Program Evaluation Overview

David Brightwell, Ph.D.
Illinois Commerce Commission
dbrightw@icc.illinois.gov
June 6,2017

Goal of EM&V

- EM&V = Evaluation, Measurement, & Verification

- To examine & improve customer experiences (Process Evaluation)
- To measure savings (impact of the program)
- To verify that measures are actually installed (and working properly)

This discussion focuses on Impact Evaluation.

Impact Evaluation

- Goal: Find out how much energy is saved because of the program?
- Savings: The difference between what is used and what **would have been used if the program did not exist**?
- How much energy would have been used if the program did not exist?

Impact Evaluation, cont’d.

- Ultimate objective is to try to simulate two worlds. In one world, the program exists. In the other world, the program does not exist. Everything else between the worlds (up to the point of program inception) is identical.

- Under the above circumstances, all differences between the worlds after program inception can be seen as the result of the programs.

Impact Evaluation, cont'd.

- Creating an alternative world is obviously not possible.

- What can be done to approximate an alternative world?
  - Randomized Controlled Trials ("RCT") – "gold standard." Used in drug trials, web design, marketing, etc.
    - Quite often not practical? Requires random assignment.
  - Quasi-Experimental Design ("QED") – not as good as RCT but is still "scientific" in nature.
  - Engineering Simulations ("ES"): Look at individual measures and estimate energy use based on design specs and lab testing.

Impact Evaluations, cont’d

- Personal Preference is RCT or QED:
  - ES methods tell measure savings under optimal conditions.
    - Does not examine any behavioral changes of program participants.
    - Does not consider what did not happen because of program participation.

Impact Evaluation, cont'd

- Engineering Simulations need to be combined with Net to Gross estimates.
  - NTG adjusts engineering estimates to account for whether or not participants would have done the same thing without program intervention.
  - NTG relies on surveys.
    - Surveys are expensive to administer
    - Surveys rely on responses of participants
    - Response rates are often low. (If response is not random, survey results are biased).

Impact Evaluations, cont’d

- RCT and QED
    - Both use a control group
        - The Control Group is people who are similar to participants but who did not participate (essentially the control group serves to approximate the alternate world).

Impact Evaluations, cont'd:

Comparing energy use of only customers who participated in the program does not provide a reliable estimate of savings.

- Weather differences
- Economic differences
- Political differences
- Numerous other factors that change over time but affect energy use.

Quasi-Experimental Design

- The more data related to energy use that is available, the better the matching.
- Ideal:
  - historical energy use
  - # of occupants
  - Income of Family
  - Education Level
  - Square Footage
  - Age of Residence
  - Work Status
  - Location of Residence
  - Building type

Quasi Experimental Design Issues, cont’d.

Data usually available:

- Historical energy use
- geographical location (zip code or census tract)

Quasi Experimental Design Issues, cont'd.

Potential Issue with creating matches.

- If non-income-qualified persons are different (with respect to energy use) than income qualified persons, estimated savings may be incorrect.
  - Live in better insulated buildings
  - Less Downward impacts from downturns in economy
  - More Positive impacts from improved economy
  - Different square footage
  - Different # of residents in household
- Are most zip codes with IQ persons homogeneous (with respect to building ages, family sizes, etc.) (for programs that serve a census tract not as big an issue)

Quasi Experimental Design Issues, cont'd.

- Regression Discontinuity method can address some (but not all) issues
  - Programs have eligibility cutoffs (either income or priority queues due to fund limits). Compare change in behavior of those who "just barely" qualified to those who "just barely" did not qualify.

- How do you measure savings for those who participated but are far from the cutoff criteria?

Conclusion

- Impact evaluation is imperfect.
- RCT's are the best method but not always practical.
- QED is similar to RCT. The more relevant information that is available, the better the results.
- Engineering Simulations are an option.