LBL-36917
UC-1600

ERNEST ORLANDO LAWRENCE
BERKELEY NATIONAL LABORATORY

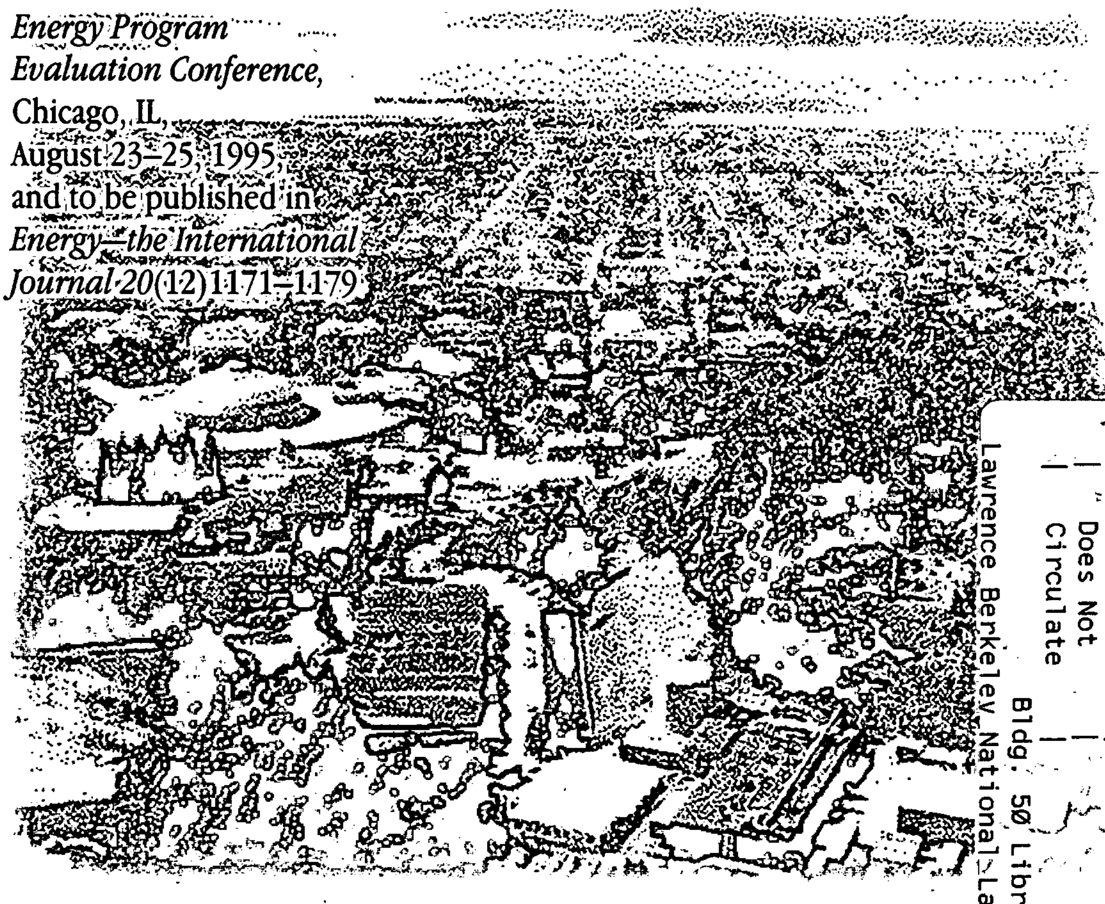
The Reliability of DSM Impact Estimates

Edward L. Vine and Martin G. Kushler

Environmental Energy Technologies Division

May 1995

Presented at the
1995 International Energy Program Evaluation Conference,
Chicago, IL,
August 23–25, 1995,
and to be published in
Energy—the International Journal 20(12) 1171–1179



LBL-36917
UC-1600

The Reliability of DSM Impact Estimates

Edward L. Vine
Martin G. Kushler

Energy Analysis Program, Energy and Environment Division
Lawrence Berkeley Laboratory. Building 90-4000
University of California, Berkeley, CA 94720

May 1995

Presented at the 1995 National Energy Program Evaluation Conference
Chicago, IL, August 1995

This work was supported by the Assistance Secretary for Energy Efficiency and Renewable Energy, Office of Utility Technologies, of the U.S. Department of Energy under Contract No. DE-AC03-76SF00098.

THE RELIABILITY OF DSM IMPACT ESTIMATES

Edward L. Vine, Lawrence Berkeley Laboratory, Berkeley, CA

Martin G. Kushler, Michigan Public Service Commission, Lansing, MI

Abstract

Demand-side management (DSM) critics continue to question the reliability of DSM program savings and, therefore, the need for funding such programs. In this paper, we examine the issues underlying the discussion of reliability of DSM program savings (e.g., bias and precision) and compare the levels of precision of DSM impact estimates for three utilities. Overall, the precision results from all three companies appear quite similar and, for the most part, demonstrate reasonably good precision levels around DSM savings estimates. We conclude by recommending activities for program managers and evaluators for increasing our understanding of the factors leading to DSM uncertainty and for reducing the level of DSM uncertainty.

Introduction

Supporters of energy efficiency assert that demand-side management (DSM) is a reliable resource that should be included in a utility's resource planning activities. However, DSM skeptics continue to question the reliability of DSM program savings and, therefore, the need for funding such programs. DSM planners and evaluators will need to address this issue in the coming years if DSM is to continue. Thus, the purpose of this paper is to examine the issues underlying the discussion of reliability (e.g., examining other areas of uncertainty in demand-side and supply-side planning, and the potential negative side effects of focusing on levels of precision). In addition, we compare the levels of precision of DSM impact estimates for three utilities.

As described below, a few state and federal regulators have addressed the reliability of DSM programs and the precision of energy-savings estimates (e.g., California, Massachusetts, New Jersey, Pennsylvania, and the U.S. Environmental Protection Agency). As DSM comes under even greater scrutiny in a more competitive environment, it is expected that additional state and federal regulators will have to respond to these issues (e.g., as part of the development of statewide or national measurement and evaluation protocols). In the last few years, a number of practitioners in the field of DSM program evaluation have discussed the issue of reliability and uncertainty in the literature, in regulatory hearings, and in program evaluations (e.g., Buller and Miller 1992; Hanser and Violette 1992; Horowitz 1992; Messenger et al. 1994; NEES 1994; Quantum Consulting 1994; Raab and Violette 1994; Schlegel et al. 1991; Sedmak et al. 1994; Sonnenblick and Eto 1995; Violette 1991; and Xenergy 1993 and 1994). After reviewing this material, we recommend activities for improving our understanding and, possibly, reducing the uncertainty of DSM as a resource.

Putting DSM Uncertainty in Perspective

There is always uncertainty about the savings associated with DSM measures. This level of uncertainty permeates all DSM program evaluation activities: e.g., designing samples, collecting and analyzing data, and interpreting and reporting the results of evaluations. Given this level of uncertainty, one must understand the limits of evaluation - what can and cannot be accomplished by evaluation.

Many factors influence the value of a resource to a utility, and uncertainty across all of these factors influences the investment risk related to that resource. According to a recent report prepared for the National Association of Regulatory Commissioners (NARUC), uncertainties are inherent in both supply-side as well as demand-side resources and, therefore, the evaluation of demand-side resources may in principle be no more uncertain than the evaluation of supply-side resources (Raab and Violette 1994):

Supply-Side Uncertainties

- Future prices of fuels (natural gas, oil and coal)
- Future availability of supply-side plants (major outages at baseload plants and forced and unforced outage rates)
- Capital costs of plants
- Operating costs of plants
- Changing environmental regulations
- Demand and energy forecasts
- Development time frames
- Licensing and construction time frames
- Public opposition
- Future regulatory structure

Demand-Side Uncertainties

- Uncertainty in DSM impacts (operating assumptions, interactions between measures, persistence of savings, projections of program participation, estimated technical potential)
- Uncertainty in DSM costs (program marketing, measure installation)
- Level of free riders, spillover (including free drivers), and snapback
- Demand and energy forecasts
- Future regulatory structure

Compared to uncertainties in long-run forecasted (incremental) demand levels (which may be off as much as +/- 50%), estimated DSM impacts may be viewed as relatively accurate (*ibid.*). For example, the North American Electric Reliability Council has experienced large errors in forecasting demand: summer peak demand was projected in 1973 to reach 734 GW by 1983, but actual demand was 448 MW, a 39% error (DOE 1995). Uncertainty bands around *load growth* may be even greater and is a more important variable because load growth projections (especially, peak load growth as well as transmission and distribution growth) are used by system planners to determine incremental resource acquisition needs.

Nevertheless, compared to the supply side, demand-side uncertainties are relatively new and, therefore, are of more interest to regulators, utility staff, and intervenors. Measures of reliability are needed for characterizing the uncertainty of DSM, as discussed in the next section.

Measures of Reliability

The "mean" (average) estimate is typically regarded as the best estimate of DSM impacts. However, two other determinants, the level of bias (e.g., systematic omission of key variables, see below) and precision (acceptable variation around the estimated mean load impacts) of DSM program impact estimates, are needed to assess the reliability of DSM as a resource. A good example of the relationship between bias, precision and reliability is presented in the following:

> ". . . imagine a bull's eye target. If your shots are tightly placed in the center, then your rifle (measurement and evaluation study) is unbiased and precise. It is reliable. If however, your shots are in a tight pattern but all to the left of the target, then your rifle is precise, but biased. An imprecise but unbiased rifle would produce a pattern that is centered about the bull's eye, but widely scattered. Consequently, a study may be quite precise but unreliable because it is biased; or unbiased but still unreliable because it has too great a variance." (Messenger et al. 1994)

Most practitioners feel that evaluations should produce credible results that focus on the elimination of bias and attempt to be precise, but are flexible in levels of precision.[1] It is hoped that these unbiased estimates become more precise over time.

---

[1] Examples of categories of bias include the following: nonrepresentative samples, self-selection in participating in programs, contamination from other programs, measurement error in variables examined, and omitted variables.

The degree of credibility that may be attached to results is expressed by the level of statistical confidence (e.g., 90% confidence). This is in contrast to the precision of the estimate, which is gauged by the width of the confidence interval itself. Confidence and precision are competing ends. For a fixed sample size and variance, a reduction in the interval width, causing greater precision, can be achieved only at the expense of reducing the level of confidence, and vice versa. The only way to increase both confidence and precision is to collect a larger sample, but there are costs associated with this (see below). Thus, precision levels (and our confidence in savings results) are typically driven by budgets, not *a priori* accuracy criteria. And the budgets will affect the type of evaluation methods (e.g., econometric methods based on whole-premise billing data, or metering methods utilizing information on specific equipment installed) used to estimate energy savings (and vice versa), also affecting the uncertainty of the evaluation results and program cost-effectiveness (Sonnenblick and Eto 1995).[2]

As noted above, the accuracy of estimates of DSM impacts is generally reported using a measure of precision at a given level of confidence. In conventional medical studies, where the risks of error can be life threatening, researchers are "confident" that their findings are correct if there is less than a one percent chance that the true population mean is 2.5% or above, or 2.5% below, the study estimate (this is called a "99/5" decision rule) (Horowitz 1992). In social sciences, researchers often report their results at the 95% confidence level (implying a willingness to accept a 1 in 20 random event to be misinterpreted). Many suggestions regarding what might be reasonable confidence and precision levels for DSM impact evaluation are based on the experience of load research. Load research has long had a standard of 90% confidence and 10% precision ("90/10" rule), but load research simply measures a level of consumption, and not a change in consumption which is less predictable and which is relatively small. Thus, the 90/10 standard is too stringent for evaluating DSM savings. In general, most experts agree that the precision guideline for load research is not suitable for the DSM savings situation where changes are being measured, in contrast to the one-time, static precision of consumption (as in load research) (Hanser and Violette 1992).

[2]Different methods are subject to different uncertainties that will result in different estimates of precision. For example, some methods (e.g., multivariate regression) have relatively strict assumptions (e.g., normality and little correlation among independent variables).

Moreover, these specifications of confidence are conventions only:

> ". . . there is no absolute standard for when to be confident in study findings; acceptable precision levels depend on the conventions of the field of study. The costs of error, the costs of measurement, what is technically achievable given the available measurement tools, and finally, the tolerable level of uncertainty, all play a role in establishing reasonable and prudent standards. . . . the field of DSM impact evaluation has yet to agree on standards for confidence or required precision levels." (Horowitz 1992)

A recent study assessing the uncertainty in estimates of DSM program cost-effectiveness when evaluation methods of varying precision and accuracy are used found that the imprecision in the cost of conserved energy was significant for programs with mean total resource cost ratios close to one, while higher ratios seem to guarantee cost-effectiveness even with significant estimate imprecision (Sonnenblick and Eto 1995). However, they found that biased savings could threaten the confidence of cost-effectiveness estimates for programs with ratios approaching 2.0, especially when estimate imprecision was considered. Thus, the bias of the savings estimates is probably just as important (if not more) than the precision of the estimates.

Another issue not examined is at what level should precision standards be applied? For the whole DSM portfolio, all DSM programs for a specific sector, an individual DSM program, or individual measures? They all have different methodological and cost implications.

In conclusion, confidence intervals and precision levels are important, but have sometimes been over emphasized. A practical approach is needed for determining the level of precision and should reflect realistic expectations given the technical, economic, and practical limitations in the measurement of energy savings (Schlegel et al. 1991; Sonnenblick and Eto 1995).

Side Effects of Precision Requirements

DSM evaluation is not an exact science and requires the use of judgment and interpretation for assessing the performance of utility DSM programs. Making sure that the evaluations of these programs can provide the level of confidence required by regulators to make decisions about investments in certain types of DSM programs is critical; however, too much focus on the precision of the results may be detrimental if the "big picture" is not taken into account. As a result, the setting of precision levels may become threshold precision values that the utility will believe it has to exceed with its impact evaluation before investing in a DSM program:

> "If these accuracy levels are not achieved, the utility may be concerned about penalties. Under these conditions, if commissions require high levels of precision, utilities will be given an incentive to only invest in those programs that produce large, readily identifiable savings where impact evaluations are likely to meet these potentially restrictive accuracy requirements. Other programs that are likely to be cost-effective but, for example, represent small percent reductions in consumption when compared to a customer's total consumption may not be undertaken." (Raab and Violette 1994)

Similarly, the production of precise impact results to qualify for shareholder incentives may discourage utilities from attempting to evaluate some of the more difficult program parameters, such as spillover effects and market barrier costs, as described in a recent report on California's measurement and evaluation load impact studies (Messenger et al. 1994). Market evaluations that use baseline studies, conjoint methods, and trade ally and manufacturers' data should be able to estimate precision around program spillover effects, and the neglect of such measures may lead to biased (and, therefore, less credible) results.

Regulatory Policies

NARUC Report

Although the National Association of Regulatory Commissioners (NARUC) has not formally adopted a policy on reliability and uncertainty, a recent NARUC white paper has offered the following recommendations:

> "PUCs should not require higher levels of accuracy than is cost-effective to achieve. Moreover, PUCs should not set standards for accuracy that are more stringent than those required of utility management when making investment decisions in supply-side resources or other utility investment decisions. PUCs may also want to consider setting some lower bound accuracy levels, but should be flexible to account for the wide range of program types and other factors." (Raab and Violette 1994)

EPA's Conservation Verification Protocols

The U.S. Environmental Protection Agency (EPA) recently designed a set of Conservation Verification Protocols (CVP) as part of its mission to implement the Acid Rain Program of the Clean Air Act Amendments of 1990 (EPA 1993). The CVP was designed

> ". . . to be rigorous without being burdensome on the utility or the regulator. The CVP has the added benefit of helping to ensure the cost-effectiveness of utility conservation programs and SO_2 emission reduction measures, as well as the reliability of energy savings from the measures." (EPA 1993)

For purposes of the emissions allowances, the objective of the CVP is to award allowances for savings that occurred with reasonable certainty. The CVP requires that the savings are expressed in terms of the utility's confidence that the true savings were equal to, or greater than, those for which it applied. Thus, the CVP uses a 75% level of confidence using a one-tailed test (no specific precision level is targeted): in other words, the reporting entity must be statistically confident (at the

75% level) that the minimum level of energy savings has been achieved. The authors of the CVP note that their approach differs from the more stringent procedures employed by some electricity rate regulators, but argue that the CVP procedure ". . . offers utilities more flexibility, smaller sample sizes, and the opportunity to claim some legitimate savings even when the evaluation itself was not as successful as planned" (*ibid*).

The authors conclude by noting the following:

> "The CVP takes this approach because while it is not based on usual statistical standards, it reflects the state-of-the-art for reasonable impact evaluation of savings from utility conservation programs." (*ibid*)

In summary, EPA considers a one-tailed test appropriate for most DSM applications because the real concern is not that there are too much savings, but rather that there are too little savings for the program to be cost-effective. Emphasis is thus placed on the lower bound only.

The Massachusetts DPU Decision

The Massachusetts Department of Public Utilities (MDPU) is the only state regulatory commission to have formally addressed the issue of uncertainty and precision. The MDPU reviewed the issue of confidence and precision in the early 1990s and recommended that a 90/10 rule be used. In 1992, the MDPU reviewed and revised its earlier order. In the 1992 proceeding, the Boston Edison Company presented evidence that precision levels in impact evaluation could be as high as +/- 70% at a 90% confidence level for some programs (MDPU 1992). In that proceeding, the MDPU retreated from the 90/10 standard, stating:

> "The Company had correctly pointed out statistical errors underlying the Department's earlier reliance on the 90/10 standard. The Department directs the Company to seek the best precision it can expect to attain with a 90% statistical confidence, subject to the constraint that the marginal value of the precision attained should not exceed the marginal cost of attaining it." (*ibid*)

In that order, the MDPU also indicated that it expected kWh to be measured more precisely than kW, due to the greater costs of measuring capacity. Furthermore, the MDPU "found it appropriate and cost-effective to seek similar precision levels from one program to the next in terms of absolute kWh or kW, rather than in percentages." This results in tighter relative precision levels (i.e., precision expressed in percentage terms) for programs with greater savings than for programs with smaller kWh and kW savings.

In the NARUC report mentioned above, the authors note that it still was not clear that "even achieving a targeted absolute kWh precision rather than a constant targeted relative precision across most programs would be reasonable given the differences in programs" (Raab and Violette 1994). The authors assert that different programs target end-uses within different customer groups, and the relative savings impact will be different depending on total consumption and predicted impact.

California, New Jersey, and Pennsylvania

In 1992, the California Collaborative (the major investor-owned utilities, the California Energy Commission, the Division of Ratepayer Advocates, and the Natural Resources Defense Council) prepared a set of measurement and evaluation protocols that were later approved by the California Public Utilities Commission (CPUC 1993). Confidence intervals and levels of precision were discussed in workshops on the development of the protocols, but were not included in the protocols for reporting energy savings for the following reasons (personal communication from Don Schultz, CPUC, Division of Ratepayer Advocates, Sept. 26, 1994):

- It was impossible to determine a *reasonable* standard (reference point).
- Applying a reference point would be burdensome and challenging (e.g., holding a utility (or an evaluator) to a certain standard).
- Nobody wanted to rely on one statistical reference point.
- There is always a potential for legitimate error - it was best to try to minimize error during the process of collecting and analyzing data.

The New Jersey Board of Regulatory Commissioners (NJBRC) adopted evaluation protocols that are almost exactly the same as the verification protocol developed by the National Association of Energy Service Companies (NJBRC 1993), but the protocols have no guidelines on confidence and precision levels for savings (personal communication from Bill Brady, NJBRC, Sept. 29, 1994).

Finally, the Pennsylvania Public Utilities Commission is in the process of preparing evaluation guidelines that must be followed by the state's utilities, and the level of precision to be required of DSM impact evaluations is set at a target confidence level of 75%, similar to the EPA standard (Hastie 1995).

Utility Estimates of Relative Precision

Until recently, it was hard to find evaluations that reported estimates of relative precision. Examples from a few utilities are presented below.[3]

Consumers Power Company

In 1991, the Consumers Power Company (CPCo) initiated large-scale DSM programs in the residential, commercial, and industrial sectors (Kushler 1993). In 1992-93, a comprehensive evaluation of these programs was conducted and is the source of the data described below (Vine 1994). In Table 1, we present the relative precision of the net annual energy savings estimates for CPCo's residential and non-residential programs at the 90% confidence level. In the residential sector, the Residential Free Install was the most precise (10%), while the Appliance Recycling program was the least precise (81%).[4] The levels of precision for the Mail Order Catalog and the Water Heater Conversion program were in the 20-26% range, while the Residential Rebate Coupon program had a 41% relative precision. In the non-

[3]The programs were selected after reviewing program evaluation reports in the DEEP Library maintained at Lawrence Berkeley Laboratory. The goal of the Database on Energy Efficiency Programs (DEEP) Project is to compile and analyze the measured results of energy efficiency programs in a consistent and comprehensive fashion (Vine et al. 1993). The DEEP Library contains over 600 evaluation reports from around the country and Canada. The three programs examined in this paper were selected because they contained the data needed for measuring the precision of DSM impacts, data that, until recently, were often missing in DSM program evaluation reports.

[4]The precision estimate indicates the relative magnitude of the difference between the low (or high) estimate and the mean estimate. For example, at the 90% confidence level, the average net energy savings for the Residential Free Install program were 13,005 MWh and ranged from 11,700 MWh to 14,316 MWh. The precision level was 10%: (13,005-11,700)/13,005.

residential sector, the Non-Residential Free Install and Direct Rebates programs, which were primarily lighting programs, each had a reasonably good level of precision (16%). The Custom-Designed Rebates program, due to the diversity of measures applied, was the least precise (57%).

Pacific Gas and Electric Company

Pacific Gas and Electric (PG&E) analyzed the relative precision of the gross annual energy savings estimates from its 1991 and 1992 commercial and industrial programs at the 90% confidence level (Table 1) (PG&E 1993). The most precise annual savings estimate was from its lighting rebate program (14%). The refrigeration rebate program was the least precise (33%), and the HVAC rebate program had medium precision (28%).

New England Electric System

The New England Electric System (NEES) reported estimates of the relative precision of its 1993 DSM programs offered by one of its companies, the Massachusetts Electric Company (Table 1) (NEES 1994). Some programs had relative precision levels of 100%, but most of these had small load impacts. Excluding these programs, the Residential Space Heating program was the most precise (23%) residential program while the Appliance Recycling program was the least precise (58%). In the non-residential sector, the most precise was the new construction program (Design 2000) targeting motors (23%) and the least precise was the Design 2000 program targeting lighting (38%).

Table 1. The Relative Precision of Energy Savings of Selected DSM Programs

Utility/Reference	Sector	Program Name	Program Year	Net Annual Energy Savings (GWh)	Gross Annual Energy Savings (GWh)	Relative Precision at 90% Confidence Level (%)
CPCo/Xenergy 1994	Residential	Appliance Recycling	1992-93	15		81
CPCo /Xenergy 1994	Residential	Free Install	1992-93	13		10
CPCo /Xenergy 1994	Residential	Rebate Coupon	1992-93	8		41
CPCo /Xenergy 1994	Residential	Mail Order Catalog	1992-93	0.3		26
CPCo/Xenergy 1994	Residential	Water Heater Conversion	1992-93	4		20
CPCo /Quantum 1994	Non-Residential	Free Install	1992-93	10		16
CPCo /Quantum 1994	Non-Residential	Direct Rebates	1992-93	128		16
CPCo /Quantum 1994	Non-Residential	Custom-Designed Rebates	1992-93	91		57
PG&E/Xenergy 1993	Non-Residential	CIA Rebates - Lighting	1991-92		485	14
PG&E/Xenergy 1993	Non-Residential	CIA Rebates - Refrigeration	1991-92		49	33
PG&E/Xenergy 1993	Non-Residential	CIA Rebates - HVAC	1991-92		124	28
NEES/NEES 1994	Non-Residential	Design 2000 & Energy Initiative - Lighting	1993	600		38
NEES/NEES 1994	Non-Residential	Design 2000 & Energy Initiative - Motors	1993	67		23
NEES/NEES 1994	Non-Residential	Design 2000 & Energy Initiative - Variable Speed Drives	1993	112		30
NEES/NEES 1994	Non-Residential	Design 2000 & Energy Initiative - HVAC	1993	31		100
NEES/NEES 1994	Non-Residential	Design 2000 & Energy Initiative - Food	1993	3		100
NEES/NEES 1994	Non-Residential	Design 2000 & Energy Initiative - Custom	1993	167		26
NEES/NEES 1994	Non-Residential	Performance Engineering	1993	73		34
NEES/NEES 1994	Non-Residential	Small C&I	1993	215		31
NEES/NEES 1994	Residential	Appliance Recycling	1993	15		58
NEES/NEES 1994	Residential	Complementary Program	1993	0		100
NEES/NEES 1994	Residential	Energy Crafted Home	1993	4		50
NEES/NEES 1994	Residential	Energy Fitness	1993	33		100
NEES/NEES 1994	Residential	Multi-Family Retrofit	1993	106		53
NEES/NEES 1994	Residential	Residential Lighting	1993	83		38
NEES/NEES 1994	Residential	Residential Space Heating	1993	128		23
NEES/NEES 1994	Residential	Water Heater Rebate	1993	2		100

Cross-Utility Comparisons

Comparing the precision levels of utility programs within a utility as well as between utilities is a difficult endeavor, since precision levels vary by evaluation methods, savings estimate (gross versus net savings), targeted measures and markets, program evaluation budgets, and utility experience in program design and evaluation. In Table 2, we compare precision levels for CPCo and NEES at a more aggregate level - by sector (residential and non-residential) and for all programs. For NEES's programs, the relative precision at the 90% confidence level was 16% for all of its programs, 20% for its non-residential programs, and 21% for its residential programs.[5] For CPCo's programs, the relative precision at the 90% confidence level was 18% for all of its programs, 21% for its non-residential programs, and 32% for its residential programs. Overall, the precision results from the companies appear quite similar and, for the most part, demonstrate reasonably good precision levels around DSM savings estimates.

Table 2. The Relative Precision of Energy Savings By Sector

Utility/Reference	Sector	Program Year	Net Annual Energy Savings (GWh)	Relative Precision at 90% Confidence Level (%)
CPCo/CPCo 1994	Residential	1992-93	40	32
CPCo/CPCo 1994	Non-Residential	1992-93	229	21
CPCo/CPCo 1994	TOTAL	1992-93	269	18
NEES/NEES 1994	Non-Residential	1993	1,268	20
NEES/NEES 1994	Residential	1993	371	21
NEES/NEES 1994	TOTAL	1993	1,639	16

[5]The combined relative precision for all of the programs is less than the relative precision for the residential and non-residential sectors because random discrepancies tend to offset one another (NEES 1994).

Recommendations for Reducing Uncertainty

A number of opportunities exist for increasing our understanding of the factors leading to DSM uncertainty and for reducing the level of DSM uncertainty. We briefly recommend below some activities that the evaluation and regulatory communities can undertake to take advantage of these opportunities.

Historical data analysis

- Devise tests and develop procedures to detect and correct bias.
- Support statewide, regional, and national efforts that compile and synthesize results from many impact evaluations (e.g., DEEP) (Vine et al. 1993).

Future program evaluations

- Conduct process evaluations that focus on key uncertainty parameters (including evaluation methods) affecting precision and bias.
- Conduct evaluations of the same program periodically to see if uncertainty decreases over time.
- Use multiple evaluation methods to see how uncertainty varies with methodology (Sonnenblick and Eto 1995).

Program design

- Conduct social science experiments to determine critical parameters of uncertainty.
- Once critical uncertainty parameters have been identified, redesign program to reduce uncertainty and then evaluate new program.

Risk assessment and decision analysis

- Assess the risk of not pursuing DSM because of uncertainty problems. Evaluate the risks associated with not pursuing DSM versus pursuing DSM, or with reduced levels of DSM (Buller and Miller 1992)[6]

- Weigh the value of reducing uncertainty? Since there will always be uncertainty, focus on achieving the optimal decision given certain levels of confidence, precision, and risk. Decision theory may be helpful for this issue, including examining the "penalty" (cost) associated with a utility ignoring DSM, even though savings may occur from DSM.

Measurement and evaluation guidelines

- Prepare guidelines for achieving cost-effective accuracy levels (similar to MDPU decision). Consider how to obtain a given reduction in uncertainty in the most cost-effective manner.

- Prepare guidelines for reducing measurement bias, so that key factors are accounted for (e.g., program spillover).

Program evaluation is recognized by many for reducing the uncertainty of DSM program impacts and for enhancing the value of DSM as a resource (e.g., Messenger et al 1994; Sedmak et al 1994). By analyzing billing data and end-use metered data, using large participant and comparison samples, conducting on-site visits, assessing market shipment or sales data, and designing and implementing better customer survey questionnaires, evaluators have been able to reduce the uncertainty of DSM estimates associated with engineering results and self-reported data. On the other hand, these same methods have increased our awareness of the remaining uncertainties underlying the evaluation methods used for estimating energy savings and program cost-effectiveness. Admittedly, much work remains to be done.

As noted above, precision levels vary by evaluation methods, savings estimate (gross versus net savings), targeted measures and markets, program

---

[6]Using sensitivity analysis, Eric Hirst has suggested that DSM programs generally reduce uncertainties in a utility's resource portfolio compared to a portfolio without DSM. And even increasing the cost of DSM programs by 50% had little effect on the conclusions from the sensitivity analysis (Hirst 1992).

evaluation budgets, and utility experience in program design and evaluation. Accordingly, it is difficult to recommend a single precision standard for impact evaluations, due to the importance of individual circumstances. As a constructive recommendation, however, evaluators should estimate and report confidence intervals, as well as means, and they should include a discussion of how the issues of bias and precision were addressed in their research. Finally, it is also important to reiterate that one needs to keep DSM uncertainty in perspective: there is lots of uncertainty in all aspects of resource planning, on the supply side as well as the demand side. That fact, together with the relatively good precision levels demonstrated by the three utilities examined in this paper, suggests that there is no justification for broadly attacking DSM as a resource on the basis of uncertainty of evaluation savings estimates.

Acknowledgments

This work was supported by the Assistant Secretary for Energy Efficiency and Renewable Energy, Office of Utility Technologies, of the U.S. Department of Energy under Contract No. DE-AC03-76SF0098. The authors are especially grateful for the review comments on a previous draft of this paper from the following people: Les Baxter, Joe Eto, Liz Hicks, Ken Keating, Alan Meier, Mike Messenger, George Penn, Diane Pirkey, Rich Sonnenblick, Ken Train, and Robert Uhlaner.

References

1. Buller, S. and W. Miller, "How Should We Treat Factors Contributing to Uncertainty in Measurement and Evaluation of DSM?" *Proceedings of the ACEEE 1992 Summer Study on Energy Efficiency in Buildings*, Vol. 7, pp. 31-35. Washington, D.C.: American Council for an Energy-Efficient Economy (1992).

2. California Public Utilities Commission (CPUC), Decision D.93-05-063, May 19, 1993, Interim Opinion on Ex Post Measurement and Evaluation. San Francisco, CA: California Public Utilities Commission (1993).

3. Consumers Power Company (CPCo), Testimony in Case No. U-10554. Lansing, MI: Michigan Public Service Commission (1994).

4. Hanser, P. and D. Violette, "DSM Program Evaluation Precision: What Can You Expect? What Do You Want?" *Proceedings of the Fourth National Conference on Integrated Resource Planning*, pp. 299-313. Washington, D. C.: National Association of Regulatory Commissioners (1992).

5. Hastie, S., "Upheaval in Pennsylvania Over DSM Cost Recovery," *Evaluation Exchange* 5(1):1-2 (1995).

6. Hirst, E., "Effects of Utility DSM Programs on Risk," ORNL/CON-346. Oak Ridge, TN: Oak Ridge National Laboratory (1992).

7. Horowitz, M., "Savemetrics: The Science of Measuring DSM Impacts," *Proceedings of the 1992 International Energy Efficiency & DSM Conference,*. Bala Cynwyd, PA: SRC International (1992).

8. Kushler, Martin G. "A Two-by-Four and a Pound of Cheese: A Case Study of the Effect of Regulatory Incentives on a Reluctant Utility." *Proceedings of the 1993 International Energy Program Evaluation Conference*, pp. 362–367. Chicago, IL (1993).

9. Massachusetts Department of Public Utilities (MDPU), Case DPU 90-335, April 8, 1992.

10. Messenger, M., N. Stone, B. True, A. Kandell, P. Purcell, and J. Lang, "Utility DSM Program Measurement and Evaluation Studies," Staff Report, Docket No. 93-ER-94, Sacramento, CA: California Energy Commission (1994).

11. New England Electric System (NEES), "1993 DSM Performance Measurement Report," Westborough, MA: New England Electric System (1994).

12. New Jersey Board of Regulatory Commissioners (NJBRC), "Measurement Protocol for Commercial, Industrial and Residential Facilities," Trenton, NJ: New Jersey Board of Regulatory Commissioners (1993).

13. Pacific Gas and Electric Company (PG&E), "Annual Summary Report on Demand Side Management Programs in 1992 and 1993: Technical Appendix, April 1993," San Francisco, CA: Pacific Gas and Electric (1993).

14. Quantum Consulting, Inc., "Consumers Power Company's REDUCE THE U$E Program Non-residential Evaluation Report." Berkeley, CA: Quantum Consulting (1994).

15. Raab, J. and D. Violette, "Regulating DSM Program Evaluation: Policy and Administrative Issues for Public Utility Commissions." Washington, D.C.: National Association of Regulatory Utility Commissioners (1994).

16. Schlegel, J., R. Prahl, and M. Kushler, "Measurement in the Age of Incentives." *Proceedings of the 1991 International Energy Program Evaluation Conference*, pp. 182-190. Chicago, IL (1991).

17. Sedmak, M, R. Uhlaner, and B. Smith, "Building Reliable DSM Resources with Program Evaluation," *Proceedings of the ACEEE 1994 Summer Study on Energy Efficiency in Buildings*, Vol. 8, pp. 177-185. Washington, D.C.: American Council for an Energy-Efficient Economy (1994).

18. Sonnenblick, R. and J. Eto, "Calculating the Uncertainty in Program Cost-Effectiveness Estimates," *Proceedings of the 1995 International Energy Program Evaluation Conference*. Chicago, IL (1995).

19. U.S. Department of Energy, "Performance Issues for a Changing Electric Power Industry," DOE/EIA-0586. Washington, D.C.: Energy Information Administration, U.S. Department of Energy (1995).

20. U.S. Environmental Protection Agency (EPA), "Conservation Verification Protocols," EPA 430/8/B-92-0 02. Washington, D.C.: U.S. Environmental Protection Agency (1993).

21. Vine, E., "Comparative Evaluation of Consumers Power Company's Energy Efficiency Programs." LBL-36202. Berkeley, CA: Lawrence Berkeley Laboratory (1994).

22. Vine, E., C. Payne, and R. Weiner, "Comparing the Results of Energy Efficiency Programs: The Creation of a National Database on Energy Efficiency Programs (DEEP)," LBL-33654. Berkeley, CA: Lawrence Berkeley Laboratory (1993).

23. Violette, D., "Analyzing Data," in "Handbook of Evaluation of Utility DSM Programs," pp. 51-72. E. Hirst and J. Reed eds., ORNL/CON-336. Oak Ridge, TN: Oak Ridge National Laboratory (1991).

24. Xenergy, "Evaluation of the CIA Retrofit Rebate Program, Final Report," San Francisco, CA: Pacific Gas and Electric Company (1993).

25. Xenergy Evaluation Team (Xenergy), "Consumers Power Company's Residential Demand-Side Management Programs, Final Report." Madison, WI: Xenergy, Inc. (1994).

Ernest Orlando Lawrence Berkeley National Laboratory
One Cyclotron Road | Berkeley, California 94720

Prepared for the U.S. Department of Energy under Contract No. DE-AC03-76SF00098