

Title: Evaluating Whole-Building Programs: It is harder and easier than you think!

This paper addresses two trends in our industry: 1. More programs are being introduced that seek deep savings, specifically whole-building or whole-house programs, and 2. Many in the industry are looking for more standardization of methods of evaluating the impacts of programs. An example of the latter is the Uniform Methods Project (UMP), among others. We find that there are design issues specific to whole-building programs that merit extensive discussion, and that the UMP stops short of what is needed. We also find that many evaluators are confused about the issues and definitions of key terms. Our paper focuses on comparison groups and the unique advantages and disadvantages of finding appropriate comparison groups for the whole-building type of program, most of which are not mentioned in the UMP. Evaluators commonly choose the example of a single-measure program, such as an air conditioner rebate program, to think through what kinds of comparison groups are necessary to estimate program impacts. But the issues surrounding comparison groups are different for programs that have a lot of potential measure types and where multiple, expensive measures are required to qualify for the program.

Some of the issues we have identified in the UMP's chapter on whole-building programs include: 1. Insufficient consideration of the complexities in using future participants as a comparison group for the evaluated program cycle. For example, might future participants have installed some qualifying measures in the year prior to their participation, and would that lead to a partial estimate of net effects? 2. Insufficient consideration of both the advantages and disadvantages of choosing a comparison group in a program that requires multiple, expensive measures to qualify, and where there is a long list of qualifying measures. For example, is a customer who installed insulation, did air sealing, and replaced an air conditioner a good comparison customer for a participant who put on a cool roof and installed efficient windows? 3. Omission of some possibly legitimate designs for producing net effects, 4. A lack of clarity on what constitutes the "eligible population" for a whole-building program, and 5. Lack of guidance on the importance of checking the assumptions behind the use of any comparison group. This paper makes those critiques, but bases them on a substantial explication of the relevant design issues, especially those that are unique to this type of program.

Ours is one of two papers being written to address these issues, and submitting abstracts for the 2017 IEPEC (the other is titled, "Compared to What? Practical Tools for consumption Data Analysis Mitigating Self-Selection Bias.") The other paper focuses on modeling program impacts for whole-building programs, while this paper focuses on the comparison groups that should be used in the impact analysis. The two papers are linked in that what type of modeling to use and what variables to include will depend, in part, on what kind of comparison group is in the analysis.

Submitted 11/14/16

Evaluation Area: Impacts

Sector: Multi-Sector

Focus: Methods