Key Issues for Illinois Evaluation Framework

Ralph Prahl & Gil Peach
Evaluation Advisors to Stakeholder Advisory Group

November 18, 2008

Overview of Presentation/Discussion

1. Application of evaluation results: retrospective vs. prospective approaches (Ralph)
2. Application of net savings results (Ralph)
3. Approaches to deeming of savings parameters (Gil)
4. Sampling and measurement error (Ralph)
5. Principles governing allocation of resources (Gil)
6. Methods for estimating net savings (Ralph)

NB: We have ordered the issues the way we have because of the interdependencies between the underlying concepts and recommendations in each section. This may not seem clear at first but will become clearer as we proceed.

Overview of Approach

- For each issue we:
  - Introduce and explain the issue
  - Give a broad sense of the range of practices in other states.
  - Briefly discuss strengths and weaknesses of various approaches
  - Recommend an approach for Illinois, and provide our rationale

1. Application of Evaluation Results: Retrospective Vs. Prospective Approaches

- Retrospective means the results are used to determine savings for past program years.
- Prospective means the results are used to adjust savings claims for future program years.
- Results can be applied both prospectively and retrospectively, so the two aren't mutually exclusive.
- Under prospective system:
  - Savings are based on tracking estimates with adjustments and refinements based on ex-post evaluation results derived in an earlier period.
  - Focus shifts to outcomes that can be meaningfully applied to a future time period (e.g., refinement of individual parameter values, overall realization rates)
  - A distinction may be drawn between impact evaluation (empirical measurement of savings) and verification (confirmation that measure was installed and program rules followed). Even under prospective evaluation system, retrospective verification may be required.

Retrospective vs. Prospective: Key Advantages and Disadvantages

	Retrospective	Prospective
Pluses	•Arguably the gold standard in accountability.	•Tends to be more cost-effective •Faster reporting •More certainty, hence less potential for counter-productive risk-avoidance behavior.
Minuses	•Exposes administrator to high level of risk, potentially causing counter-productive behaviors. •Leads to delays in reporting. •May be costly, in that it tends to limit ability to apply results across multiple years.	•Some risk that parameters will change from year to year, leading to errors in savings reporting. •May create incentives for gaming behavior

Retrospective Vs. Prospective: What Some Other States Do

Wisconsin	Prospective, plus retrospective for a maximum of one year.
California (’06-’08 cycle)	Retrospective only, covering three-year program cycle.
Vermont	Prospective only.
Massachusetts	Hybrid. Each year at time of annual report for preceding year, most recent available impact evaluation results are applied to tracked savings for all programs. Results may be from past years (prospective) or current year (mildly retrospective).

Retrospective Vs. Prospective: Recommendations

- We recommend that Illinois adopt a primarily prospective evaluation framework, for the following reasons:
  - Illinois' circumstances lead to high potential for retrospective system causing counter-productive risk-avoidance behavior.
  - Fact that Illinois is starting from scratch argues for a strong focus on gradually and incrementally improving reliability of savings estimates, and for this purpose a prospective system is most effective.
  - Programs will not be mature for several years, making overall ex-post impact estimates unstable; prospectively applied adjustment factors are more likely to remain applicable.
  - Prospective system better able to cope with the 3% evaluation spending cap.
  - Prospective system allows faster determination of outcome of the statutory penalty mechanisms at the end of reporting periods.
- This recommendation is quite specific to the current policy environment, and could change if the policy environment changes.

A Primarily Prospective System: How To Do It

- Apply *impact evaluation* results mainly prospectively, but consider requiring retrospective *verification.*
- Focus impact evaluation on measurement of individual parameters and/or realization rates that can be applied going forward.
- Could be some limited retrospective application of impact results
  - For example, apply all impact results that are available as of the due date for annual report for year X both to future years and to year X.
- Under a prospective system, essential to have a binding impact evaluation schedule, as utility may have incentive to delay activities in order to limit risk.

2. Application of Net Savings Results

Our Read of What the ICC Com Ed and Ameren Orders Say

- The orders indicate that
  - There is *some* role for deemed net-to-gross values, at least initially.
  - Net savings must be estimated empirically at least once during the 1st 3-year cycle, and at least once every 3 years thereafter.
  - The specific deemed net-to-gross values initially proposed by the utilities were not approved.
  - The net-to-gross values derived by the end of the 1st 3-year cycle should be used as the initial assumptions for the 2nd 3-year cycle.
- Beyond the above, there is much regarding the application of net savings results that the orders appear to leave unresolved, including:
  - The scope of the role for deemed values
  - Whether net-to-gross results are to be applied only prospectively, or also retrospectively

Application of Net Savings Results: What Do the ICC Orders Say About Goal Attainment?

- Orders do not appear to explicitly address whether goal attainment should be based on net vs. gross savings.
- Orders do appear to specify that net savings be used for TRC B/C (e.g., Order 07-0540, pp. 3-4, 9)
- Orders later devote several pages to how NTGRs should be derived and applied (e.g., Order 07-0540, pp. 42-45) without any discussion of specific applications of the results.
- My interpretation is that the ICC probably intended that goal attainment be based on net savings.
  - If the intention was that different outcomes be used for different applications, wouldn't this have been mentioned in the Orders?
  - However, other interpretations are possible.

Application of Net Savings Results: General Recommendations

- View the most important functions of estimating net savings as being to:
  - Incentivize administrators to get savings that would not otherwise occur
  - Support the efficient allocation of resources across programs and measures
  - Improve program design and implementation
- Apply net savings results in exactly the same, primarily prospective manner we have recommended for all other parameters.
  - No good reason to have a different system than for gross savings parameters
  - A primarily prospective system represents a reasonable balance between appropriately incentivizing administrators and potentially leading to counter-productive risk-avoidance behavior.
  - Use the same approach for all applications of net savings results (e.g., assessing goal attainment, redesigning programs, B/C analysis)

Application of Net Savings Results: Benefit-Cost Analysis

- Most states east of the Mississippi use net savings in TRC test; many states west of the Mississippi use gross.
- Unclear why a river should have this mysterious effect.
- Ralph and Gil have somewhat different perspectives.
- Ralph:
  - Net savings should be used, because B/C analysis should be focused on the costs and benefits that the program has *caused* to occur. Why should we be concerned about costs and benefits not caused by the program?
  - Com Ed and Ameren orders appear to specify that net savings be used.
- Gil:
  - Gross savings should be used, because B/C analysis should be focused on the goal of getting results in place and as urgency increases many social institutions and forces will be pulling together to create success. The goal is survival past peak oil, peak gas, peak uranium, and climate change.

3. Approaches To Deeming Of Savings Parameters

- Under a primarily prospective system, there is no choice but to deem savings parameters for first year, as there are no existing evaluation results to apply prospectively yet. The focus is then on refining deemed values over time, and/or developing realization rates that are applied prospectively to the deemed savings result.

- To develop initial deemed values, need to: (1) determine for each measure whether you're deeming unit savings, deeming individual parameters, deeming calculation methods, or focusing on development of realization rates to be applied to custom savings estimates; (2) look at what a range of other jurisdictions are assuming, and carefully assess transferability of each source, taking into account factors such as program design, target populations, service territory, climate, etc; (4) look for relevant empirical studies, assessing transferability in the same manner as in #3; (5) develop recommendations. Given the number of measures involved and the number of sources that must be considered, this is not a trivial task – expect it to take contractor teams several months.

Approaches To Deeming Of Savings Parameters

- Today the methods for developing ex ante estimates of gross program savings are well developed. In many cases, simple engineering algorithms (formulas) can be used. In others, simulation models give the best results because the simulations can take many variables into account. But the two main approaches use algorithms or simulation models. (Examples: Connecticut & New York)

- For large industrial settings where DSM savings occur through improvement in manufacturing processes (such as through improvements in handling compressed air, the deeming basically has to satisfy the plant manager and the plant engineers). This usually takes us back to the engineering estimates approach, grounded in the realities of a particular industry and a particular production process

Approaches To Deeming Of Savings Parameters

- Deeming can be as useful as direct impact measurement – depending on level of rigor.

- These approaches blend together in simulation models (model based statistical estimation and model based engineering estimation).

4. Sampling and Measurement Error

BACKGROUND

- Sampling error
  - Typically, funding does not allow for direct examination of all cases in the population of interest, necessitating sampling.
  - Any time sampling is done, there is uncertainty associated with luck of the draw. This is called sampling error.
  - Statistical precision refers to the magnitude of the uncertainty in results due to sampling error.
  - States sometimes have specific standards for the level of statistical precision that impact evaluations should provide.
- Measurement error
  - Can be defined as the likelihood that you are not actually measuring exactly what you think you're measuring, or not measuring it accurately.
  - There are many kinds of measurement error, and it is a MAJOR issue in EE evaluation.
  - Many key controversies can be viewed as disputes about measurement error.
- Together, sampling and measurement error form the two biggest sources of uncertainty surrounding evaluation results.

Examples of Measurement Error

- You use self-reports to estimate the NTGR and the participant cannot give an accurate answer due to cognitive or recall issues.
- You use comparisons of CFL sales in a program and a no-program state to estimate the NTGR, and there are demographic differences between the states that limit comparability.
- You survey HVAC contractors and end up getting mainly those interested in efficiency to respond, causing the results to be unrepresentative of the overall population.
- You use the wrong baseline in estimating gross savings from a custom C&I measure.
- You do a billing analysis to estimate savings from weatherization, and don't fully control for changes in weather.

Standards for Sampling Error Vs. Standards for Measurement Error

- Much of good evaluation planning and implementation boils down to minimizing measurement and sampling error.
- However, there is a radical difference in the ease with which the procedures for minimizing these two sources of uncertainty can be characterized.
  - Methods for limiting sampling error are relatively easy to characterize.
  - Methods for limiting measurement error are not.
- We believe this difference in ease of developing rules sometimes leads to an inordinate focus on statistical precision at the expense of measurement error.
  - Likened to looking for one's keys under the street light because that's the only place you can see, despite the fact that you know that's not where you left them.
- Furthermore, statistical precision improves with increasing sample sizes and aggregation, but measurement error does not, increasing the importance of measurement error at a policy level.

Sampling and Measurement Error: Recommendations

- Do not have specific quantitative standards regarding statistical precision.
  - Such standards provide a “hard” requirement, while requirements regarding measurement error must be more qualitative – and resources are limited.
  - As a result, quantitative precision standards can lead to inadequate attention to measurement error.
  - We believe this might be aggravated given the sharp resource requirements imposed by the 3% spending rule.
  - Another factor is that it is hard to capture all the variations in situations in a simple set of quantitative standards.
- Instead, focus on the overall process for minimizing uncertainty (including both sampling and measurement error) within the constraints of limited resources.

Sampling and Measurement Error: Some Possible Principles

- Planning for impact evaluation should include systematic consideration of sources of both sampling and measurement error.
- Across programs, limited impact evaluation resources should be allocated in a manner that minimizes overall uncertainty (including both sampling and measurement error) about total portfolio impacts.
- Similarly, across impact evaluation activities within an individual program, resources should be allocated in a manner that minimizes overall uncertainty about total program impacts.
- Efforts to minimize sampling and measurement error should be explicitly balanced.
- Impact evaluation activities should be designed and staged to lead to a systematic, cumulative reduction in uncertainty over time.

5. Principles Governing the Allocation of Resources

- The largest risk to most DSM programs is not in their direct results in terms of gross energy savings. It is in the way free rider results can take away from the tangible results of programs by assigning causation of a portion of savings to non-program forces. For example, in a lighting program, sales may exceed goals and the gross results appear strongly cost effective. But an evaluator can come in and ask people who have purchased bulbs if they would have done so in the absence of the program, or if the program was the reason they bought the energy-efficient alternative. Likely, a good reason analysis would show a complex of remote and proximate reasons for a customer choice. For a program administrator, the evaluators may appear as too easily accepting representations of customers and converting them into facts that take away from program effort and successes. The free rider and net-to-gross methods are the areas of methods that are likely to have the biggest effect on net savings results. So, it is worthwhile to have a strong focus on getting these methods right.

- At the same time, evaluation is more than monitoring for compliance. It should contribute to development of stronger measures, more effective programs, and new technologies and approaches, and is necessary to help us move from "Plan B" DSM (like Energy Star) to "Plan C" DSM (like the "Go Deep" 1000 Homes Project).

- We need to think and talk about the implications of the 3% spending rule in light of the ramp-up in program budgets over time: need to be very strategic in how we allocate resources over time.

5. Principles Governing the Allocation of Resources

- It is important to do Process Evaluation the first year. Illinois is on the learning curve.
- Market characterization is also important (and is outside the 3% budget limitation)
- But impact is most important for this first DSM cycle.

6. Methods for Estimating Net Savings: A 30,000 Foot View

- Methods for the estimation of net savings are a frequent source of controversy.
- The fundamental reason for this controversy is that net savings are inherently challenging to measure:
  - Human behavior and motivation are complex.
  - Net savings focuses on a counter-factual (what would have happened in the absence of the program) which is always challenging.
  - Not a simple yes or no; in the absence of the program, the timing and scope of EE actions might have been different, and the participant might have adopted an intermediate level of efficiency.
  - Multiple parties may have influenced the EE adoption.
- The issue of what to do about these challenges is value laden, and thus opinions differ widely.
- Assumption here is that we will be estimating net savings, and need to establish broad principles regarding the selection and implementation of individual methods.

Methods for Estimating Net Savings: Background

- There are 3 main classes of approaches:
  - Self-reporting (ask participants what they would have done without the program)
  - Econometric (model differences in the observed behavior of participants and non-participants)
  - Market-based (compare overall levels of adoption in the program area and a comparison area)
- Each class encompasses a wide range of individual approaches
- Methods for an individual program may include multiple approaches, sometimes even drawn from multiple classes.
- Selecting the right approach or combination of approaches is a crucial first step, and may be at least half the battle.

Methods for Estimating Net Savings: What Issues Should the Framework Address?

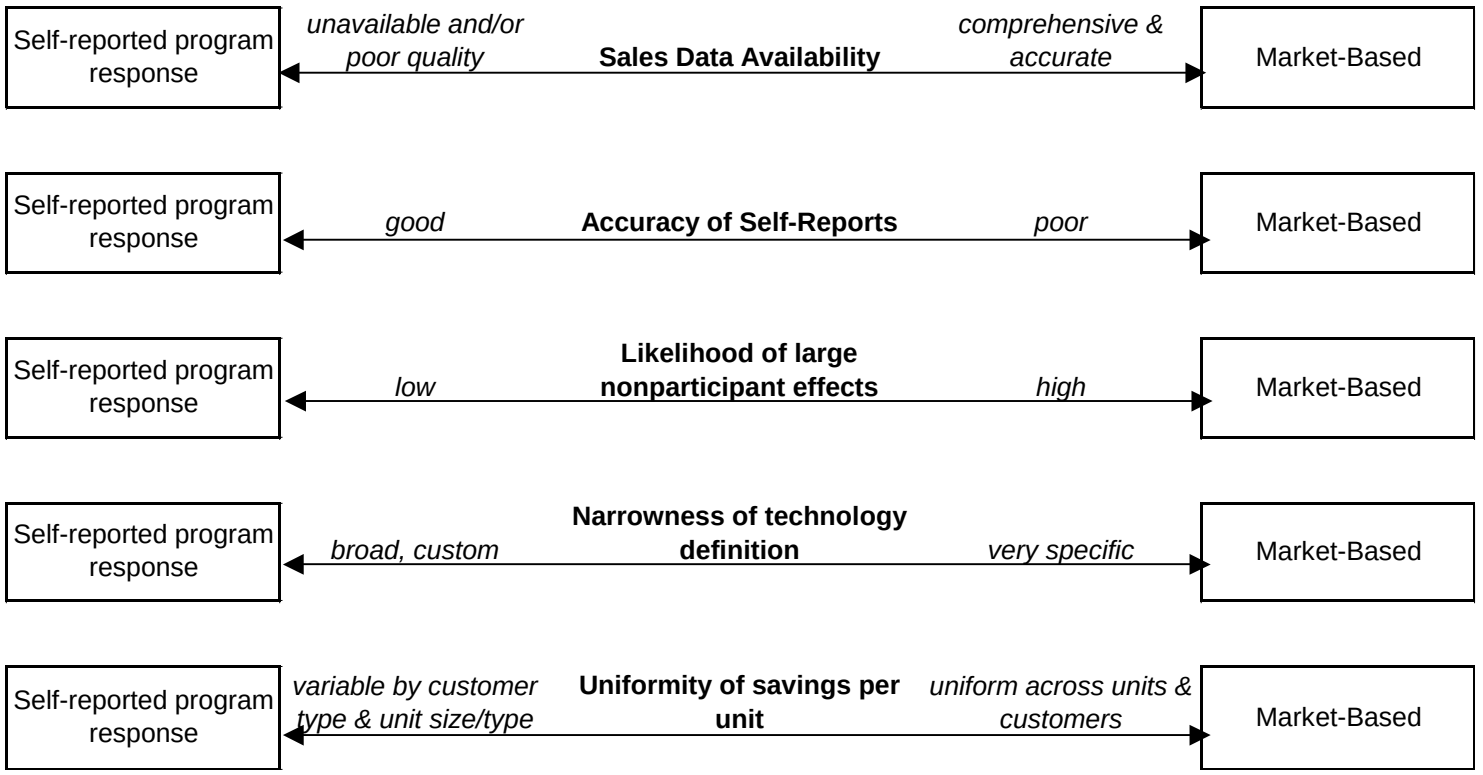
- May be useful to break the methods development process into three steps:
  1. Which class(es) of approach to use for each program.
  2. Which individual approach(es) to use for each program.
  3. How to implement the individual approach(es).
- In developing framework, we suggest focusing for now on establishing principles regarding Step 1.
- While Steps 2 and 3 are very important, in other states it has taken a good deal of time to agree on overarching principles regarding them.
- Therefore, in order to avoid holding up the evaluation planning process, for now we recommend a pragmatic approach: simply delegate the SAG evaluation consultants to work collaboratively with the evaluation contractors on Steps 2 and 3 at a program- and utility-specific level.

Principles Regarding Which Class(es) of Approach to Use: An Example from Wisconsin

- Over the years, Wisconsin has developed detailed guidelines for the selection and implementation of net savings approaches.
- Wisconsin seldom uses econometric methods, leaving self-reporting and market-based methods as the primary classes.
- A systematic approach for choosing between these two classes is a fundamental part of Wisconsin's net savings framework.
- Next slide provides some details on this methods selection framework.

Wisconsin High-Level NTG Method Selection Screening Criteria

(Credit: Wisconsin Focus on Energy Evaluation Team)



Methods for Estimating Net Savings: Additional Recommendations

- Balance investment in the estimation of net-to-gross ratios with investment in the estimation of gross savings parameters.
- Invest the most in estimation of net savings in cases where the NTGR is the most uncertain.
- In cases where the NTGR is likely to be uncertain and the savings are substantial, consider using multiple methods.
- Don’t over-do it. Keep in mind that extreme accuracy is typically neither feasible nor necessary.
- When it comes to uncertainty, worry the most about measurement error that operates consistently in the same direction across programs. At the portfolio level, most other uncertainties will tend to come out in the wash.
- Anticipate that NTGRs will evolve over time as the program matures.
- Plan on multiple rounds of NTGR analysis, both to provide early feedback to be used in improving program design, and to capture changes in NTGRs.
- To the extent self-reporting is used, develop standardized instruments at the statewide level to ensure consistency and comparability.

Illinois Evaluation Framework: Recap of Key Recommendations

- Regarding the application of impact evaluation results:
  - Use a primarily prospective system.
  - Consider requiring retrospective *verification.*
  - Develop a binding schedule for impact evaluation activities.
  - View the most important functions of net savings evaluation as being to incentivize administrators, support resource allocation and inform program redesign.
  - Apply net savings results in the same, primarily prospective manner as other savings parameters. Use the same approach for all applications of net savings.
- Regarding handling of uncertainty:
  - In developing any requirements for the minimization of uncertainty, give equal attention to measurement error and sampling error.
  - Do not have specific quantitative standards regarding statistical precision.
  - Design and stage impact evaluation activities so as to lead to a systematic, cumulative reduction in uncertainty over time.
  - In allocating impact evaluation resources, focus on minimizing overall uncertainty, taking into account both measurement and sampling error.

Illinois Evaluation Framework: Recap of Key Recommendations, Cont.

- Regarding guidelines for net savings estimation:
  - Begin by establishing guidelines for determining which broad class(es) of approaches should be used for individual programs.
  - Invest the most in net savings estimation in cases where the NTGR is most uncertain; when there is great uncertainty and large savings are at stake, consider using multiple methods.
  - Plan on multiple rounds of net savings estimation, both to provide early feedback and to capture changes in NTGR.
  - Don’t overdo it!