Status Quo and Gaps for Impact, Attribution and Retention of Behavioral Programs and Beyond

Lisa A. Skumatz, Ph.D., Skumatz Economic Research Associates, Inc., Superior, CO

Introduction

The California Public Utilities Commission (CPUC) and the California Institute for Education and Environment (CIEE) recognized that evaluation / attribution methods have reached a point that they must evolve in order to provide credible evaluation results for the next generation of programs. Two primary factors have complicated the methodologies that have been applied to energy efficiency programs:

- Transition to more behavioral, outreach and other non-measure-based programs (education, advertising), making it especially hard to "count" impacts, and
- Increased chatter in the marketplace, in which consumers may be influenced by any number of utility programs by the host / territorial utility (the "portfolio") as well as influences from outside the territorial utility (national, neighboring programs, movies / media, etc.).

The project conducted a comprehensive review of current state of the art, and identified best practices and gaps in methods for attribution and evaluation of traditional and behavioral programs. The project used interviews, literature review, and analysis to examine technical, research, and policy issues associated with attribution of savings to programs:

- impact evaluation, or gross savings estimates;
- net-to-gross (NTG) ratios and its components, free ridership, spillover, and takeback, and
- retention of savings.

The Findings

Imapct / Gross Savings: The five main impact methods (M&V, Deemed, statistical modeling, market share / sale, and survey approaches) have generally served to provide gross estimates of programs, even if there are a few issues arising because of the switch toward market and behavioral programs. What we do find is that in many cases, the behavioral programs have not been set up to facilitate evaluation; random assignment is not used, and sample sizes are not strong. However, results for a few programs make it clear the methods are valid. Our poster also summarizes a number of other issues associated with the impact side of evaluation, including best practices suggestions, gaps and methodological improvements, baseline and overlap isies, and the problems of assigning "protocols" to programs that need more "tailored" and flexible approaches.

NTG: The project reviewed results of net-to-gross (and component) estimations from around the US to identify patterns in results for "categories" of programs, and examined best practices in estimation methods used to date for traditional measure-based programs. We found considerable variation in NTG and component results. We examined policies used by different states – whether NTG or its components are used at all, whether "deemed" levels are used, or whether

the regulators endorse or include NTG figures based on primary research. We reviewed and compared protocols from several states, and examined strengths and weaknesses of the approaches. Beyond reviewing the "state of the art" in traditional attribution work, we also examined research and policies related to attribution for behavior, education, and training-based programs, and issues related to disentangling the effects from multiple programs and deliverers (marketplace "chatter"). Most importantly, little work on the NTG associated with pure behavioral programs has been conducted, and the lack of random assignment has made it difficult to explore the results for programs. Most importantly, we note that many states and utilities ignore spillover in computations of net-to-gross, and since spillover can be an important element of outreach and behavioral programs, the omission hurts (or *would* hurt, if it were more commonly measured) the net savings figures assigned to these programs.

Retention: Measure lifetimes are another critical element in the computation and attribution of savings to programs – computations that are important in credibly assessing remaining energy generation needs, as well as rewards and incentives for providers of programs. The measure lifetime analysis literature and methodology is fairly robust. More than 100 studies have been conducted, examining *in-situ* median lifetimes for residential and non-residential measures. However, we found virtually no measure lifetime studies addressing behavioral programs. We note there are adjustments needed for the traditional measurement methods (timing, and recognition of "partial" retention of behaviors. We also addressed technical degradation, and remaining useful lifetimes, in our research, and present key results in the poster.