SUMMIT BLUE
CONSULTING

Perspectives on DSM Program Evaluation

DSM Informational Workshop
Public Utilities Commission

Denver, Colorado
February 8, 2007

By: Daniel Violette, Ph.D.
Summit Blue Consulting
Boulder, Colorado 80302
e-mail: dviolette@summitblue.com

DSM Programs – A View

- Delivery of DSM programs are challenging and take time to roll out.
  - Program concepts
  - Value propositions and customer assessments
  - Marketing
  - Make the Sales
  - Delivery Channels – appropriate industry infrastructure
  - Fulfillment (get the service or technology to the customer)
  - Quality control
  - Financial accounting
- Designing a new DSM program is similar to the development of a new product or service with the same set of challenges.
- Not all programs will be successful, and many will need a "shake-out" period before they become successful.

Agenda

- Background on Role of Evaluation
- Types of Evaluation
- Baselines in Impact Evaluation
- Net Impacts – Free Riders and Spillover
- Cost-Effective Evaluation
- Rigor and Accuracy
- Non-Energy Impacts (NEIs)
- Trends in Evaluation

Why Evaluate Demand Side Resource (DSR) Programs?

- All programs pose challenges. Evaluation helps ensure that objectives and expectations for programs are attained.
- Evaluation transforms guesses, initial estimates, and tracking data into information on program performance.
- Evaluation provides for accountability – Implementators can view evaluation as a negative designed to diminish what they have accomplished.
- Instead, it should be viewed as the proof of accomplishment and a way to improve programs
- Taken together, evaluations are tools for improving program performance.

SUMMIT BLUE
CONSULTING

Why Evaluate DSR Programs? (cont.)

- A good, *on-going* evaluation effort can help focus efforts on attaining achievable impacts with cost-effective program designs.
- Evaluation helps to improve programs:
  - *Where* are the energy savings being achieved?
  - At *what cost* are impacts being achieved?
  - Which consumers *do* participate and why?
  - Which consumers are *not* participating and why?
  - Which consumers *drop out* and why?
- And, assess program impacts:
  - Overall, examine what is changing in the market for DSM products and services as a result of the offered programs.

SUMMIT BLUE
CONSULTING

Evaluation Objectives

- The *overall objective* is to provide the information necessary to make good decisions regarding investments in DSM.
- Evaluations can:
  - Estimate energy and capacity reductions (kWh, kW, therms, or Btu).
  - Assess any changes in quality and reliability of service.
  - Determine the costs of programs.
  - Determine consumer satisfaction and acceptance of the program.
  - Translate program impacts into environmental changes.
  - Assess other non-energy benefits/impacts – e.g., comfort, safety, ease of resale for a home or building.

Types of Evaluation and Assessments

1. **Program Design or Theory Assessment:**
   - Most recent program designs begin by developing a "program theory" and a "logic diagram" that shows how the program will work in the market (i.e., influence customers, trade allies, cross-sell, fulfill, and attain objectives) – it lays out the assumptions and logic of program.
   - A program design or theory assessment examines these initial hypotheses for consistency with what is observed in the field (e.g., workshops with ESCOs will increase the ability of ESCOs to close deals with the support of a utility program)
   - These efforts can become more important as programs trend towards market transformation efforts by setting out market indicators that can be tracked over time.
2. **Process Evaluation:**
   - A systematic assessment of program operations and processes with recommendations for improving the programs' delivery efficiency.

SUMMIT BLUE CONSULTING

Types of Evaluation & Assessment (cont.)

3. Market Characterization and Assessment:

– Characterization generally is a description of a specific market or market segments, including the types and number of buyers and sellers in the market, key actors, type and number of purchases and transactions that occur.

– Assessment generally looks at trends in the market over time including 1) changes in the structure or functioning of a market; 2) the behavior of participants in a market that result from one or more program efforts.

– Typically these efforts are designed to consider market indicators that can be tied to the adoption of energy-efficient products, services or practices related to program activities.

4. Impact Evaluations:

- Estimates the change in energy usage and other targets (such kWh, kW, or Therms) due to DSM programs.

SUMMIT BLUE
CONSULTING

Evaluation Types – Recent Trends

- Program theory, process, market and impact evaluations / assessments are not new.
  - In many cases, they were embedded in the process and impact evaluation efforts conducted previously.
  - Defining these efforts helps to provide more focus on the workings of the market in which a DSM program is meant to function.
  - Increases the emphasis on measurable market indicators:
    - Number of LEED certified builders.
    - Awareness of products and technologies.
    - Market share of energy efficiency products and practices
    - Number of firms engaged -- A&E firms, ESCOs, lighting designers, etc.
- All of this information can be used to support estimates of the accomplishments of a DSM program or portfolio.

Impact Evaluation

- Impact evaluations are based on either:
  - M&V results for a sample of sites or projects within the program portfolio and/or
  - Statistical/regression analysis of utility consumption data
  - Impact evaluation often includes assessments to adjust gross demand savings in order to derive net demand savings.
- Key Issues:
  - To measure a change in energy or demand usage, a baseline must be estimated against which the change can be calculated.
  - Estimating appropriate baselines has proven to be one of the most significant challenges for both EE and DR.
  - A correct baseline will address issues like spillover and free riders.

Impact Evaluation (cont.)

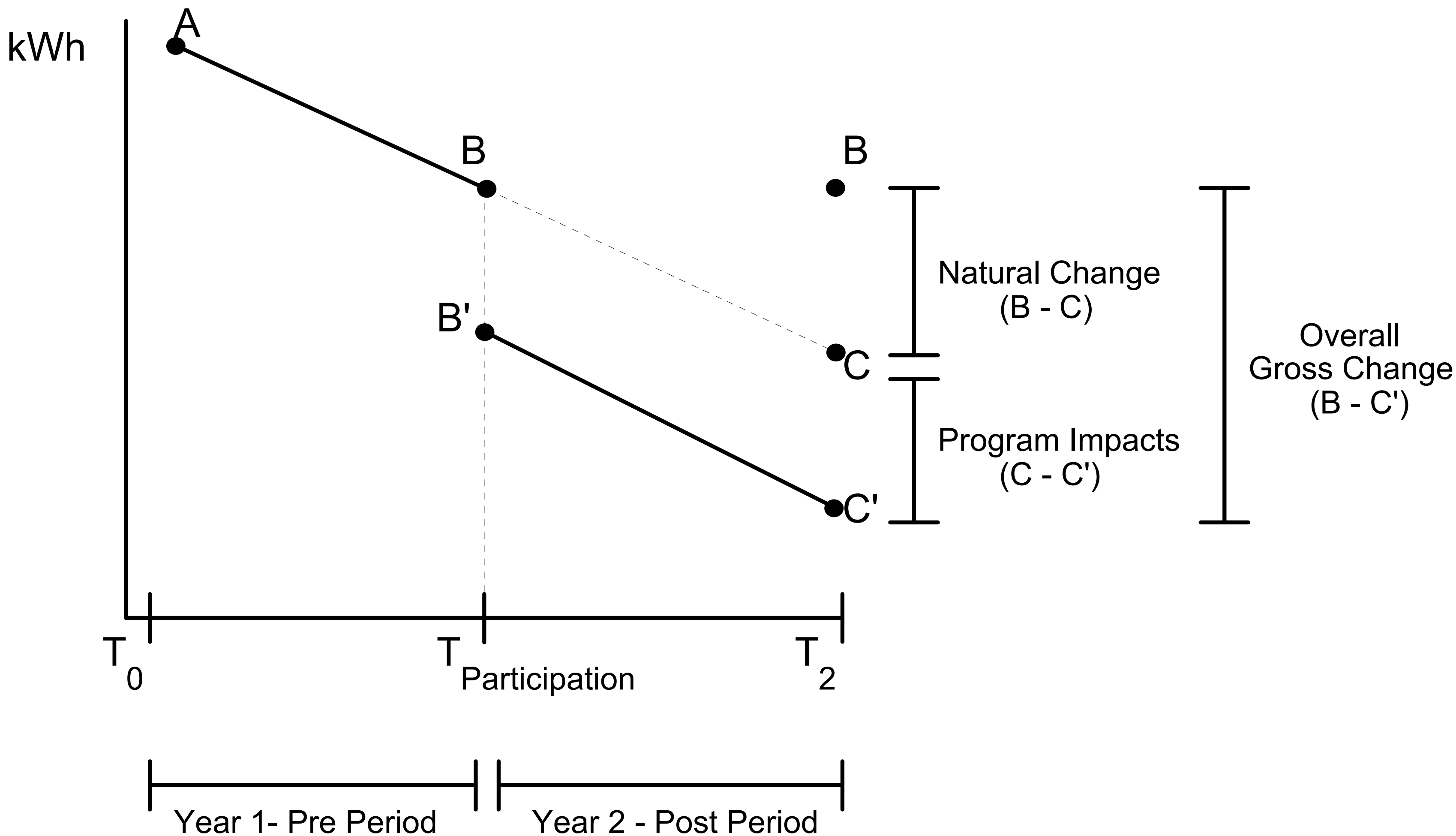
- Data Sources
  - Program tracking databases
    - ♦ Critical to the success of any evaluation – identifies participants, number of units, site data (all as appropriate).
    - ♦ Should include initial information on savings expected per unit (e.g. CFL) or sites as they participate in the program.
    - ♦ Now, tracking gives you a “best” estimate of program accomplishments given design data.
    - ♦ Impact evaluation validates these program tracking estimates.
  - Participant applications
  - Participant surveys
  - Energy consumption information – billing data, interval & equipment specific metering, engineering algorithms and algorithms accompanied by measurements.

SUMMIT BLUE
CONSULTING

Impact Evaluation - Baseline

- What is a baseline?
  - The change in overall energy use or demand that would have occurred in the **absence** of the program.
- Common approaches for estimating baseline:
  - *Engineering Approaches.* Estimates are made of the energy use with and without the DSM measure using engineering algorithms and/or models. (e.g. DOE-2, eQuest)
  - *Statistical Comparison (or Control) Approaches.* Estimates of baseline energy use are derived from a "comparison" group of consumers or other data that serve as a proxy for the actions program participants would have taken, had the program not been offered. (e.g. billing data analysis, nonparticipant surveys)

Illustration of Impact Calculations



Impact Evaluation – Net Impacts

Net Impacts = Net-to-Gross Ratio x Gross Impacts

- Gross Impacts – total savings of participants in program
- Net-to-Gross (NTG) Ratio – adjusts the gross impacts for the impacts attributable to the program,
  - The NTG ratio attempts to take out the natural change that would have occurred without the program.
- NTG ratio is a function of
  - Freeridership: % of participants that would have made the efficiency improvement (in total or in part) without the program
  - Participant Spillover: additional savings achieved by participants due to their participation in program
  - Nonparticipant Spillover: nonparticipants that made efficiency improvements as a results of the program's outreach and marketing.

SUMMIT BLUE
CONSULTING

Importance of Net-to-Gross (NTG)

- Net impacts are gross program impacts adjusted downward for free riders and upward for participant and non-participant spillover.
- Most every program has some free riders and also some spillover.
- The longer a program is in place:

  Freeridership tends to increase; but, spillover also tends to increase.

- This has led to diverse treatments by states:
  - Some states assume that free riders and overall spillover cancel out.
  - Some require impacts be calculated net of free riders, and no spillover is counted (but spillover may be estimated for informational purposes).
  - One state only does not discount for free riders, but adds in spillover.
  - One state includes free riders and only participant spillover.
  - The trend is to get the best estimate possible and more states are moving toward adjusting for both free riders and spillover. **HOWEVER:**

Importance of Net-to-Gross (NTG)

- Some evaluation methods will only provide estimates of net impacts (regression models with control groups over time).
- Self-report approaches comprise most other methods of estimating freeridership and spillover include:
  - Ask the customer (but do so near the time they made the decision)
  - Ask trade allies (A&E firms and builders)
- These provide good ball park estimates and valuable information for program decisions on such things as rebate levels and market transformation, but costs need to be balanced against benefits.
- The accuracy is hard to quantify. Using a sample, one can calculate a confidence interval/precision, but these are around "what the population would have answered had they been asked the free rider question" -- not the actual freeridership value.
  - Due to various types of survey and self-response biases.

Cost Effective Evaluation

- Keys to a cost-effective evaluation of DSM programs:
  - The most expensive evaluations are those that have to attempt to recreate program data that was not gathered when at the time it was most cost-effective to gather that data.
  - The data tracking system must be developed not only with implementation in mind, but also with evaluation in mind – and it must be maintained.

Steps necessary for a successful evaluation effort:

Step 1. Commitment to evaluation at the outset – at day 1.

Step 2. Development of the DSM tracking system(s).

Step 3. Real-time management of the tracking system.

Step 4. Finalize evaluation strategy based program developments

Step 5. Execute the evaluation strategy (sampling, measurement, etc.)

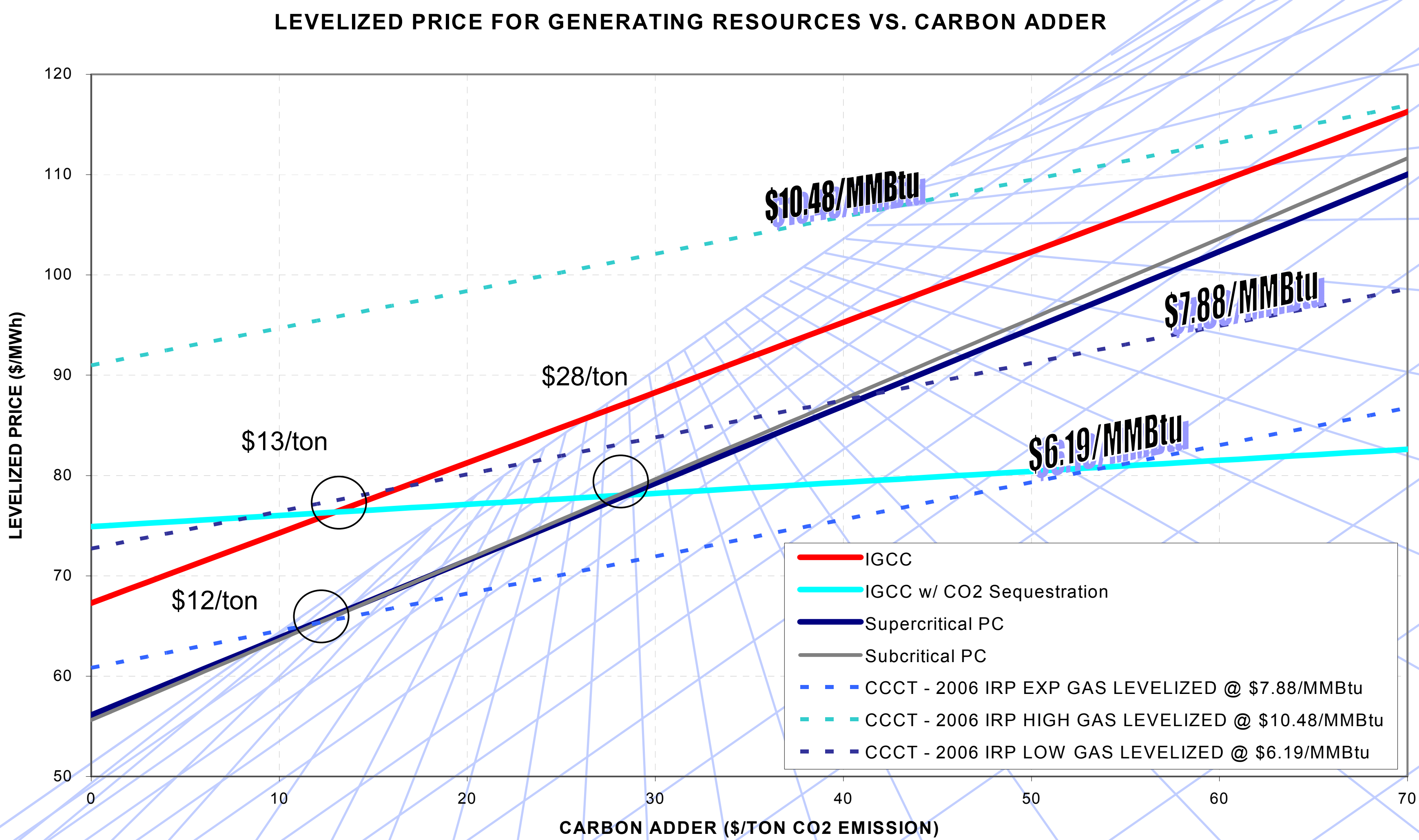
Step 6. Effectively communicate the results

Rigor and Accuracy

- Different programs will be able to attain different levels of precision and accuracy due to inherent variability in the program.
- What level of accuracy is adequate:
  - Since samples and statistics are used in DSM evaluation, it is natural to report confidence intervals and precision levels.
  - Accuracy targets for different outcomes are generally set in the plan, but often are not mandatory. (See California EE Evaluators' Protocols, p.163).
  - Many uncertainties on the supply-side are not amenable to statistical treatment, and the cost of a MWh from planned future supply-side resource may be more uncertain than one might think (see next slide).
  - Examples of accuracy, California EE Evaluators' Protocols use 90/30 for engineering algorithms at the premise level and 90/10 for "enhance rigor studies" at the premise level – moving from the premise level (i.e., one observation) to the program level increases the variance. (p.165)
  - **The goal is to make good decisions** – if an evaluation indicates that a DSM program is a cost-effective resource in the integrated plan at a 80/30 confidence and precision, is this adequate?

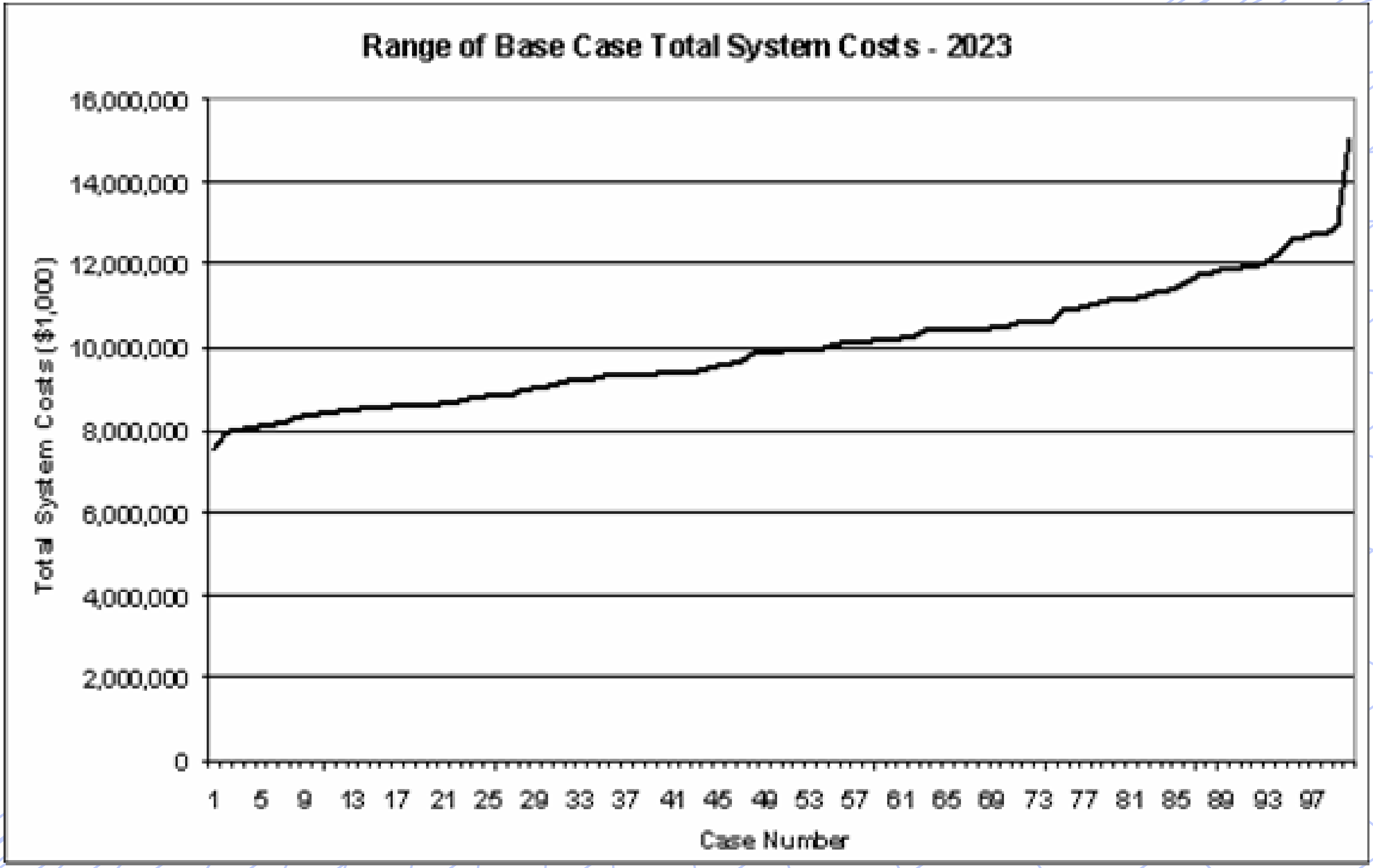
Generation Costs vs. Carbon Adder

(from Idaho Power 2006 IRP: contact Karl Bokenkamp)



System Cost Range for Base Case
(East Cost System)

Monte-Carlo capacity expansion simulation model (100 futures)



Ranges of System Costs by Year
(East Coast System)

- Generally, a 100% increase from low to high in system costs over time.

Ranges of System Costs for Select Years

Range of Total System Costs for Selected Years - Base Case ($ Billions)

Year	2010	2012	2015	2018	2020	2023
Maximum	7.7	8.2	10.2	10.3	12.4	15.0
Minimum	3.5	3.8	5.1	5.6	6.5	7.5
Range	4.2	4.5	5.1	4.6	5.9	7.5
Ratio	118.5%	118.8%	101.7%	82.2%	89.9%	99.3%

Non-Energy Benefits (NEBS) or Non-Energy Impacts (NEIs) Evaluation

- Definition: An evaluation to identify and quantify the non-energy benefits (or costs) associated with a DSM program.
- Some examples of NEB/NEIs include:
  - Reduced emissions & environmental benefits
  - Quality building, materials and construction
  - Productivity improvements
  - Reduced debt and lower levels of arrearage
  - Higher comfort / convenience level of participant
  - Job creation (e.g., among ESCOs)
- Methods include conjoint analyses and self-reports linked to energy savings, e.g., how important was X in your decision to install EE compared to the energy savings.

Evaluation Costs

- Typical evaluation budgets for DSR program portfolios run 4%-10% depending on portfolio and stage of rollout.
- Performing cost-effective evaluation is an important issue
- In many cases evaluation will more than pay for itself.
  - An evaluation costing 8% of program expenditures, that results in improved selection of actions and improved delivery may increase the cost effectiveness of the program by far more than the 8% spent on evaluation.
- A recent trend has been a lowering of budgets for evaluation – parallels an early 1990s trend that was reversed in the mid-1990s – to much was missed in evaluation.

Evaluation Frequency

- Impact evaluations – initial test of concept and then possibly every two years or as required.
- Process evaluations – within a year of launching new programs or after more substantial changes to program occur.
- Market Assessment – generally needs to be ongoing to track annual changes over time in market indicators.

Trends in Evaluation

- Increased use of metering and monitoring as costs have come down.
- Engineering site approaches for a sample of sites is increasing for many programs were only billing analyses had been done.
- Increased use of multiple methods where the data are readily available – confirming site analyses with billing analyses.
- Better integrating evaluation with the implementation process:
  – Surveys sent to people shortly after participating to assess factors that influenced their decision and to gather information on free riders.
- Better engineering analyses and algorithms now used in most implementation processes (e.g. for estimating how much a participant will save).
  – Allows for more cost-effective evaluation methods to be used.

Contact Information

Daniel M. Violette, Ph.D.
Principal, Strategy Practice
Summit Blue Consulting
Boulder, Colorado 80302
Phone: 720-564-1130
E-mail: dviolette@summitblue.com