Key Evaluation Framework Issues:
Recommendations from SAG Evaluation Consultants

1. Application of evaluation results: retrospective vs. prospective approaches

   - Use a primarily prospective impact evaluation system.
   - 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
   - Consider requiring retrospective *verification.*
   - Develop a binding schedule for impact evaluation activities.

2. Application of net savings results

   - 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 as for all other parameters.
   - Use the same approach for all applications of net savings results (e.g., assessing goal attainment, redesigning programs, B/C analysis)

3. Approaches to deeming of savings parameters

   - Originally we did not say "deemed values." We simply said that we began with technical, economic, and practical potential studies that made use of what we called "engineering estimates," placeholders until impact results were available. In the 1980s-1990s, there was a strong emphasis on DSM impact evaluation, just as today. But the image of evaluation at that time was much more on "hard measurement" than it is today. It took some years for program effects to become better understood and for evaluators and regulators to become comfortable with gross savings results. The focus was on measurement, direct metering, and - where necessary - simulation or model based statistical methods.
   - 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.

- Added to this is a very workable shortcut. If a state with similar measures that are not affected by weather has done several DSM cycles, their deemed values have incorporated several layers of review over time plus corrections for sequential evaluation studies. The California DEER database is a good place to look to leverage years of effort and millions of dollars of work to arrive at reasonable deemed values.
- If the measures are affected by weather and building type then results from states with similar weather can be used and weather zones and building types can be worked in to simulation models for more exact estimates.
- 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.

4. Sampling and measurement error

- Do not have specific quantitative standards regarding statistical precision.
- 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 allocation of resources

- One general principle is to focus more resources in the areas that seem to have the greatest effects in making results uncertain. Sometimes this is called taking the "signal to noise" ratio into account. It is useful to focus where noise is strong and signal is weak.
- 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).

6. Methods for estimating net savings

- In developing any framework rules regarding estimation of net savings, focus for now on establishing principles regarding which broad *classes* of methods (e.g., self-reporting, econometric, market-based) to use for which kinds of programs and situations.
- Delegate SAG consultants to work collaboratively with evaluation contractors on specific methods for individual programs, and how to implement them.
- 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.