indicated above are ignored in the scenarios contained in the Scenario Report. This clearly is inappropriate. The Commission's objective is to determine, from a comprehensive list of scenarios, a reasonable assessment of the utilities' 2006-2008 energy efficiency programs. The Scenario Report, in its current form, fails to produce a comprehensive list of scenarios, and is biased in favor of scenarios that rely solely upon the most controversial and egregious measurement flaws.

In an effort to correct this bias, the Sempra Utilities provide the Commission with a scenario that will help eliminate some of the major shortcomings contained in the ERT. By mitigating the effect of those flawed measurement studies that are most problematic, the Sempra Utilities believe the Commission can evaluate a more appropriate assessment of the 2006-2008 program cycle.

B. The Evaluation Reporting Template (ERT) Is Strewn With Systematic Flaws And Should Be Corrected

The ERT is strewn with such systematic errors that it significantly misrepresents the Sempra Utilities' accomplishments from the 2006-2008 program cycle. Furthermore, the Draft Report contains non-transparent methodologies that in every case, and despite all logical rationale, drive utility savings and benefits drastically lower. Such errors are manifested in every facet of the ERT and are discussed in detail in each utility's comments submitted on the Draft 2006-2008 Energy Efficiency Evaluation Report on May 17, 2010, and incorporated herein by reference. Some examples are presented below:

- The ERT Does Not Include The Benefits Associated With 2006-2008 Codes and Standards Activity: This section is discussed in detail above.
- The ERT Systematically Reduces All Residential Lighting Program Avoided Cost Benefits To The Lowest Climate Zone Value: The ERT systematically undercounts the avoided cost benefits associated with the utilities largest program. Instead of utilizing the climate zones in which customers actually purchased the efficient lighting products, the ERT instead systematically assigns the lowest valued climate zone. Clearly this is inappropriate and serves no purpose other than to devalue the utilities' largest program.
- The ERT Does Not Properly Account For Emerging Technologies Program Costs: The ERT improperly includes the program costs from the Emerging Technologies program in the net benefit calculation despite the clear direction from Decision 07-09-043 which specifies how the net benefits should be calculated. "With the exception of the Emerging Technologies Program and LIEE, all energy efficiency portfolio costs including associated evaluation, measurement and verification (EM&V) shall be included in the

calculation of PEB."[20] Such an error only serves to undercount utility net benefits by several million dollars.

- The ERT Contains "E3 Calculator FALSE" Errors Which Assign Zero Avoided Cost Benefits To Valid Measure Installations: The ERT contains hundreds of "E3 Calculator FALSE" errors that undercount the energy savings and benefits associated with the utility programs. In short, the E3 Calculator, which calculates the program's net benefits, needs to agree functionally for the calculations to be made appropriately. The ERT contains a bug which is propagated throughout whereby the climate zone, building type, and load shape fields are not in functional agreement and therefore no energy savings or avoided cost benefits are assigned.
- The ERT Alters The Known Location Of Installed Measures By Unknown And Unwarranted Parameters: There are many inconsistencies when comparing the ERT's zip code to climate zone mapping methodology. In the utilities' measure-level reporting, the utilities relied upon the latest zip code to climate zone mapping received from the California Energy Commission (CEC). However, the ERT relies on a completely different and un-documented methodology to map customer location zip codes to climate zones. The ERT instead links the customer zip code to its own lookup table (with no reference to where it is from). Since the CEC is the definitive source on this issue, the ERT should utilize the CEC look up table, as the utilities did, and not an undocumented, unprovenanced source.
- Furthermore, the ERT errs in the calculations of SoCalGas' PEB. Specifically, SoCalGas' PEB should not change depending on the inclusion or exclusion of interactive effects. This is consistent with the results shown in the 2006-2008 Verification Report issued through Resolution E-4272.

With such readily apparent systematic errors, the ERT should not be relied upon in any meaningful way as representative of the Sempra Utilities' 2006-2008 program accomplishments. As a result, all scenarios contained in the 2006-2008 Energy Division Scenario Analysis Report resulting from the ERT are flawed, should be neither expressly nor implicitly relied on, and therefore, discarded.

V.
CONCLUSION

For the above stated reasons, all scenarios contained in the 2006-2008 Energy Division Scenario Analysis Report resulting from the ERT are flawed, should be neither expressly nor implicitly relied on and therefore, discarded. The Sempra Utilities respectfully request that the Commission accept the Sempra Utilities' Scenario into the record and use employ this scenario in its determination of the final 2010 true up earnings payment. The Sempra Utilities' scenario

---

[20] Decision 07-09-043, O.P 2, p.215

represents a more appropriate assessment of the 2006-2008 utility energy efficiency programs. Furthermore, it corrects for the controversial issues that have been well-documented in this proceeding.

Dated: May 18, 2010

Respectfully submitted

By *_/s/ Steven D. Patrick_*
Steven D. Patrick

Attorney for:

SAN DIEGO GAS & ELECTRIC COMPANY and
SOUTHERN CALIFORNIA GAS COMPANY
555 W. Fifth Street, Suite 1400
Los Angeles, CA 90013-1011
Telephone: (213) 244-2954
Facsimile: (213) 629-9620
E-mail: sdpatrick@semprautilities.com

ATTACHMENT

SAN DIEGO GAS & ELECTRIC COMPANY
SOUTHERN CALIFONIA GAS COMPANY

Comments on the April 15, 2010
"Draft 2006-2008 Energy Efficiency Evaluation Report"

1. SoCalGas Seeks Clarification on Tables 11 and 29.

It is not clear to us how the CPUC gets to the "0.30%" value from the 22,212,713,417 Therm total they say they've used along with any of the Therm values they've provided in the corresponding row of Table 11. We do not agree with the base gross number of 22,212,713,417. Below are our recommended changes to the calculations in Tables 11 and 29:

Table 11. SCG Savings Impacts 2006-2008	Annual Impacts		Lifecycle Impact		% sales
	Gross	Net	Gross	Net	Gross
Reported Savings					
Therm	75	67	1,094	975	0.30%
Evaluated Savings					
Therm	54	32	574	344	0.24%
Goal					
Therm		57	*No Lifecycle Goals*		0.30%
Emissions					
Tons of CO2 Avoided	319,344	171,916	3,337,159	207,558	

The base of 22,212,713,417 therms should be replaced by 16,022,250,000 which is the sum of 537,337 MDth for Yr2006, 537,493 MDth for Yr2007and 527,395 MDth for Yr2008, converted to Therms (multiplied by 10,000).

Attached below is the source of our recommendation.

Table11_Scg-Comm ents-R-C&I_Tot...

We have additional comments on the table for your consideration:

- In Table 11, the row labeled "Therm" probably should be labeled "MMTherm" for "million Therms". The "75" value in Table 11 that is labeled "Therm" is the same order of magnitude of the data in Table 1 labeled "MMTherm".

- The corresponding "% Sales Gross" of the "75" value (interpreted as 75 MMTherm) of Table 11 using the 22,212,713,471 Therm total provided by Carmen Best then yields "% Sales Gross" of 0.34; however, the 75 is indicated as an annual number while the denominator (22,212,713,471 Therms) is the sum of data for three years. We believe that the denominator value should also be on an annual basis (i.e., divide it by 3). With this correction, the "% Sales Gross" value becomes 1.01
- We also believe that all the IOU tables (for Electricity as well as for Gas) probably have the same mistakes in methodology.

2. The Verification Report Errs in Using the 2006-2008 Load Impact Evaluations

SDG&E and SoCalGas believe that relying on the 2006-2008 load impact evaluations to update the program achievements due to numerous deficiencies and errors in the reports.

Background

The evaluation protocols generally include the following three components.

- Verification of installation/operation of the energy efficiency measures.
- Estimation of ex post gross savings, with a comparison to ex ante gross savings.
- Adjustment for free ridership through application of the net-to-gross ratio (NTGR).

Errors may occur at any one of the three evaluation stages, although verification is generally fairly straight-forward. Thus, the focus of attention is on estimation of ex post energy savings and the NTGR. The common errors are grouped into seven somewhat overlapping categories (sampling, metering/monitoring, compliance modeling, econometric estimation, the self-report approach, compliance with M&V protocols, and reporting).

Errors Common to Many of the Load Impact Evaluations

Sampling Issues

- Small sample sizes
  - Results in inappropriately large error bounds, that often include the ex ante estimates. Thus, the null hypothesis (IOU achieves claimed savings) cannot be rejected.
  - Difficult or impossible to make inferences to the population.
  - Sample sizes are often less than the protocols require.
  - Measurement error is much more significant in situations in which the sample size is small since each observation is weighted more heavily.
- Nonrandom sample selection
  - Samples based solely on those individuals that were willing to participate.

    - Biased results not representative of the overall population so inferences to populations are not appropriate.
- Out-of sample predictions
    - The sample is representative of the population given a specific program design, time frame, etc.
    - Inappropriate to make inferences to populations outside the specific program (e.g., use information on CFLs to make inferences about LEDs) or the specific time frame (e.g., use information from the 2004 – 2006 program cycle to make inferences about 2006 – 2008 program cycle) without accounting for these differences in the statistical analysis.

Metering/Monitoring Issues

- Site-specific measurements are subject to significant measurement error.
    - Dependent on when the measurements are taken (e.g., during the economic downturn or during the most benign period of the year).
    - Dependent on the length of time the metering/monitoring was conducted (e.g., 4 – 6 weeks)
    - Dependent on who conducted the metering/monitoring and ultimately interpreted the results.
    - The effect of measurement error is exacerbated in situations in which the sample sizes are small.
- Site-specific measurements are subject to the Hawthorne effect.
- Spot measurements cannot capture fluctuations in variables of interest (e.g., related to day, week, month, seasonality or variation in either the energy efficiency application or the local environment. Thus, spot readings cannot be used reliably for extrapolation.
- Extrapolation from logger results is problematic because they are operational for insufficient periods of time. This creates two obvious problems.
    - A short installation/operation period (e.g., two weeks) will be unlikely to capture the fluctuations in use patterns that occur over the year that correspond to variations in weather, daylight hours, macroeconomic conditions, etc. In essence, an analysis based on one to two weeks of logging data is exactly akin to basing the analysis on spot readings, which cannot be used reliably for extrapolation were defined by the study authors to be woefully insufficient.
    - Appropriate sampling design must consider both the number of sites and the time period over which the loggers are installed/operational (e.g., achieving a 10 percent precision with a 90 percent requires in excess of 29 weeks of logger information).
- Significant measurement error and imprecision in the measured/monitored variables of interest (e.g., estimated hours of use) undermines the estimation of ex post savings.
- Program evaluators make biased interpretations/decisions (e.g., one-sided trimming in the evaluation of Small Commercial).

Compliance Modeling Issues

- Modeling is based on metered/monitored data that is measured imprecisely.
- Models perform poorly compared to actual usage (e.g., see Residential New Construction).
- Model results are incomplete (e.g., the treatment of interactive effects is undefined).
- Model results are inappropriately used (due to small sample size, large error bounds, etc.) to aggregate up to produce population savings estimates.

Econometric Estimation Issues

- Econometric analysis is severely outdated
- Methods used are not scientifically credible. A litany of problems would include the following.
  - Lack of theoretical justification for modeling design (i.e., *ad hoc*).
  - Data sets are inappropriately small.
  - Hypothesis testing is conducted inappropriately (e.g., insignificant variables are eliminated from estimated equations or step-wise regressions are used).
  - Empirical results are not subject to robustness testing or sensitivity analysis. Such testing would include, but not be limited to, determining the relative importance of alternative sets of independent variable sets, alternative functional forms, alternative estimation methods, measurement error, outliers, and influential observations, etc.
  - Econometric results suggest model misspecification (e.g., see Commercial Steam traps) such as omitted variable bias.
- Econometric modeling is based on incomplete knowledge of the relevant literature or an incomplete understanding of the methods involved (e.g., conjoint analysis or revealed preference modeling).
- Data sets created for 2006 – 2008 program evaluations are inappropriately combined with data from previous evaluations without justification or an understanding of the relative impact on ultimate results.
- Econometric results produce inconsistent and counter-intuitive results that are offered without explanation.

Self-report approach (SRA) Issues

- The method is plagued by numerous inherent biases. A list would include the following.
  - Self-report bias in which the respondent attempts to please the surveyor or to create the appearance of "socially acceptable behavior." This bias is especially relevant when being questioned regarding socially acceptable activities (e.g., quitting smoking, recycling, adopting energy efficiency, etc.).
  - Starting point bias in which the final respondent answer is closely tied to the suggested starting point (see Codes & Standards evaluation).
  - Non-random selection bias, in which respondents self-select into the survey or are only those individuals willing to participate.
  - Decision-maker bias, in which the survey is conducted with a single individual but decisions are not made in this manner (rather the process

includes many disparate influences) or the appropriate decision maker cannot be identified (e.g., no longer works for the entity).
    - Program-influence bias, in which the NTGR for non-residential applications seems to be limited at upper end due to the "program influence" question. Specifically, it seems that this question inappropriately anchors the respondent to the 50 – 50 attribution between the "program" and "other factors."
    - Program evaluator bias, in which the evaluators make post-survey interpretations/adjustments regarding respondents' answers that are either *ad hoc* or inappropriate (e.g., converting "Don't Know" into a numerical value).
    - Survey implementation bias, in which surveys are conducted by overly-experienced survey personnel that achieve a pre-determined result.
    - Reality bias, in which SRA results are inconsistent with actual data (e.g., sales information, prior lack of action by participants, stocking practices, etc.).
- There is a general lack of supporting evidence to calibrate the SRA findings.

Compliance with M&V Protocols

- Protocols are often not followed or not followed to completion (e.g., billing analysis in Small Commercial or the econometric analysis in Residential Retrofit).
    - Failure to think through the evaluation process prior to beginning the evaluation effort or "unforeseen" problems undermine the analysis.
    - Less rigorous analysis completed.
    - Evaluation effort produces less than expected at a cost per result much higher than expected.
    - Lack of supporting evidence reduces the validity of the work presented.
- Evaluators create *ad hoc* measures of program success that are outside the M&V protocols (e.g., "vulnerability" and "leakage" in the Upstream Lighting evaluation).

Reporting Issues

- There is a general lack of information provided that is necessary to assess the validity of the study (e.g., econometric estimation missing information includes, but is not limited to, R-square vales, number of observations, standard errors, etc.
- There is unexplained variation across IOUs on such variables as hours of use, NTG ratios, etc.
- Specific IOUs seem to perform consistently worse than others without explanation.
- NTG ratios that vary widely (see HVAC HIM and Specialized Commercial Evaluation) without explanation.
- There is insufficient evaluation of the likely effect of sampling error, measurement error, modeling error, etc.
- Inappropriate conclusions are drawn (e.g., inferences from small samples, inferences from data measured with error, inferences that use out-of-sample predictions, etc.).
- In appropriate attribution (e.g., in every case in which ex post savings differs from ex ante savings it is assumed that the ex ante figures are incorrect yet the

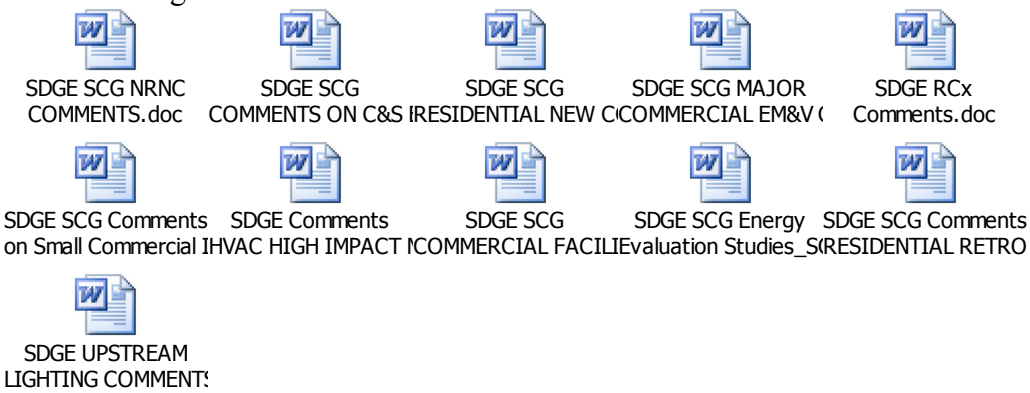
evaluation, whether it be data collection, modeling, estimation, etc., is fraught with errors and is likely the reason for the difference).

- Evaluation results are inappropriately used in policy making.
- Recommendations for improving performance, especially as it pertains to inter-IOU performance, are often missing.

Conclusion

The Load Impact studies posted to date have numerous problems that undermine the relevance of the findings. In fact, the problems are so large that it appears that all the evaluation results lack the rigor necessary to yield results anywhere near to being usable for program evaluation and subsequent design.

The following are the comments SDG&E and SoCalGas filed on the various studies.



ATTACHMENT A

Table 11

Extracted from Data Request Response provided to Richard A. Myers, CPUC's Energy Division, on July 2nd, 2009.

SOUTHERN CALIFORNIA GAS COMPANY
ANNUAL AVERAGE DELIVERED NATURAL GAS VOLUMES
IN MILLIONS CUBIC FEET PER DAY

	Year	Delivered Volumes (MMcf/d)	Est'd. Delivered [6/] MDth	(Dth/Mcf)	Est'd. Delivered MDth	FERC-Form 2 (MDth Total Deliveries)	Est'd. Diff. Rel. to FERC Form 2 MDth	
Residential [1/]	2006	678	254,944					
	2007	673	253,137					
	2008	659	248,430					
Core Commercial/Industrial [1/]	2006	301	113,183					
	2007	312	117,353					
	2008	300	113,094					
Noncore Commercial/Industrial [2/]	2006	450	169,210		**537,337**	**631,936**	**-94,599**	(Totals here are for "R+C+I" for respective years 2006, 2007 and 2008.)
	2007	444	167,003		**537,493**	**654,600**	**-117,107**	
	2008	440	165,871		**527,395**	**641,463**	**-114,068**	
Electric Generation [3/]	2006	769	289,162					
	2007	849	319,336					
	2008	907	341,921					
Wholesale [4/]	2006	394	148,153					
	2007	406	152,710					
	2008	422	159,086					
TOTAL [5/]	2006	2,592	974,652	1.0302	**974,652**	**974,918**	**-266**	
	2007	2,684	1,009,540	1.0305	**1,009,540**	**1,009,884**	**-344**	
	2008	2,728	1,028,401	1.0300	**1,028,401**	**1,028,189**	**212**	

NOTES:

1. Residential includes master metered & CARE. Core Commercial & Industrial includes rate category G-10, Gas Engines, Gas Air-conditioning and NGV load.
2. Noncore Commercial & Industrial includes rate category G-30 but ***excludes*** refinery CoGen; and ***excludes*** EOR Cogen, but includes EOR Steaming
3. Electric Generation ***excludes*** electric generation and cogeneration load of SDG&E and Southwest Gas
4. Wholesale load includes DGN. Wholesales load in April 2005 includes SDG&E's large billing adjustments for the prior years.
5. Exchanges, UAF and Company use volume are not included.
6. Calculated as: "MDth" = "MMcf/d" x (Days/Yr) x ("Dth/Mcf"), for respective years

SDG&E SCG Non-Residential New Construction Program Impact Evaluation

COMMENTS OF SAN DIEGO GAS AND ELECTRIC COMPANYAND SOUTHERN CALIFORNIA GAS COMPANY

NON-RESIDENTIAL NEW CONSTRUCTION PROGRAM IMPACT EVALUATION

1) The study does not explain how the sample was stratified across the service territories. It is not clear why the sample percentage relative to the IOU program population and gross KWH savings for SDG&E would have a relatively higher number of sample points compared to SCE and PG&E. It is interesting that the final sample selection results in relative precision rates for PG&E and SCE that are at least twice that of SDG&E. This would imply that if the results were applied on a utility-basis, that each utility's results do not have the same level of reliability. This discrepancy should be adequately explained in the study and a thorough discussion how this disparity does not ultimately affect the reliability of individual IOU results. On the other hand, because the study does not explain how the results are to be used for the final verification report, **the evaluation study should clearly discuss the applicability of the results when used either on a statewide basis or on an IOU-basis.**

	Percent of Sample Relative to IOU Program Population Count (Table 1-2)	Percent of Sample on a Energy Savings Basis (Table 1-3)	Relative Precision of Ex Post Gross Savings (MWH) (Table 1-3)
PG&E	19.5%	41.3%	11.5%
SCE	26.6%	48.7%	9.1%
SDG&E	57.1%	82%	4.2%

2) Overall, the evaluation suggests that the non-residential new construction program is a success. However, the SDG&E and SCG results do not have the same realization rates as PG&E and SCE. There is no explanation of why SDG&E and SCG programs do not achieve at the level of SCE and PG&E's programs. Is it perhaps a result of the sample design issue raised in the first comment or are there legitimate program design or program participant differences (e.g., mix of building types, greater free-ridership, modeling issues) or some other factors that account for these differences? **The evaluation report should provide explanations for their findings.**

3) The Joint Utilities find with respect to the natural gas estimation the same sampling issues brought up in Comment 1 related to the electric savings are the same if not more egregious Furthermore, the evaluation study (at page 10) notes that the "basic gas savings estimation techniques used by the utilities needs to be fundamentally re-examined. The lack of relationship is so poor that error bounds and relative precision have essentially no meaning." As a result, in table 1-5, no relative precision is shown. The Joint Utilities believe this is because the study

was designed based on electric savings and no separate sample designed for natural gas savings. There is no analysis to show that the electric sample suffices as a substitute for a natural gas sample.

	Percent of Sample Relative to IOU Program Population Count (Table 1-2)	Percent of Sample on a Energy Savings Basis (Table 1-5)	Relative Precision of Ex Post Gross Savings (therms) (Table 1-5)
PG&E	19.5%	29.3%	0%
SCG	26.6%	76.9%	0%
SDG&E	57.1%	82.4%	0%

The Joint Utilities conclude from this that it is impossible to draw any inference about how close the utilities came to achieving their target natural gas goals. It indicates the modeling is so flawed that it produces no usable information. **If this is indeed correct, then the Joint Utilities recommend that the gas model, and this portion of the results, *not* be accepted as reliable or used for updating DEER or used to measure utility performance in the ERT and VRT process.**

4) On a measure-specific level, the error bounds for some of the measure savings calculated exceed the point estimate for ex post savings. This is evident in Tables 3-12, 3-14, 3-17 and 3-20, primarily for SCE and PG&E-specific measures. Again, this raises the question regarding the validity of the sample design and ultimately the reliability of study findings for gross electric and natural gas savings.

5) In the Study Appendix, there is documentation provided for each of the study participants. In the case of participant G120006, it states that the customer did not provide requested data to adequately evaluate the site. In spite of this, the evaluators proceeded with their estimation using a different measure to determine its realization rate, resulting in an extremely low realization rate of 12%. The study does not explain either, why this is not an outlier and therefore removed from the study or why the substitute measure provides adequate representation and therefore the results are useable. **The study should explain their treatment of outliers or non-responsiveness if they continue to use these sample points.**

6) It is interesting that a program that is extremely complicated and requires several decision making points over potentially several years (and as noted in the study over several program cycles) with T24 code changes potentially impacting the building project, that the study would resort to depending on a mere 3-question self report procedure. In addition, the methodology documentation (in Appendix D) does not adequately explain the statistical basis for aggregating the responses from the 3 questions and coming up with a reliable statistic to be used in

measuring free-ridership (at page4). Furthermore, the scoring methodology for the questionnaire is on a relative not absolute scale, discrete versus continuous scale and therefore the results could vary depending on how the scale is determined. Again, there is no explanation why the choice of the scale is adequate. For example, participant D63257 has a NTGr of 87% based on the 0-6 scale, would the NTGr be different if scale had been 0-10 or 0-100? A sensitivity analysis would have been useful to make this determination and provide the program with a reliable estimate of free-ridership.

The following provides an example of why the NTGr self-report and the interpretation accorded the response is questionable and may indicate an inherent bias in the evaluation. The discussion for Participant G70010 goes as follows:

> "Discussions with the facility owner indicated that while SBD was influential in the implementation of the measure. The site contact said the Program is particularly useful because it verifies the validity of energy efficiency ideas generated within the company. The site contact stated that SBD representatives are helpful because they either "confirm what we think [about an energy efficiency project], refute it, or cause us to rethink it". The site contact also indicated that in the absence of the program, there is a "50/50" chance they would have gone with different equipment because the installed ovens were very expensive. The site contact's combination of answers yielded a free ridership score of 3.0 out of 6, or 50.0% free ridership. Site net savings were therefore 50.0% of gross savings."

It appears that the participant attributed many benefits to participating in the program and working with program staff. However, the one question that the interviewee responded, that it was "50/50" chance they would have gone with different equipment because the installed ovens **were very expensive"** resulted in this project having a NTGr of 50%. How does the methodology account for the extremely positive responses and the one indecisive response of 50/50 resulting in 50% NTGR?

It is problematic that when the interviewee does not know or does not remember or is uncertain in Question 23, the participant gets a score of 0. This is common for projects with extended timelines. A score of "0" is a significant in that the most you can score if they responded 10 to both Question 22 and Question 24 is 66%. The study does not explain why this adequate treatment for a response that is more akin to a non-response. If the respondent selects a Q23 response that gives them a score of 1, the total score becomes 83%. This is a huge step function and the program would never get a NTGr between 66% and 83%. What exacerbates the problem is that this is done on a measure basis, if the respondent installed several measures, how does the methodology do internal consistency validation? Below is the Question 23:

Q23.
How did Savings By Design influence the implementation of (maximum of 2 points) Open ended Question that is coded:

1 = SBD had no influence on this measure 0 points
2 = SBD representative first suggested/introduced measure 2 points
3 = SBD performed simulations and/or design analysis 2 points
4 = SBD incentive made this measure an “easier sell” 1 points
5 = SBD incentive helped the measure meet investment criteria 2 points
6 = Prior SBD projects have had success with this measure 1 points
7 = DK, Not Certain, Can’t Remember 0 points
50= other individually assessed

The Joint Utilities, therefore, conclude that the NTGr estimate or the net savings impacts may not be as reliable as the results show and the Joint Utilities do not recommend the use of these results.

Overall, the Joint Utilities find that the sample design and NTG methodology raise significant questions regarding the validity of the ex post results. The Joint Utilities recommend that these issues be addressed and the results re-evaluated prior to using any of the study results for DEER updates or for the Verification Report.

SDG&E SCG Comments on Codes & Standards Evaluation

COMMENTS OF SAN DIEGO GAS AND ELECTRIC COMPANYAND SOUTHERN CALIFORNIA GAS COMPANY

CODES AND STANDARDS EVALUATION

The joint utilities have issues with three different areas of the report:

1. Elimination of Savings Due to Federal Standards

The study notes that "the evaluated potential savings for two standards were estimated to be zero because federal standards established a new baseline: Tier 2 Large Packaged Commercial Air Conditioners and Pre-rinse Spray Valves." The elimination of these savings resulted in the study concluding that the evaluated natural gas savings were "a little less than half" the Savings Estimate Spreadsheet (SES, developed by HMG)[1] estimate.

The impact of eliminating the potential savings for these items implies that the utilities will not get credit for any savings related to the installation of Tier 2 Large Packaged Commercial Air Conditioners and Pre-rinse Spray Valves.

One issue that arises here is whether the Federal standard was a direct consequence of the codes and standards adopted by California. If this is indeed the case then attribution to the IOUs needs to be re-examined. For example, the Federal Energy Program website, http://www1.eere.energy.gov/femp/technologies/eep_low-flow_valves.html, notes that in the case of pre-rinse spray valves, research and product testing was conducted by the Food Service Technology Center (FSTC) with funding by California utility customers. Specifically, the website notes, "The Food Services Technology Center (FSTC) has an online database of pre-rinse spray valves that have been tested in accordance with ASTM 2323-03. FSTC in San Ramon, California conducted research and product testing on pre-rinse spray valves. FSTC is funded by California utility customers and administered by the Pacific Gas and Electric Company under the auspices of the California Public Utilities Commission."

The Joint Utilities believe that this is a major flaw in the findings of the study and therefore should be corrected in the final study. To not address the issue could result in the study being unreliable and not provide accurate estimates of program results and therefore should not be used in the ERT and VRT process or be used for DEER updates.

2. Elimination of Savings for Time Dependent Valuation Standards

[1] Heschong Mahone Group, Inc. 2005 (Revised November 1). *Codes and Standards Program Savings Estimate for 2005 Building Standards and 2006/2007 Appliance Standards.* Prepared for Joint Utilities.

The report notes that, "no potential energy or demand savings were estimated for the two Time Dependent Valuation (TDV) standards because the evaluation team could find no evidence that the standards would drive building design toward more on-peak savings to meet the requirements of the standards."

In Appendix J, the authors of the study discuss their rationale for eliminating the savings associated with TDV standards. First they note "the original (TDV) savings estimate was calculated by assuming that in residential buildings the standards would lead to installation of air conditioners with an EER of 11, rather than an assumed base level of 10." Based on field data used to determine compliance with the TDV standard the authors found that 54% of the air conditioners installed had an EER of 11 or higher. Furthermore, the authors found that higher EERs were being used to meet the TDV requirement, but in as little as 18% of the cases. It should be noted however, that the authors also found that just less than 50% of respondents said the approach used to meet the TDV standard varied or they did not know what approach was used. Thus, it is quite possible that EER's were being used in **significantly** more than 18% of the cases to meet the TDV standard. The joint utilities believe this is a major error. This needs to be corrected in the study.

3. General Service Incandescent Lamps, Tier 1

The authors of the study conclude there are no savings associated with the General Service Incandescent Lamps, Tier 1 standards. In Appendix I, the author's justify the no savings claim by noting the following, "The data from our appliance standard compliance research in 2008-09 (discussed later) provides supporting evidence for the observation that producers were likely to increase output without reducing wattage. We found, more than two years after the Tier 1 standard went into effect, that between 56% and 63% of bulbs sold had the same wattage as the traditional wattage category they were in; that is, 75 watt bulbs were still being sold and, to comply, their lumen output was increased. There are no unit savings associated with these bulbs. As noted above, the adopted Tier 1 standard was less stringent than the proposed Tier 1 standard. Based on our analysis, the result of the two outcomes—unchanged wattage with reduced lumens and a less stringent standard—was that there were essentially no energy savings from the adopted Tier 1 standard."
If approximately 60% of bulbs provided no units savings, what about the other 40%? Even with the reduced standard, some savings should be attributable to at least the 40% of bulbs with unit savings. The fact that the authors of the study attributed no savings to General Service Incandescent Lamps, Tier 1 standards resulted in a loss of 20.7 GWh/year of claimed savings by the utilities (see Table 36). This should be corrected in the final report.

4. Sampling Issues:

In general, the study is based on very small sample sizes, *ad hoc* weighting, and *ad hoc* assignment of relevance to survey responses. Consider the following examples.

- The first-round NOMAD analysis for commercial dishwasher pre-rinse spray valves is based on responses from eight individuals, two of which were characterized as outliers. However, the criteria utilized to determine outlier status (their responses seemed unreasonable) is *ad hoc*. How would the results differ if alternative criteria were used? Also, the Bass curve fit produces estimates of NOMAD that are significantly different from the prior study. Are these results robust to alternative functional forms?

- In the area of surveying experts about stakeholder allocation in attribution, most standards were analyzed with three to five responses from experts (small sample size issue), weighted by "involvement of the respondent in the development of the standard and his or her presumed knowledge of stakeholder activities." The determination of the weights seems completely *ad hoc.* In addition, strong agreement between experts about resource allocation is determined by a standard deviation less than 0.5. This value, as are the values that determine moderate and low agreement, is also completely *ad hoc*.

Attribution was determined by a panel of independent evaluators. The discussion began with a score proposed by the facilitator. There is a large potential for starting point bias in this type of protocol (i.e., the final score is essentially determined by the initial proposed score). Were alternative protocols considered? Was there an attempt to investigate starting point bias? How often and by how much did the final score deviate from the initial score?

The author's also note that in many cases, their final sample was based simply on who was willing to talk to them or provide them access to their files or inventories. Thus, the sample was not randomly chosen and potentially suffers from severe non-random sample selection bias. Once again this raises questions about the ability of the authors to draw any meaningful conclusions about the population. The Joint Utilities believe that this is a major flaw in the findings of the study and therefore should be corrected in the final study. To not address the issue could result in the study being unreliable and not provide accurate estimates of program results and therefore should not be used in the ERT and VRT process or be used for DEER updates.

Conclusion/Recommendation

Tables 36 through 41 present the final evaluated savings estimates for the C&S Program statewide for the period 2006 through 2008. The tables present the evaluated savings in columns 1 through 3 for years 2006, 2007, and 2008, respectively and the SES for 2006 (which corresponds to the utilities claimed savings) in column 4. While the tables are illuminating, they are hard to interpret since no confidence intervals are provided for the evaluated savings. Specifically, consider the following methodology for determining whether utilities have met their claimed savings. Make the SES for 2006 the null hypothesis (i.e. the hypothesized level of savings). Next, calculate the evaluated savings and the standard deviation of evaluated savings. If a 95 percent confidence interval for

evaluated savings includes the 2006 SES value then we fail to reject the null. In other words, we can not reject the null hypothesis that evaluated savings are the same as claimed savings. In that case, the utility should credit for its claimed savings. Given the methodology employed in the study and the issues identified above, it is our belief that the standard errors associated with evaluated savings will be large. Consequently, the joint utilities believe it is unlikely that the authors could reject that evaluated savings are statistically different from claimed savings. If this is not done, the Joint Utilities believe and would recommend that this study *not* be accepted as reliable or used for updating DEER or used to measure utility performance in the ERT and VRT process.

SDG&E SCG Residential New Construction Program Impact Evaluation Comments

RESIDENTIAL NEW CONSTRUCTION PROGRAM IMPACT EVALUATION
COMMENTS OF THE
SAN DIEGO GAS AND ELECTRIC COMPANYAND THE SOUTHERN CALIFORNIA GAS COMPANY

San Diego Gas & Electric Company and Southern California Gas Company (Joint Utilities) submit the following comments on the draft Residential New Construction (RNC) Program Impact Evaluation. The following are the major concerns of the Joint Utilities regarding the study:

1) The result from the compliance models, relative to the metered data, suggests that the models are highly inaccurate and subject to relatively wide variation, much of which is unexplained. The authors of the study note this directly on page 1-20. Specifically, they state:

> "The metered data indicate that the MICROPAS compliance software, on average, overestimates the amount of heating energy consumed at a site and underestimates the amount of cooling energy consumed. Because the baseline for the utilities' ex ante estimates of energy savings is a home that meets the minimum of California's building code, and the compliance software that is used to estimate baseline energy use for this purpose does not accurately predict energy use, ex ante energy savings from the utility tracking databases may not reflect actual savings."

This raises serious questions concerning the reliability of the study. Can this be corrected? If not, should this study continue to be used in the VRT process?

2) Samples sizes are extremely small, which raises the question of whether any real inference can be drawn about the population. This issue is heightened when small state-level sample sizes are used to make inferences about specific utility service territories and/or specific climate zones. Moreover, because the study was forced to reduce the sample size due to the economic downturn, the target precision for the meter-to-model ratios were significantly reduced. Specifically, the original plan (prior to the reduction in sample size) called for target precision for the calculation of meter-to model ratios of 25% at the 90% level of confidence (90/25). However, due to the small actual sample sizes the achieved level of precision for the meter-to model ratio was only 70% to 18% at the 90% confidence level. **Thus, the small sample sizes added a significant amount of uncertainty to the meter-to-model ratios and cast doubt on the validity of the results. Do the authors agree with this interpretation? If so, should the study continue to be used in the VRT process?**

The application of the non-participant meter-to-model ratio to participants undermines the notion that participants are different from non-participants. That is, one would expect that participants have characteristics and behaviors that separate them from non-participants. If that is not the case, how do we explain participation for some individuals

and not others? This casts doubt on the results since there is no data available to inform us as to whether this is a valid assumption or how the results would vary with alternative assumptions. Again, should the study continue to be used in the VRT process?

Conclusion: The concerns and questions raised above put considerable doubt on the accuracy and reliability of the load impacts estimated in the draft report. If these concerns cannot be addressed by the final report, the Joint Utilities recommend that this study *not* be accepted as reliable or used for updating DEER or used to measure utility performance in the ERT and VRT process.

SDG&E SCG Major Commercial Impact Evaluation Comments

SAN DIEGO GAS AND ELECTRIC COMPANY
AND
SOUTHERN CALIFORNIA GAS COMPANY

MAJOR COMMERCIAL IMPACT EVALUATION
COMMENTS OF THE

The joint utilities have several concerns with the Major Commercial Program Impact Evaluation. The most important of these are the following:

Overall Concerns

1) The realization rates in the study are provided for each stratum within each program. Disaggregating in this manner results in very small sample sizes at the individual stratum level (sample sizes that range from a low of one to a high of nine) – very similar to problems discovered in almost all of the Impact evaluations being reviewed. As a result, the 90% confidence intervals, because of the small sample sizes, would tend to be very large. These limited sample sizes severely undermines the relevance of the findings. Even at the aggregated level (i.e., averages over all strata), the realization rates tend to be very large and often at the 90% confidence intervals include the value of 0.9 or higher. To illustrate these points consider the realization rates obtained for the SCG3513 program in the table below.

Table 1: SCG 3513 First Year Gross Savings Parameters by Stratum with Domain Statistics

		Measures		Stratum Boundaries Ex Ante Savings (Therms)		Gross Ex Post Unit Energy Savings1			Gross Savings Realization Rate1		
Program ID	**Stratum**	**Population**	**Sample**	**Lower**	**Upper**	**kW**	**kWh**	**Therms**	**kW**	**kWh**	**Therms**
SCG3513	1	623	4	646	27,758	0	0	14,928			1.56
SCG3513	2	54	2	29,125	102,794	0	0	13,753			.38
SCG3513	3	20	6	110,839	286,680	0	0	47,213			.33
SCG3513	9	12	7	339,670	2,177,246	0	0	228,180			.24
SCG3513	Excluded	411	0	0	578						
SCG3513	**All Sampled**	**709**	**19**	**0**	**2,177,246**	**0**	**0**	**17,133**			**.62**
					Statistics						
					Standard Error			**7,519**			**.27**
					90% Confidence Interval			**12,369**			**.45**
					Relative Precision			**.72**			**.72**

1 Stratum level results are based on sampled sites only.

Note that standard errors and 90% confidence intervals are not reported by individual stratum. However, the sample sizes in every case are extremely small,

ranging from two to seven. In many of these cases it appears likely that the 90% confidence interval will be large, implying that the interval may well include a population value of 0.9 or higher for the realization rate. Also note that even for the aggregated results (final row labeled All Sampled) the realization rate estimate is 0.62 and 90% confidence interval is 0.45! Thus the 90% confidence interval includes a population value of the realization rate of 1.07 (.45+.62). In this case, the null hypothesis that evaluated savings are the same as claimed savings *cannot* be rejected and the utility should receive credit for its claimed savings.

2) Some of the measures with the highest claimed savings have very small samples sizes. These measures were supposed to be sampled with certainty (stratum 9) but evaluators fell short. Examples include SCG3513 with a population of 12 and a sample of seven (see table above), SDGE3010 with a population of three and a sample of only one, SDGE3025 with a population of six and a sample of two. Again, such small sample sizes put into real question the validity of the results of the evaluation and these studies should either be rejected or redone using proper evaluation protocol.

3) The free ridership estimates for these programs use the self report approach (SRA) for non-residential programs. NTG rates are provided for each stratum in each IOU program. Again, this disaggregation results in extremely small sample sizes in each stratum. Consequently, the relative precision values are relatively large. Also, the SRA is beset with significant survey issues such as self-report bias, recall error, failure to survey the appropriate "decision-maker," etc. Identification of the relevant "decision-maker" may be especially important for large commercial projects. Specifically, it is likely that a "decision process" that involves multiple individuals representing different divisions of a large firm is used to evaluate energy efficiency projects rather than the decision being made by one individual. In this case, the survey team will likely interview the advocate for the project, who, of course, suggests that the project would have been completed in the absence of the program. However, this ignores the various hurdles that had to be overcome (e.g., financing, prioritization, etc.) and it is inconsistent with reality. How does the evaluation team account for these issues in its calculation of NTG??

SDG&E Program Specific Concerns

M01180 ESB project 06-03-002

Controls of Garage fans utilizing CO sensors – AC Energy
As the evaluator notes "there is no precise evidence of baseline operations", and based on discussion with only one person, he assumes that the fans were not operating 100% of the time as was assumed by the project sponsor. The evaluator assumes that the supply and exhaust fans were operating only 40% and 44% respectively. The IOU position is that this pre project fans run time appears to be low. The parking garage is below grade and does not appear to have sufficient natural ventilation. Since the post data indicate that the

fans were on 27% of the time it is logical to think that the pre run time is at least double that value (54%) assuming 50% of the sensors were faulty.
The measure load by the evaluator is 430 kW compared to 612 kW which was arrived by the vendor. The IOU does not doubt the readings but as was stated, the facility has been working on the fans controls to reduce the fans energy further.
One of the findings is that the project cost appears to be excessive for the size of the project and there was no invoice to the customer in the file. The IOU position is that ESB program as whole do comply with the TRC guidelines set by the PUC.

M01181 ESB 06-03-002

Controls of Garage fans utilizing CO sensors – AC Energy
The evaluator acknowledge that the fans were operating 100% of the time @ 2/3 rd of maximum speed prior to installation. The speed determination is based on estimate following the interview of the site contact. As noted the VFD failed providing no readout of the motor speed and there were no records of the manual setting of the drives.
The assumption by the evaluator yielded a fan power draw during pre construction of 28.3 kW and post construction post construction of 15.8 kW vs. the IOU estimate of 79 kW fan motor power draw for both periods. The operating times of the fans are in agreement. As the power draw measurements show the average power draw according the evaluator at 54 Hz is 64 kW which for 60Hz brings the power draw to 73 kW as assumed by the utility. The IOU does not agree with the 2/3 of maximum fan speed as a baseline.

M01157 ESB 06-03-002

Controls of Garage fans utilizing CO sensors – AC Energy
According to the evaluator the power draws of the fans appears to be 1/3 of name plate values. Such low power draw will result in low motor efficiency and power factor. The IOU assumed a total fans motors power draw of 221 kW vs. the evaluator position that the combined power draw is 70.9 kW. It is postulated by the IOU that the fans speed may have been altered since the installation by changing the pulley ratios between the motors and the fans. It is hard to understand why such large motors were installed on this system when they are way under utilized.

M01073 ESB 06-03-114

This project is a part of a larger bid project submitted by CCSE. The specific project premise was to deliver 289,430 kWh savings out of 2,946,061 kWh savings for the entire collections of projects. The evaluated savings were found to be 28,103 kWh for normal and 220,339 kWh for early. The M&V for this project has shown savings of 201,860 kWh and 33.2 kW in demand reduction. The reason for the drop between the estimated savings and the as found is the drop in operating hours from 8760 hrs/yr to 6,088 hrs/yr. The operating hours discovered during the M&V are even lower than the 7,147 hours pointed out by the evaluator. As a result from this project and other projects within this

ESB projects grouping, the claimed M&V savings dropped to 2,075,033 kWh from the project premise to deliver 2,946,061 kWh.

M49205 SPC 3548

This project is new installation of boiler and heat exchanger in a brand new cheese factory. The IOU did calculate the savings by obtaining future operating data from the facility staff. The heat recovery was based on the new boiler efficiency and on the process streams flows, temperatures and annual operating hours. The evaluator is talking about old plant, but there was not an old plant. The plant has not reached maximum production rates following the SPC project. The energy savings did assume that once the plant starts up they will operate at design level, it is assumed that either bottlenecks remained in the plant or the market was not there.

M49153 SPC 3357-07

The project is installation of one new 1280 ton chiller replacing 700 ton Trane chiller Model CVCA890. This is one of three centrifugal chillers at the facility. The new chiller is designated Chiller #3 by the site.
Please refer to section 6.1.Results Summary- Program claimed 740000 kWh and 171.6 kW. Please readjust 'Program kWh and "Program kW" in this section.
Please refer to section 3.1 Utility Algorithms- EQuest, not SPC program was used to calculate energy savings.
New chiller efficiency of .533 kW/ton was used in utility energy savings calculations, not .48 kW/ton as stated in 3.1.
For the EQuest model energy evaluation, the new chiller was compared to a 1998 Title 24 min (.676 kW/ton) chiller since chiller being replaced was overhauled in 1992, qualifying for early retirement.
Section 3.1 lists new chiller NPLV of .355 kW/ton. New chiller NPLV is .338 kW/ton- (per Trane equipment submittal).

Section 3.3 Evaluation Algorithms – Energy Savings
Correlation only yielded .42 R^2 value for outside air temp VS cooling load. Alternate method should be used to characterize cooling load.

Section 2.7 Site chillers – Customer has replaced 700 ton Trane Chiller (CVVA890, originally 1000 tons) w/1280 ton chiller (CVHF1280). This new chiller is designated chiller #3. This chiller will be staged with 2 existing 1000 ton electric chillers (Trane CVCA890). The new chiller is not staged with two 1300 ton chillers as currently stated in section 2.7
Section 3.1 states "the calculation does not take into account the lead/lag switching of the chillers that occurs every 7 days"
Customer signed post installation inspection confirmed "chiller 3 comes on first and. chiller 1 and chiller 2 come on after when load calls for it". Inspection report signed by site contact, (customer). Please Confirm with customer chiller 3 is lead chiller and operation does not alternate every 7 days

M01358 SPC 3034-06

The project is replacement of existing boiler with a more efficient model.
Section 6.2 Key Findings includes Table 5. In table 5 the headings - Baseline Input therms per day and Post Input therms per day should contain different values. Post input therms per day will be lower.
In Section 4.2 Data Collection Methods - Boiler sub meter gas usage was measured over a 16 day period September 28 through October 15. Sub metering should be carried out seasonally to capture variable seasonal loads.

M049261 SPC 3562

This project upgraded lighting and replaced HVAC programmable thermostats with EMS control.
During SPC review by utility, lighting hours that vendor proposed were similar to SPC characteristic schools lighting hours. Please furnish lighting logger information.

Section 3.3 Evaluation Algorithms - Evaluation Measure ID: M49261 HVAC thermostat upgrade –M&V has not been completed on this measure yet for the utility. Final utility approved savings for this measure has not yet been reported.
Section 3.3 Evaluation Algorithms states "Inputs to the baseline model were identical to those of the installed model with the exception of the HVAC operating schedules and heating/cooling set point schedules" How was the HVAC model compared to baseline utility data to insure that baseline T stat operation (with 7 day programmable t stats) was characterized accurately? Baseline equipment programmable thermostat set points, schedules and reference clock time can all be out of alignment with expected scheduling and set points. Billing analysis with corrections for lighting modifications would more accurately capture baseline HVAC operation.

On the EM&V reports reviewed many words and phrases are blanked-out and we have not been able to identify the SPC or BID project name or number. Additional assistance in identifying these projects by SPC or BID name and project number is requested. Following are a list of these report Evaluation ID Numbers.

720
1313
665
1357
1195
42147
42418
42498
47046
48395
965
965
1182
1183

1185
1186
1191
1193
48734

SoCalGas Program Specific Concerns

One overriding question for the auditors is that it appears that the auditors disregarded the calculation methodology used by the utility and chose to go with their own. This leads to questions of their methodology, who reviewed their assumptions, and why were the utility calculations so easily disregarded? In addition, for some of the projects, auditors chose to go with a simple bill regression analysis versus using actual data from metered equipment. It appears that the results were not validated against each other.

- M49914 & M49388 - 2 Furnace projects at same location for different production lines. Evaluator continually indicates production lower than "anticipated" as if production did not meet capacity. The evaluators own report indicates "due to a decreased product demand" (M49388). Original calculations were in accordance with production levels before decrease and downturn in economy. There is no reference to indicate previous production levels before decrease. Assumption that economic downturn leading to decrease product demand will remain stagnant for the life of the units is misleading and incorrect.

- M00919 - The evaluator indicates a realization rate of 79% due to the spillover effect yet in the evaluation the auditor themselves indicate "the primary site specific data collected for this evaluation was insufficient to capture the measure savings because of the myriad factors affecting production efficiency". The auditor agrees with rationale but reduces savings due to reduction of 5 - ?? unsure - in addition to 4 -?? - unsure items. Difficult to comment when the very reason for the reduction is not revealed. Auditor has difficulty measuring savings but can readily measure spillover affect.

- M00726 - Auditor indicates the burner that was replaced must meet SCAQMD standards and the existing burner was out of compliance with those requirements. As such, the burner is normal replacement. All based on questions posed to a manufacturer of burners. Questions arise. What questions were asked and were the requirements mentioned required at the time of replacement or implemented prior to date of requirements - if requirements are currently in effect or will be in the future. Is it the auditors assumption that all burners are equal with no efficiency differences? Also, does a customer have the option to not replace? This last question is not discussed in any of the evaluations.

- SoCalGas is requesting the EM&V report files for M49994 & M49998.

SDG&E Comments on 2006-08 Retro-Commissioning Impact Evaluation

SAN DIEGO GAS & ELECTRIC COMPANY

Comments on 2006–08 Retro-Commissioning Impact Evaluation

UC/CSU Project Specific Comments

1) SBW ID: P00317, IOU ID: 2

The specific MBCx project evaluated is EEM # 1, Re-Commission Laboratory and Fume Hood Controls

a) Issue: The original savings calculations for the project utilized a discharge air temperature of 55F for the air handlers while the reviewer's calculations used a discharge air temperature of 64.3F, resulting in a reduction in gas savings associated with reheat. Page 24 of the site report for SBW ID P00317 states that the average discharge air temperature of 63.4F was used to correct the baseline model's assumption where it had been set to 55, however a summary table of discharge air temperature supplied from the BMS trended from May 6, 2008 to September 17 2009 states an average discharge air temperature of 59F. It is recommended that the reviewer revise the calculations using the correct trended average discharge air temperature which appears to be 59F, not 63.4F.

b) Issue: Page 22 of the site report for SBW ID P00317 states that, "Slight corrections were also made to the model to correct assumptions regarding the central plant kW per ton rating used", but no initial values or altered values are provided. The report also lacks any specific justification for the revision of the central plant efficiency.

c) Issue: The energy savings estimation method used assumes a fixed 5,135 CFM air flow reduction, as provided by Phoenix Controls. The CFM reduction will vary depending upon sash position, and affect energy savings. A fixed discharge air temperature (DAT) was used for the energy savings calculation, based on short term monitoring of several air handlers. All 8 air handlers DAT are automatically reset and vary throughout the day. To more completely examine the heating, cooling, and fan power energy changes provided by this measure building power, chilled water, and hot water consumption pre and post should be compared. UCSD records this information monthly.

2) SBW ID: P00319, IOU ID: 4

The specific MBCx project evaluated at is EEM #1, VSDs on CT fans, Optimize Chiller/Boiler Sequencing.

a) Issue: It appears as though the reviewer did not account for new boiler controls implemented as part of the overall MBCx RCx project. The HW system measures are mentioned on page 18 of the site report for SBW ID P00319 under section 3.3 “Evaluation Algorithms- Energy Savings”, however section 7.2. “Key Findings” of the report does not include any information about the HW system measure. The associated gas savings for the MBCx project were reduced to zero therms yet there is no post implementation trending data, calculations, or supporting justification for reducing the gas savings. It is recommended that the reviewer verify the original gas savings calculations or eliminate the ex ante gas savings associated with the boiler controls as the resulting therm realization rate of zero is inaccurate.

SDG&E SCG Comments on the Small Commercial Impact Evaluation

SAN DIEGO GAS AND ELECTRIC COMPANY
AND
SOUTHERN CALIFORNIA GAS COMPANY

COMMENTS ON THE SMALL COMMERCIAL IMPACT EVALUATION

The Joint Utilities have several concerns with the Small Commercial Program Impact Evaluation. The most important of these are the following:

1) The relative precision of wattage, operating hours, and unit energy savings (UES) for high bay lighting are very poor, ranging from 14% to 45%. The reason this occurs is a repeated problem found throughout almost all of the Impact Evaluations being reviewed, the forcing of complex results from extremely small samples sizes. Table 3-28 illustrates this problem. For High Bay Lightening, the 90% confidence interval unit energy savings (kWh) is estimated at 45%! The question here is "How can any results be accepted when the confidence intervals are so extremely large and even, in many cases, include the ex-Ante estimates"?

2) Operating hour estimates are based on results from loggers and it appears that a large fraction of the variation in relative precision and realization rates stems from issues related to the logger analysis. Specifically, the logger analysis suffers from the following problems:
    a. Operating hours and UES estimates are based on only one to two months of logger data and *NO* adjustments are made for changes in lighting hours over the year. For example, the analysis does not take into account lower daylight hours during winter months or changes in occupancy rates over the year for hotels, etc. This information is critical because the loggers were installed over the period between September 2008 through October 2009 with the majority of loggers installed in the second and third quarters of 2009. Thus, operating hours were measured prior to winter months when the days are shorter and during a time period outside the peak summer travel months when hotels and restaurants tend to see the most activity. This suggests that the logger results could dramatically *understate* actual usage, particularly for some market segments.
    b. Second, this evaluation was supposed to cover the period 2006-2008 period. However, the loggers were installed during the second and third quarters of 2009, far outside the program timeframe. This could have a substantial impact on the results since the loggers were installed during the height of the economic downturn. Thus, the logger estimates may again substantially underestimate usage for a number and possibly all of market segments.
    c. Third, and as stated earlier, disaggregated estimates of hours of use by IOU and by market segment/activity are based on very small samples, particularly for the number of *sites* sampled. This undermines the validity of the logger results and adds substantial imprecision to the estimates. For

example, for downstream CFLs in the lodging market sector, a total of only 3 sites were logged for SDG&E.

3) No billing analysis was conducted for SDG&E to verify logger estimates for realization rates and UES. The reason given for this is (again) because of the small sample sizes. The billing analysis produced much more favorable results for PG&E and SCE but SDG&E was not allowed to benefit from this analysis because no results were produced.

4) One area of improvement in this report over the other Impact evaluations was in the area of Net to Gross Ratios (NTGR). The estimated NTGR values for SDG&E are uniformly high for the HIMs. Specifically, the NTGRs for interior screw lighting, high bay fluorescent lights and linear fluorescents are 0.87, 0.90, and 0.87, respectively. This finding is in marked contrast to the findings in other evaluation studies (e.g., major commercial, upstream lighting, and residential retrofit) and is likely the result of better implementation of the SRA. These other studies were beset with numerous problems (e.g., small sample sizes, inability to identify the appropriate decision-maker, lack of internal controls, lack of secondary sources of information, etc.). These issues seem to be largely overcome in this study, especially as they pertain to SDG&E.

5) Billing Model Concerns: While a billing analysis was used to estimate realization rates in the evaluation the results were not used. The modeling itself illustrates many of the problems inherent in most, if not all, of the evaluations currently being reviewed. The joint utilities point out these problems here to illustrate the overall problems in the many of the evaluations.

    - The results for the direct install programs are likely highly unreliable. As the evaluators note, the site billing data aggregation process may have been incomplete for direct install sites adding substantial measurement error to the analysis and substantially reducing the reliability of the results.
    - A large number of observations for sites with savings in excess of 40% of lagged usage were deleted from the analysis for various reasons. Specifically, the evaluators deleted observations with "excessive savings" but did nothing to observations with very low savings. Thus, the evaluators in essence preformed one-sided trimming. This could substantially bias the estimated savings downwards. Sensitivity analysis to this trimming needs to be conducted and reported.
    - The model specifications do not appear to control for month or year fixed effects. Given the timing of the analysis, overall yearly trends in energy usage could be important. Specifically, the billing data used in the analysis covers the time period between early to mid-2004 to June of 2009. Much of the post installation period is likely to be during the later part of that time frame which includes the economic downturn. While the evaluators attempt to control for "macro" events that might influence

usage by including a variable measuring changes in the county unemployment rate, such a variable is unlikely to fully control for time specific changes in usage. This is likely to bias the estimated savings downwards.

- The evaluators use engineering estimates of savings in their billing analysis. It appears that these engineering estimates were obtained from the logger analysis. Specifically, the evaluators note that, "the engineering ex post energy savings for the linear lighting measures and CFLs, adjusted for the on-site verification rates, are the savings impacts used in the billing analysis." As noted above, there are a number of substantial problems with the logger analysis including very small samples for some market segments. As a result, the engineering estimates most likely suffer from substantial measurement error. This in turn will tend to attenuate the estimated savings toward zero (i.e. produce lower savings estimates that actually occur).

Conclusion: The concerns and questions raised above put considerable doubt on the accuracy and reliability of the results estimated in the draft report. If these concerns cannot be addressed by the final report, in particular the small sample sizes used to estimate the impacts to SDG&E and biased results of the lighting logger study, then the joint utilities recommend that this study *not* be accepted as reliable or used for updating DEER or used to measure utility performance in the ERT and VRT process.

SDG&E Comments on the HVAC Impact Measures and Specialized Commercial Programs Impact Evaluation

SAN DIEGO GAS AND ELECTRIC COMPANY

COMMENTS ON THE HVAC HIGH IMPACT MEASURES AND SPECIALIZED COMMERCIAL PROGRAMS IMPACT EVALUATION

SDG&E has several concerns with HVAC HIM and Specialized Commercial Program Impact Evaluations. The most important of these are the following:

1. The evaluators note that the goal of the "overall verification and net savings sampling strategy was to achieve 10% precision at the 90% confidence level for each measure by utility." Unfortunately the evaluators did not come close to meeting that goal. The point is illustrated in the tables below, which show the achieved level of precision for the RCA and Air Conditioner Replacement evaluations. In Table 5-13 (RCA measures), achieved levels of precision vary between 13% and 34% with one program having no level of precision because the sample size was zero (PG&E2080). Of the nine measures, six had precision levels of 26% or higher (with one of those having no estimate due to a sample size of zero). Table 6-34 and 6-35 display a similar lack of precision for Air Conditioner Replacement measures. In the residential sector, the confidence bands around the estimated savings range from +/-131% to +/-21% with the majority being larger than +/-30%. As illustrated in Table 6-35 the lack of precision was even worse for commercial measures. The precision levels achieved in this study are so far from the original protocols, that they essentially make the savings estimates reported in the evaluation useless.

Table 5-13: Estimates of Achieved Precision for RCA measures

HIM	Program ID	Pre-Post Units	Error Ratio (er)	Sample RP at 90%CI	2006-08 kWh Savings	kWh Error Bound
Res RCA	PG&E2000R	7	0.5	31%	28,966,327	9,004,433
Res RCA	SCE2507	37	0.5	14%	56,440,821	7,624,172
Res RCA	SCE2502	6	0.5	34%	3,760,920	1,262,518
Res RCA	SDG&E3035	6	0.5	34%	1,373,476	460,919
Res RCA	PG&E2078	6	0.5	34%	1,191,979	400,011

HIM	Program ID	Pre-Post Units	Error Ratio (er)	Sample RP at 90%CI	2006-08 kWh Savings	kWh Error Bound
C&I RCA	PG&E2080	-	0.5	-	9,161,619	-
C&I RCA	SCE2507	30	0.5	15%	9,758,899	1,460,650
C&I RCA	SDG&E3043	10	0.5	26%	2,944,930	764,051
C&I RCA	PG&E2068	37	0.5	13%	4,818,552	648,535

Table 6-34: Residential Confidence and Precision of Savings Estimates for ACR Measures

	Climate Zone 7		Climate Zone 8		Climate Zone 9		Climate Zone 10		Climate Zone 15	
	Early	Burnout	Early	Burnout	Early	Burnout	Early	Burnout	Early	Burnout
Number of sites	28	28	16	16	8	8	71	56	28	28.00
Savings/ton (kWh)	23	8	78	42	90	52	101	61	310	125
Savings/ton +-	13	11	26	14	29	21	21	17	93	74
Savings/ton +- %	58%	131%	33%	33%	32%	40%	21%	28%	30%	59%
Grid Savings/ton (kW)	0.033	0.014	0.124	0.069	0.101	0.048	0.146	0.098	0.256	0.153
Grid Savings/ton (kW) +-	0.015	0.011	0.038	0.029	0.015	0.018	0.038	0.041	0.079	0.069
Grid Savings/ton (kW) +- %	45%	83%	31%	42%	15%	37%	26%	42%	31%	45%

Table 6-35: Commercial Confidence and Precision of Savings Estimates for ACR Measures

	Climate Zones 6 & 7		Climate Zones 8 &9		Climate Zone 10	
	Early	Burnout	Early	Burnout	Early	Burnout
Number of sites used	21	21	15	15	12	12
Savings/ton (kWh)	136	44	181	93	483	174
Savings/ton +-	58	34	93	62	179	102
Savings/ton +- %	43%	77%	51%	67%	37%	59%
Grid Savings/ton (kW)	0.13	0.05	0.18	0.11	0.30	0.12
Grid Savings/ton (kW) +-	0.039	0.026	0.065	0.055	0.145	0.118
Grid Savings/ton (kW) +- %	30%	53%	36%	50%	48%	98%

2) The extremely high error bounds associated with the RCA and Air Conditioner Replacement evaluations are directly due to sample size issues. In nearly all cases the evaluators were forced to work with samples that were significantly smaller than called for in the original protocols. The small sample sizes not only lead to extremely high error bounds for estimated savings, they also call into question the underlying validity of the results due to nonrandom sample selection. As an example, consider the Commercial RCA field findings. When discussing the sampling issues, the evaluators note that;

> "For the SCE 2507 Program, of the 31 contractors listed as performing commercial service, six were willing to participate in the metering effort. The cutoff date for installing metering equipment was set at September 28, 2009, to allow for adequate data collection before the cooling season ended. The contractors had a limited number of sites available, as they would not perform RCA testing during the hottest months of summer. The temperatures remained unseasonably hot until late in the season, preventing contractors from pursuing RCA work until near or past the metering cutoff date. The team attempted to install as many meters as possible, which resulted with 42 meter installations. Mechanical issues, such as compressor and metering equipment failures, further reduced the sample size to 36 units in SCE territory. The SDG&E 3043 Pre/Post Program used almost all of its rebate allocation within the first quarter of 2009,

> which resulted in sites being approved on a case-by-case basis only by the program implementer. Because few contractors participated, the sample size was smaller than expected with 16 meters installed. Due to unit mechanical failures and problems with metering equipment, the final sample size was reduced to 10 units. The PGE2068 Program had few contractors with the capacity and willingness to field M&V sites, resulting in the majority of units needing to be monitored by one contractor. The HVAC team completed 53 unit installations, including several multistage units. These units presented analytical challenges that prevented many from being included in the final evaluation analysis. No units were achieved for the PGE 2080 Program due to unsuccessful coordination attempts."

Thus, across all the IOUs the evaluators were forced to work with significantly smaller sample sizes than originally expected. Furthermore, the samples that were achieved most likely suffer from significant sample selection bias since they are based solely on those contractors that were willing to participate. The small sample size problem is exacerbated because the evaluators then needed to split the sample to estimate energy and demand savings for climate zones and building types.

3) The excessively small sample sizes also led the evaluators to deviate from the planned evaluation in a number of ways that compromise the findings of the evaluation. For example, for both RCA and Air Conditioner Replacement evaluations, the evaluators note the following:

> "The amount of charge correction for the sampled units could not be controlled to be mapped onto the program-level distributions. The result was that the small sample sizes for the charge-removed categories may not well-represent the very large program population of those units. The HVAC team decided to combine the results with all available data, namely the detailed data in similar format used to develop the DEER 2008 measure savings."

Thus, due to data issues and sample sizes the evaluators resorted to using data NOT collected for the program evaluation. How this impacted the findings is unclear but further suggests the results of the evaluation are unreliable and unverifiable.

4) Another example of the problems associated with the implementation of the evaluation can be seen in the Residential RCA Field Findings (Section 5.5.1). There, the evaluators note that:

> "In some cases, a majority of runtime data were collected at relatively similar temperature conditions in both pre- and post-maintenance cases. In other cases, the pre- and post conditions covered very different temperature or occupancy patterns. For the units where the performance covered similar pre and post conditions, the average condition was used, as presented in summary tables in Appendix E, to represent the capacity and unit efficiency. For units where the post-maintenance conditions were generally different, the conditions relative to

the standard curve were used to calculate a representative capacity and efficiency value for the temperature and humidity conditions seen before maintenance."

For a valid comparison of pre and post energy usage, the evaluators need to make certain that the only difference between the pre and post period is the *change in the RCA*. That is, they need to design as close as possible a controlled experiment so that other explanations for differences in energy utilization rates can be ruled out. As the paragraph above illustrates this was far from the case. In many cases the pre and post conditions varied significantly in terms of temperature and occupancy patterns. This significantly reduces the reliability of the findings. Furthermore, the solution to this issue appears *ad hoc* in nature. Why did the evaluators believe that "the conditions relative to the standard curve" would be representative of the capacity and efficiency value for the temperature and humidity conditions seen before maintenance?" At the very least, the evaluators need to present sensitivity analysis that provides some indication of the impact of this assumption on estimated savings.

5) The calculated NTGRs vary significantly across IOUs and across programs, even when the programs are providing similar services. For example, consider the Air Conditioner Replacement program NTGR results listed in the table below. Both SCE 2507 and SDG&E 3029 were programs that provided C&I AC Replacement. However, the estimated NTGR for the SCE 2507 program was 96% whereas the estimated NTGR for the SDG&E program was 3%. Thus, two programs offering the same service obtained polar opposite NTGR estimates. One has to wonder *how such radically different findings could arise?* While the designs of the programs differed in some ways, it seems extremely unlikely that two programs offering essentially the same service could have exactly opposite NTGR estimates. Results such as this call into question the validity of the NTGR evaluation procedure implemented by the evaluators. Similarly perplexing results for the NTGR ratios exist in the RCA program. For example, for residential RCA, the estimated NTGR for PG&E's 2000R program was 9% while the estimated NTGR for SCE's 2507 program was 77%. Both program provided similar services and yet ended up with radically different NTGR estimates.

Table 6-36: Free-ridership and NTGR Findings

Free-ridership Estimates and Net-to-Gross Ratio (NTGR) Findings kWh Weighted			
		% Free Riders (FR)	NTGR % (1-% FR)
PG&E 2080	ROB C/I Upstream A/C	6%	94%
SCE 2507	C/I AC Replacement	4%	96%
SCE 2507	Res AC Replacement	44%	56%
SDG&E 3029	C/I AC Replacement	97%	3%
SDG&E 3029	Res AC Replacement	47%	53%

Conclusion

It is clear from the above comments that the verification and gross impact analysis of the RCA and Air Conditioner Replacement HIMs suffer from significant problems related to sample size, sample attrition, nonrandom sample selection, and numerous other problems. While many of these issues appear to be beyond the control of the evaluators they nevertheless undermine the validity and applicability of the evaluations findings. In fact, the problems are so large, that it appears nearly all of the evaluation results lack the appropriate sample design and level of rigor necessary to yield results that are anywhere near to being usable for program evaluation.

As a result of the problems described above the Joint Utilities strongly recommend that this study *not* be accepted as reliable or used for updating DEER or used to measure utility performance in the ERT and VRT process.

SDG&E SCG Comments on the Draft Commercial Facilities Impact Evaluation: Door Gaskets and Refrigeration Strip Curtains

DIEGO GAS AND ELECTRIC COMPANY AND THE SOUTHERN CALIFORNIA GAS COMPANY

COMMENTS ON THE DRAFT COMMERCIAL FACILITIES IMPACT EVALUATION: DOOR GASKETS AND REFRIGERATION STRIP CURTAINS

The Joint Utilities have several concerns with Commercial Facilities impact Evaluation. The most important of these are the following:

1) The door gasket gross realization rate is very low for SDG&E. Some of the low realization rate is related to (1) the gas tracer finding that leakage through existing gaskets is much lower than anticipated (see Table 5-3) and (2) there was no significant difference in gasket leakage for participants (post-installation) and non-participants. Both of these factors increase the baseline from which savings are calculated. Unfortunately, the gas tracer and participant/non-participant data is not provided in the report. The Joint Utilities request that this data be provided.

2) The gross realization rate analysis for both door gaskets and strip curtains uses short-term (two weeks) monitoring to extrapolate to annual energy use. Extrapolation from short-term logger results is problematic because they are operational for insufficient periods of time. This creates two obvious problems. First, a short installation/operation period (e.g., two weeks) will be unlikely to capture the fluctuations in use patterns that occur over the year that correspond to variations in weather, daylight hours, macroeconomic conditions, etc. In essence, an analysis based on one to two weeks of logging data is exactly akin to basing the analysis on spot readings. Second, achieving a 10 percent precision with a 90 percent confidence level requires many more weeks of logger information. Thus, sampling design must consider both the number of sites and the time period over which the loggers are installed/operational. Without this the Evaluation results are unreliable and should be rejected.

3) The NTGR results for door gaskets and strip curtains are based on relatively small samples (26 sites and 81 sites, respectively). In addition, the SRA as used in this application suffers from the same set of problems identified in other comments (e.g., self-report bias, recall error, failure to survey the appropriate "decision-maker," inherent bias related to the "program influence" score in the non-residential survey, etc.). Again, these types of issues call into question the validity of the NTGR evaluation procedure implemented by the evaluators and the results of the evaluation.

Conclusion

As a result of the problems described above the Joint Utilities strongly recommend that unless these problems can be corrected, the study *not* be accepted as reliable or used for updating DEER or used to measure utility performance in the ERT and VRT process.

SDG&E SCG Comments on Southern California Industrial and Agricultural Contract Group: Pipe Insulation and Steam Traps

San Diego Gas & Electric Company
Southern California Gas Company
Comments on Southern California Industrial and Agricultural Contract Group:
Pipe Insulation and Steam Traps

The following are the comments of the Joint Utilities:

Overview of Program Evaluation

- Pipe insulation was evaluated using on-site engineering measurements and verification of therm impacts at 66 sites. All work was conducted in the SCG service territory. Estimates of the NTG ratios were based on the SRA method for non-residential programs and included data from all service territories. Parameters of interest included installation rate, ex post gross savings, UES, and NTGR.
- Commercial steam traps evaluated using SAE billing analysis. Estimates of the NTG ratios were based on the SRA method for non-residential programs.
- There is substantial uncertainty concerning the gross savings from industrial applications. Hence, the evaluation method relies on site-specific information to clarify usage parameters (hours of operation, pressure at boilers and steam traps, boiler efficiency, orifice size, number of traps, etc.), which are used in engineering algorithms to calculate ex post savings. Estimates of the NTG ratios were based on the SRA method for non-residential programs.

Overview of Program Results

There are two high impact measures (HIM) relevant to the Joint Utilities: pipe insulation and C&I steam traps. The evaluation results for these measures as they pertain to the Joint Utilities are as follows:

- The ex ante estimate of savings from pipe insulation were 16,400,122 therms and 126,630 therms for SCG and SDGE, respectively. The gross therm realization rate for pipe insulation is very low (0.079 for SCG). This rate is significantly lower that that for PG&E (0.35). The NTG ratio for pipe insulation for SCG is relatively high (0.72) whereas PG&E achieves a NTGR of 0.49. The low net realization rate for SCG is due to over-reliance on the dry cleaner segment (78% of sites), which have lower-than-assumed operating hours, higher-than-assumed ambient temperatures, and pre-existing pipe insulation.

- The ex ante estimate of savings from C&I steam traps were 15,252,403 therms and 537,187 therms for SCG and SDGE, respectively. The gross therm realization rate for commercial steam traps is very low (0.12 for SCG). This rate is significantly lower that that for PG&E (0.30) mostly because PG&E assumes a smaller ex ante therms/trap value (45.87 v. 139). The NTG ratio for commercial steam traps for SCG is 0.70 whereas SDG&E achieves a NTGR of 0.72 (PG&E = 0.62).

- The gross therm realization rate for industrial steam traps exceeds 2.0. Even with relatively low NTG ratios (0.52 for high pressure and 0.57 for low pressure) industrial steam traps have a net realization rate that exceeds 1.0.

Issues in Evaluation of Pipe Insulation

As specified above, the pipe insulation gross realization rates are very low for SCG. Some of the low realization rate is related to the large number of installations that fail to qualify for the program (new construction, partially new pipe, and pipes with pre-existing insulation). These sites receive either a zero realization rate or a rate that is heavily discounted. However, even if one ignores the non-qualifying sites, the gross realization rate is only 25% for *qualifying sites*. This very low gross realization rate is based on data collected in the field from the qualifying sites and consists of spot readings and data collected from HOBO U12-012 and U12-014 thermocouple loggers. The key drivers in reducing the realization rate below the ex ante values are operating hours (less than expected) and ambient temperatures (higher than expected). There are several potential problems with the spot and logger measurements. Two are listed below.

- Measurement error – this is especially important for spot reading since temperatures (internal and external) fluctuate widely over a day, week, month, etc. In addition, temperatures may fluctuate widely over the pipe length. Thus, spot readings cannot be used reliably to extrapolate to create an annual use pattern. In fact, the study comes to this exact conclusion in Appendix A-5: "Logged temperatures provide more reliable temperature estimates than spot readings and were used wherever possible."

- Inappropriate extrapolation from small data sets – according to Appendix A-5 the loggers were left in place for a "minimum of a week, and typically between one and two weeks. In a few cases, loggers were left for as long as eight weeks." There are two obvious problems associated with having loggers in place for only one to two weeks and then extrapolating to estimate annual energy use. First, the type of measurement error defined above will be present. Specifically, a two-week period will be unlikely to capture the fluctuations in use patterns that occur over the year that correspond to variations in weather, daylight hours, macroeconomic conditions, etc. In essence, an analysis based on one to two weeks of logging data is exactly akin to basing the analysis on spot readings, which were defined by the study authors to be woefully insufficient. Second, define the 52 weeks in a year as the relevant population and the logging period as the sample (i.e., results are going to be extrapolated to the year based on a sample). Then, in order to achieve a 10 percent precision with a 90 percent confidence level a sample in excess of 29 weeks would be required. Even if one, uses 365 days as the population the 90/10 sample would be greater than 57 days or more than 8 weeks. Thus, the sampling design allowed for the correct number of sites (66) but the logging sample design was insufficient to produce the 90 percent confidence bounds presented in Table 3-10. In fact, the errors are much greater than those presented.

The NTGR results for pipe insulation are based on a relatively large sample (although the research protocol goals were not achieved). However, the SRA as used in this application suffers from the same set of problems identified in other comments (e.g., self-report bias, recall error, failure to survey the appropriate "decision-maker," inherent bias related to the "program influence" score in the non-residential survey, etc.). Again, these problems suggest caution should be used when applying the estimated NTG ratios in policy situations.

Issues in Evaluation of Commercial Steam Traps

Ex post gross savings are estimated using commercial billing analysis from 497 dry cleaning establishments. The results, presented in Table 4-11 suggest that the gross realization rates for PG&E and the Joint Utilities are 30.4% and 11.76%, respectively. There are several obvious problems with the billing analysis as it is conducted and reported. With regard to reporting, there are many missing details such as the final count on the number of observations and the estimated coefficients (and corresponding significance information) for the month and year fixed effects. Without this information it is difficult to fully evaluate the estimated results.

With regard to the results presented in Table 4-11 the following methodological comments apply.

- First, the estimation approach does not conform to a classic "event study," which generally includes information pre and post of the event of interest (in this case the replacement of steam traps) and the occurrence date of the event is part of the *a priori* information set. The data set is constructed around this date (e.g., billing data twelve months before and twelve months after event) and difference-in-difference estimation methods are used. It seems from the description on pages 4-9 – 4-11 that the study authors do not have information on when exactly the steam trap change-out occurred and are using the billing data to implicitly identify this date. Thus, they use the engineering estimate of savings as an independent regressor and utilize a time series of data sufficiently long to include at least 12 months of post installation information. The success of this approach is dependent on the accuracy (i.e., lack of measurement error) of the engineering savings variable. That is, as measurement error increases in the engineering savings variable the estimated coefficient in the regression will attenuate toward zero. Engineering estimates are inherently imprecise because, for example, they are based on previous studies that utilize logger data for one week extrapolated to the year. Thus, the estimated coefficient, which is interpreted as the gross realization rate, is biased downward. The magnitude of this error is unknown.

- Second, the R-square values for the estimated equations are very low (0.02 for PG&E and 0.59 for the Joint Utilities). This suggests that the variation in the independent variables capture very little of the variation in the dependent variable (Δtherms) and is evidence that the model suffers from omitted variable bias. Two variables that would be potential candidates for inclusion in the model would be individual-specific fixed effects and terms that are the interaction of the individual-specific fixed effects and measures of weather. The absence of

individual-specific fixed effects means that the model fails to capture any observable or unobservable factors that influence energy consumption but does not vary over time (e.g. square feet of living space, number of occupants in facility, etc). Furthermore, the inclusion of the interaction terms adjusts for each facility's weather-sensitive energy usage (i.e. individual specific weather related energy consumption patterns). The relative importance of these variables and other possible variables should be explored.

- Third, this study provides only one regression based on the billing data. There are no robustness tests, no consideration of alternative functional forms or independent variable sets, and no evaluation of measurement error or alternative estimation techniques. This one regression then becomes the sole basis for the determining the gross savings realization rate. This seems premature without additional information about the sensitivity of the results to a series of robustness tests. The Joint Utilities should request the billing analysis data set for in order to perform additional testing.

In essence, the empirical work is not up to current scientific standards regarding the development, use, and reporting of estimated models. It is strongly recommend that these authors review the following two papers, which, by the way, are neither new nor novel:

- Leamer, Edward E., "Let's Take the Con Out of Econometrics," *American Economic Review*, 73, pp. 31-43, 1983.
- Leamer, Edward E., "Sensitivity Analysis Would Help," *American Economic Review*, 75, pp. 308-313, 1985.

The NTGR results for commercial steam traps are based on a relatively large sample (although the research protocol goals were not achieved). However, the SRA as used in this application suffers from the same set of problems identified in other comments (e.g., self-report bias, recall error, failure to survey the appropriate "decision-maker," inherent bias related to the "program influence" score in the non-residential survey, etc.). Again, these problems suggest caution should be used when applying the estimated NTG ratios in policy situations.

SDG&E SCG Comments of the Residential Retrofit Impact Evaluation Comments

SAN DIEGO GAS AND ELECTRIC COMPANY AND
THE SOUTHERN CALIFORNIA GAS COMPANY

COMMENTS OF THE RESIDENTIAL RETROFIT IMPACT EVALUATION
COMMENTS OF THE

The Joint Utilities have several concerns with the Residential Retrofit Program Impact Evaluation. The most important of these are the following:

1) The primary conclusions regarding almost every measure (furnaces, clothes washers, dishwashers, insulation, room air conditioners, and pumps and motors) are identical. Specifically, the installation/operation compliance rates are near 100% but the overall values are reduced dramatically by excessive estimates of free ridership. All free ridership estimates for these programs use the same protocol, the self report approach (SRA). In every situation, this approach generates excessive free ridership estimates, in spite of evidence to the contrary. For example, evidence from sales of ENERGY STAR retailers on clothes washers suggests a much lower free ridership rate. As stated in the report (page 48):

   "The self-report NTGR is also substantially higher than the market share data reported by the Department of Energy (DOE). The National ENERGY STAR Retailer Partners are required to annually provide sales data to the DOE for dishwashers, clothes washers, room air conditioners, and refrigerators. In 2006-2008 the National ENERGY STAR retailer partners reported the market share data for ENERGY STAR clothes washers (which is also inclusive of all the more efficient CEER tiers) was 38% in 2006, 42%, and 24%, respectively. Additionally, the 2007 Itron Market Share Report[38] found that 45% of California clothes washer sales were ENERGY STAR rated or higher. While this is not an estimate of free-ridership, it is an indication that sales of ENERGY STAR clothes washers were in the 24%-42% range throughout the U.S., substantially lower than the self-reported estimate of free-ridership in this study".

In addition, the SRA results are inconsistent with actual behavior of consumers, particularly in the case of insulation. That is, if individuals would have put in more insulation in the absence of the program then why hadn't they already done so? Many individuals in the program have owned their homes for an extended period of time and could have added insulation prior to the program being available. But they *did not do it* *until* the program was available. The SRA also does not allow for an examination for how stocking practices may determine choice – if stocking is changed in response to the program then when individuals are asked about their purchases they state that they would have purchased the same thing regardless of program – but if they have no choice due to stocking practices then the answer is meaningless. This contrary evidence suggests that the SRA is most likely biased in the direction of producing very low NTG rates. The SRA should be subject to robustness testing or sensitivity analysis.

2) The results for refrigerator recycling deviate from the norm described above. Specifically, the estimated net-to-gross values are equal to or greater than claimed values but the ex post gross savings are significantly less than ex ante gross savings. It should be noted that the NTGR estimates for recycled refrigerators are based on a *completely different protocol than the SRA* (non-participant behavior is evaluated and the disposal of the old refrigerators is considered). Again, this suggests that the NTG values for the other measures in all of the evaluations that use the SRA methodology may be significantly biased *downward*.

3) The ex post gross savings estimates for refrigerators was done differently than the other measures evaluated and are based on a regression of *in situ* modeling data rather than a regression of DOE modeling data (the protocol used in previous evaluations). The joint utilities have the following concerns with this type of evaluation.
    a. First, because the participants in the *in situ* study know they are being evaluated (metered), results of the *in situ* evaluations are subject to the Hawthorne effect. Specifically, the Hawthorne effect refers to the situation where members of the treatment group act differently because their participation in the study makes them feel special, regardless of the treatment. This will bias the results of the evaluation towards finding smaller energy savings than would otherwise be present.
    b. Second, there is potential for significant sample selection bias, in which households that agreed to participate in the *in situ* evaluation were not randomly assigned. Thus, the evaluation may suffer because the characteristics of participants in the study (both observed and unobserved) are not representative of the larger population. For example, those that agreed to participate may be "greener" than those that did not agree to participate. This would once again bias the results of the evaluation towards finding smaller energy savings than would otherwise be present. This problem may be particularly acute given that the author's note that: "However, due to difficulty recruiting sufficient participants for the metering study, the final metering sample fell short of its goal. In particular, the evaluation was unable to solicit as many primary appliances as targeted. This was due to participants' unwillingness to postpone delivery of their replacement appliances in order to allow sufficient time for metering. There were also issues related to the timeliness of participation data provided by the utilities or its implementers. In addition, a significant number of metering study participants canceled after initially agreeing to participate. As noted above, this problem was particularly prevalent for primary appliances."
    c. Third, none of the work is up to current scientific standards regarding the development and use of empirical models. Further, these standards are not of recent vintage. For example, see the following two papers published in the early to mid-1980s.:

i. Leamer, Edward E., "Let's Take the Con Out of Econometrics," *American Economic Review*, 73, pp. 31-43, 1983.
ii. Leamer, Edward E., "Sensitivity Analysis Would Help," *American Economic Review*, 75, pp. 308-313, 1985.

d. The refrigerator recycling evaluation provides only one regression based on the *in situ* modeling. Contrary to the standards for conducting econometric analysis, there are no robustness tests, no consideration of alternative functional forms or independent variable sets, and no evaluation of measurement error or alternative estimation techniques. And this one regression is the sole basis for the new evaluation protocol and the ultimate determination that the IOUs fail to achieve the savings goals. This seems premature without additional information about the sensitivity of the results to allowance for the Hawthorne effect, sample selection bias, and a series of robustness tests.

e. Fourth, the regression results produce inconsistent and counter-intuitive results. For example, the model used to forecast yearly refrigerator use from 10 – 14 days of *in situ* modeling concludes that "no statistically significant difference was found between appliances in conditioned v. unconditioned spaces" (see p. 138). Yet, outdoor temperature is significant (see p.139) and appliances in warmer climate zones use more energy than those in cooler climate zones (see p. 142).

For these reasons, the joint utilities recommend that the results of the impact evaluations for refrigerators using the in situ methodology be rejected as unreliable.

Conclusion

As a result of the problems described above particularly in the SRA for NTG and the *in situ* methodology used in the refrigerator impact evaluation, the Joint Utilities strongly recommend that this study *not* be accepted as reliable or used for updating DEER or used to measure utility performance in the ERT and VRT process.

SDG&E Comments on the Upstream Lighting Impact Evaluation

SAN DIEGO GAS AND ELECTRIC COMPANY

COMMENTS ON THE UPSTREAM LIGHTING IMPACT EVALUATION

The following are the major comments/concerns of SDG&E regarding the draft Upstream Lighting Impact Evaluation.

1) Analysis was focused primarily on screw-in CFLs with little or no independent evaluation of non screw-in CFLs. There is no comparative analysis to indicate to the reader that CFL results can be reliably extrapolated to other lighting products.
2) The results from the analysis used to estimate the number of CFLs "not sold" was used exclusively to estimate the "not sold" sales of the other lighting products as well. This clearly is not appropriate. No independent analysis was done on either Fixtures or LEDs which are completely different products. This is a major flaw in the overall report and must be corrected in the final report. Otherwise the SDG&E should receive full credit for its reported Fixtures and LEDs.
3) The "leakage" estimates rely on a vulnerability index defined by stores that stock the product and are within ten miles of non-IOU areas. SDG&E receives a large leakage penalty due to the proximity of Mexico to the SDG&E service territory. Again, this analysis is limited to screw-in CFLs with no information offered about other lighting products. In addition, the vulnerability measure is completely *ad hoc*, there is absolutely no discussion of how results are altered if another distance is used (i.e., robustness), and it is never tested as to whether or not vulnerability is transferable across IOU areas again casting doubt on the results of the analysis. A more appropriate model would be a spatial model of the diffusion/leakage process. Finally, the report ignores the possibility of "reverse leakage" in which products from non-IOU programs leak into IOU areas.
4) The adjustment for the use of a product in non-residential applications is limited to screw-in CFLs, with no information offered about other lighting products. The analysis is based on the CFL user survey, in-store consumer intercept surveys, on-site data, and extrapolation techniques. The manner in which the survey results are used in conjunction with the extrapolation techniques is never defined. Products other than CFL are never evaluated and therefore the CFL results should not be applied to non-screw-in CFL products and therefore no adjustment to reported non-screw-in CFLs should be applied.
5) The installation rate analysis was supposed to be based on "three inter-related models" (diffusion model, purchase model, and installation model). However, these models were deemed as not useful because of poor data quality (see page 16). Instead what was used was a much simpler and less reliable "trajectory" analysis. This approach is based on simple accounting and does <u>*not allow*</u> for any changes in behavior or estimated parameters over time. This is a serious problem when attempting to transform the market and casts major doubt on the results of the evaluation. In addition, this section of the report is just one of many examples where the report promised to conduct a specific investigation but instead did something completely different using readily available but not appropriate data and modeling methods. This is unacceptable and the modeling should be rejected.

6) The consumer self-report study asks consumers only one hypothetical question – would you select CFLs if they cost twice as much. Answers to this question are beset with potential bias (the authors agree on page 27 "hypothetical, out-of-context purchase decisions are not reliable predictors of actual behavior"). As such, self-reported survey responses need to be calibrated with other information or the analysis should be deemed as unreliable and rejected.
7) The econometric analysis is based on four statistical models (hedonic price model, conjoint analysis, revealed preference models, and the total sales approach). There are significant issues with each of these models.

    a. First, a hedonic price model is designed to determine the hedonic or implicit price of the characteristics that differentiate closely related products in a product class. Hedonic price models are used in situations in which the value of a specific characteristic is unknown (e.g., the value of school quality that is capitalized into the price of nearby homes). Application of the hedonic price method requires a large data base of the price of closely related products and their corresponding characteristics. For example, it is not uncommon to have several hundred thousand observations of home prices and characteristics in order to have the confidence to place an economic value of a specific characteristic. The number of observations used in this report was never specified. In addition, the estimation results are not provided. The only suggestion is that the hedonic price on the variable "IOU Discount" is not different from the evidence from the summary statistics. This suggests a poorly structured model that is inconsistent with a very large literature and that does not provide meaningful results. As such this study should not contribute to the preponderance of evidence and should be eliminated from the report.

    b. Second, the conjoint analysis is supposed to provide stated-preference information about the purchase of lighting products. However, the process as described is not consistent with a large and growing literature (much of it used to value non-market goods) which uses multiple questions to triangulate answers and ensure statistical reliability. In addition, the actual survey and choice set design are not provided for review. It seems that the work is based on a software product that the researchers are unfamiliar with. The important choice question, which follows an extensive education process, obviously leads to biased responses that are meaningless. This study should not contribute to the preponderance of evidence and should be eliminated from the report.

    c. Third, two revealed preference models were utilized to calculate the NTGR: Revealed Preference Purchase Models and a Revealed Preference Elasticity Model. The methods relied on revealed preference survey data collected through in-store consumer intercepts. Final NTGR estimates

used by the authors are based on these two revealed preference approach models. Table 24 presents final recommendations. SDG&E has the lowest NTG of 48%. There are many concerns with these NTGR estimates. First, it should be noted that the authors directly acknowledge the limitations of these estimates. They state on page 75:

"Finally, given the timing of this evaluation (and the broad market changes occurring toward the end of 2007/early 2008, as discussed above), we are concerned that none of the NTGR results derived from the various methods can be considered representative of the 2006-2008 program. Most of the data collection that supported the various NTGR analyses was implemented between mid-2008 and mid-2009. The only NTGR estimate that was defined as representative of the full 2006-2008 program effect was based on the supplier self-report approach. However, we do not believe that these estimates, which tend to be the highest of all of the estimates, are accurately capturing the effect of this difference in timing – rather, it is likely that the supplier self-report estimates are higher than other estimates as a result of the respondent biases discussed in this report."

Thus, the author's note that the only NTGR estimate that was defined as representative of the full 2006-2008 program effect was based on the supplier self-reported approach, yet in the end, the author's *disregard* those estimates and never use them to calculate their recommended NTGR estimates.

The models the authors actually rely on also have numerous deficiencies. For example, the estimates obtained from the Revealed Preference Purchase Models are based on logistic estimates. Numerous questions and concerns arise from these estimates including: (1) What was the R-square values of the various regressions and therefore how reliable are these estimates? (2) Regressions seem very *ad hoc* – based on stepwise regression techniques and *ad-hoc* assumptions about having "well determined coefficients." (3) The authors construct estimates of the NTGR based on coefficients from the logistic regression and average values of independent variables with and without the program. However, the average values of the independent variables with and without the program are nearly identical (see Table 90) and thus very little is learned from the exercise or the calculation of the NTGR. (4) The estimated coefficient on the average price of a CFL is positive, implying higher average price increases probability of purchase in hardware home improvement category. This seems very counterintuitive and suggests substantial model misspecification. (5) Many of the estimated coefficients make no sense and furthermore, specifications cannot be compared since they include different variables.

In addition to the logistic regression approach to calculating NTGR based on the Revealed Preference Purchase Models, the authors also use a simple contrast method based on the sales data themselves. In their final calculation of the NTGR based on this method, the authors make the unrealistic assumption that total CFL sales are not substantially changed by the program. Essentially, the authors are assuming that lowering the price of CFLs with the discount does not affect the total number of CFLs sold, implying that customers are unresponsive to price. This seems particularly unrealistic given that in their revealed preference elasticity model results they find that 61% of customers surveyed stated that they would have purchased fewer bulbs if the price of bulbs were doubled. Because of the author's assumption that total CFL sales are not changed by the program, the author's calculation of the NTGR represents a lower-bound of the actual NTGR. That is, this assumption leads to a conservative estimate of the NTGR.

The second revealed preference method used by the author's was a Revealed Preference Elasticity Model. Essentially, the authors asked the revealed preference survey respondents to indicate how many CFLs they would have purchased compared to their actual purchase at double the price. Based on those answers, the authors calculate an "elasticity," which is then used to construct an estimate of the NTGR. The first big concern here is that this really is NOT a revealed preference approach. Rather it is more like a *stated* preference approach, and therefore suffers from all the problems associated with stated preference analysis (problems the authors use to discount the results of stated preference estimates of the NTGR). Specifically, the authors did not set up an experimental design whereby consumers were offered CFLs at one price and then another price that was double the original and then observed the purchasing behavior of those consumers. Such a design would be a revealed preference approach. Rather, the authors simply asked consumers what they WOULD purchase IF the price were double – this is fundamentally a stated preference approach to estimating willingness to pay.

A second concern with the Revealed Preference Elasticity Model approach is that the authors make the *ad hoc* assumption that if a respondent reported they would purchase fewer CFLs if the price doubled the authors coded those individuals as having been willing to purchase 80% of the number of CFLs they had originally. This assumption appears completely *ad hoc* in nature and serves to drive down the final estimates of the NTGR. For example, for SDG&E, assuming individuals would have purchased 80% of the CFLs they had originally leads to an estimated NTGR of 34%. In contrast, assuming individuals would only have purchased 20% as many CFLs as they did originally would lead to an estimated NTGR of 62%, or nearly double the alternate estimate.

In conclusion, the revealed preference modeling does not provide reliable estimates of net-to-gross and does not contribute to the preponderance of evidence. This section should be eliminated from the report.

d. The final method used by the authors to calculate the NTGR is the Total Sales (Market-Based) Approach. As the author's note at page 185, "The primary methodology for the total sales (market-based) approach was a regression model to predict CFL sales as a function of program activity, while controlling for demographic, household, and economic factors that can also influence sales. The analysis presented was based on data from 1,046 onsite lighting inventories conducted in 11 areas in the U.S. Some of these areas have no CFL programs, some have modest or newer CFL programs, and some have longstanding aggressive CFL programs." One major concern with this approach (and there are many) is that California is NOT included in the regression sample. Thus, the estimates for California obtained from this method are all out-of-sample estimates. Such estimates can be very unreliable, particularly if the underlying regression model is incorrectly specified. Another concern is that no details are provided concerning the econometric model used to obtain the estimates, nor the actual estimated coefficients or their standard errors. Thus, no assessment can be made about the reliability of the regression model used in the study. A third concern is that the authors make use of data from one year, namely 2008. Thus, the model can not estimate the program impacts over the relevant time frame, namely 2006-2008. A fourth concern is that, given that that authors utilize data from only one year, the regression model is simply based on cross sectional data (i.e., data for one year from and different states). This calls into question the internal validity (reliability) of the estimated parameters since the authors are unable to control for state-specific time-invariant unobservable variables that are likely correlated with the explanatory variables of interest. In short, this method does not provide reliable estimate of the NTGR and does not contribute to the preponderance of evidence.

8) The conclusion regarding channel shift is unreliable (see p.181): "While we cannot say with certainty whether channel shift happened between stores, we can at least rule out stores where it is unlikely to happen …." Unfortunately, the authors cannot determine channel shift or the lack thereof because they do not have any real data pertinent to the question. Rather, the channel shift analysis is based entirely on distance between stores. This is a completely *ad hoc* exercise that does not, in any manner, add to the discussion regarding substitution between stores. This section of the report should be eliminated and the utilities should suffer no reduction in claimed sales/savings due to this hypothetical effect.

Conclusion

The list of issues provided above, although not exhaustive, illustrates that this report is certainly a very poor example of an evaluation of an on-going energy efficiency program.

The report is filled with promises of what we are/were going to do but never actually did. Moreover, the work actually performed is completely *ad hoc*, devoid of any theoretical or empirical foundation. Very little, if any, of the work corresponds to the current literature or protocols established for completing these types of evaluations. The study should be completely re-worked, eliminating all the analysis incorrectly done or not done at all. If this is not done, the Joint Utilities recommend that this study *not* be accepted as reliable or used for updating DEER or used to measure utility performance in the ERT and VRT process.

CERTIFICATE OF SERVICE

I hereby certify that I have this day served a copy of the foregoing **COMMENTS OF SAN DIEGO GAS & ELECTRIC COMPANY (U 902 M) AND SOUTHERN CALIFORNIA GAS COMPANY (U 904 G) ON THE ASSIGNED COMMISSIONER'S RULING PROVIDING ENERGY DIVISION REPORT AND SOLICITING COMMENTS ON SCENARIO RUNS *(RE-FILED AND RE-SERVED TO REFLECT CORRECT PROCEEDING NUMBER IN THE CERTIFICATE OF SERVICE ONLY)*** on all known interested parties of record in **R.09-01-019** via email to those whose email address is listed in the official service list and via first-class mail to those whose email address is not available.

Copies were also sent via Federal Express to Commissioner John Bohn and Administrative Law Judge Thomas Pulsifer.

Dated at Los Angeles, California, this 26th day of May 2010.

/s/ Marivel Munoz
Marivel Munoz

CALIFORNIA PUBLIC UTILITIES COMMISSION
Service Lists: R.09-01-019 - Last changed: May 21, 2010

donaldgilligan@comcast.net; SDPatrick@SempraUtilities.com; larry.cope@sce.com; dil@cpuc.ca.gov; marcel@turn.org; bfinkelstein@turn.org; lhj2@pge.com; M1ke@pge.com; wbooth@booth-law.com; wem@igc.org; ABesa@SempraUtilities.com; PVillegas@SempraUtilities.com; mrw@mrwassoc.com; tam.hunt@gmail.com; gandhi.nikhil@verizon.net; jerickson@summitblue.com; fstern@summitblue.com; Scott.Dimetrosky@cadmusgroup.com; ckmitchell1@sbcglobal.net; david@nemtzow.com; darren.hanway@sce.com; don.arambula@sce.com; kathleen.a.qumbleton@sce.com; tory.weber@sce.com; case.admin@sce.com; monica.ghattas@sce.com; jennifer.shigekawa@sce.com; liddell@energyattorney.com; CentralFiles@SempraUtilities.com; sephra.ninow@energycenter.org; JYamagata@SempraUtilities.com; bob.ramirez@itron.com; dmano@enalasys.com; Jeff.Hirsch@DOE2.com; ddavis@cecmail.org; hprince@rsgrp.com; john.stoops@rlw.com; jeanne.sole@sfgov.org; FSmith@sfwater.org; mramirez@sfwater.org; tburke@sfwater.org; jchou@nrdc.org; lettenson@nrdc.org; nlong@nrdc.org; pmiller@nrdc.org; cjn3@pge.com; efm2@pge.com; yxg4@pge.com; filings@a-klaw.com; ldri@pge.com; nes@a-klaw.com; cpuccases@pge.com; sls@a-klaw.com; SRRd@pge.com; SRH1@pge.com; bdille@jmpsecurities.com; cassandra.sweet@dowjones.com; sdhilton@stoel.com; cem@newsdata.com; RegRelCPUCCases@pge.com; slda@pge.com; rsridge@comcast.net; ghamilton@gepllc.com; jak@gepllc.com; cadickerson@cadconsulting.biz; Michael.Rufo@itron.com; rmurray@us.kema.com; achang@efficiencycouncil.org; sschiller@efficiencycouncil.org; jskromer@qmail.com; mmyers@vandelaw.com; dwang@nrdc.org; sberlin@mccarthylaw.com; brbarkovich@earthlink.net; bill@jbsenergy.com; erik@erikpage.com; mjaske@energy.state.ca.us; rliebert@cfbf.com; grover@portland.econw.com; Allen.Lee@cadmusgroup.com; ppl@cpuc.ca.gov; aeo@cpuc.ca.gov; cbe@cpuc.ca.gov; cf1@cpuc.ca.gov; cxc@cpuc.ca.gov; css@cpuc.ca.gov; jl2@cpuc.ca.gov; cln@cpuc.ca.gov; jst@cpuc.ca.gov; jnc@cpuc.ca.gov; kwz@cpuc.ca.gov; keh@cpuc.ca.gov; kmb@cpuc.ca.gov; lp1@cpuc.ca.gov; mwt@cpuc.ca.gov; mmw@cpuc.ca.gov; mkh@cpuc.ca.gov; pw1@cpuc.ca.gov; pcf@cpuc.ca.gov; rhh@cpuc.ca.gov; srm@cpuc.ca.gov; tcx@cpuc.ca.gov; trp@cpuc.ca.gov; tcr@cpuc.ca.gov; zap@cpuc.ca.gov; ztc@cpuc.ca.gov; awp@cpuc.ca.gov; mulloa@semprautilities.com; luluw@newsdata.com; mmunoz@semprautilities.com

Mike Yim
2920 Camino Diablo, Suite 210
Walnut Creek, CA 94597