Cracking the Code:
Impact Evaluation Methods of ARRA-Funded Energy Code Initiatives

Michael Rovito, ERS, New York City, NY
Tim Pettit, DNV-GL, Washington, D.C.

ABSTRACT

Under the American Recovery and Reinvestment Act (ARRA; the stimulus bill) of 2009, the U.S. Department of Energy allocated funding through the State Energy Program (SEP) that was dedicated to advancing and enhancing code compliance. This paper highlights the methods of the subsequent impact evaluation overseen by Oak Ridge National Laboratory. As one component of the broader SEP evaluation, the authors' team evaluated the savings from energy code advancement and training programs across the country. In addition to facing diverse climates, building stocks, codes, politics, and enforcement regimes, the team's code-specific budget precluded in-field work or custom methodologies.

The team developed a uniform method that was sufficiently adaptable for state-specific conditions. The savings algorithms built on the publically available Pacific Northwest National Laboratory's code savings estimator. Data were primarily collected through interviews with public and private sector experts, who estimated building practices. Quantitative responses were captured through a Delphi process and were used as inputs into the model. Expert estimates were combined with data on building stock growth for major sectors and energy use intensity deltas between codes and state-specific base cases to arrive at savings by state. Attribution was also considered – how essential were SEP efforts to the eventual adoption of the new codes and improvements in compliance?

Taken together, the many aspects of the evaluation provide a reasonable, cost-effective approach for developing and attributing savings from code-based initiatives.

Introduction

Program administrators and regulators are paying greater attention to improving the level and enforcement of building energy codes in their states. As recently as a few years ago, some states had no building energy code at all and many had compliance rates estimated to be less than 50%.[1] The savings potential associated with strengthening codes and improving compliance is not difficult to intuit, but with regulators beginning to allow programs to count those savings towards their energy efficiency goals the question arises of how to precisely quantify and attribute the savings. This paper presents a model for doing just that.

Under the American Recovery and Reinvestment Act (ARRA; the stimulus bill) of 2009, the U.S. Department of Energy allocated funding through the State Energy Program (SEP) that was dedicated to advancing and enhancing code compliance; the funding was also contingent upon the states updating their energy code to the most recent national model code. This paper highlights the methods of the subsequent impact evaluation overseen by Oak Ridge National Laboratory. As one component of the broader SEP evaluation, the authors' team evaluated the savings from energy code advancement and training programs across the country. In addition to facing diverse climates, building stocks, codes, politics, and enforcement regimes, the team was limited by the allocation of evaluation funds, which precluded in-field work or custom methodologies.

[1] http://www.imt.org/uploads/resources/files/IMT_Report_Code_Compliance_Savings_Potential.pdf

The team developed a uniform method that was sufficiently adaptable for state-specific conditions. The savings algorithms built on the publically available Pacific Northwest National Laboratory's code savings estimator. Data were primarily collected through interviews with public and private sector experts, who estimated building practices. Quantitative responses were captured through a Delphi process[2] and were used as inputs into the model. Expert estimates were combined with data on building stock growth for major sectors and energy use intensity deltas between codes and state-specific base cases to arrive at savings by state. The challenging topic of attribution was also considered – how essential were SEP efforts to the eventual adoption of the new codes and improvements in compliance when compared to other non-ARRA code adoption influences and evolving building practices such as increasingly "green" consumer sentiment?

Taken together, the many aspects of the evaluation provide a reasonable, cost-effective approach for developing and attributing savings from code-based initiatives.

This code evaluation was the broadest of its kind to date, was deployed across a range of political and climactic environments, and was performed without in-field work. The evaluation built on known code evaluation models, protocols, and tools and was able to provide energy impact estimates while reducing the costs more typically associated with code impact evaluations. As such, the evaluation approach can serve as a cost-efficient model for those states and utilities hoping to claim savings from code acceleration or training when the evaluation resources do not allow for more rigorous approaches.

Background on Activities Being Evaluated

Under ARRA, the U.S. DOE allocated funding through the SEP that was dedicated to advancing and enhancing code compliance. The funding was also contingent upon the states updating their energy code to the most recent national model code. Prior to ARRA, some states had no building energy code at all and many had compliance rates estimated to be less than 50%.[3] Consequently, there is significant energy and cost savings potential in advancing the code (i.e., adopting a code where none has been adopted or adopting a more advanced code where one already has been adopted) and of enhancing compliance with the code (i.e., increasing the rate at which builders follow code requirements in construction).

In practical terms, SEP funding went to a variety of activities, which differed by state and which were chosen by the SEP representatives from each state.

For code advancement, the SEP funding provided the resources necessary to adopt a new code. This included:

- Provision of technical resources in an advisory capacity to address questions raised by the decision making entity responsible for adopting a code whether that be a legislative body or an appointed committee
- Development of reports and studies that examine the cost-benefit impact of adopting a code or of adopting a new and more advanced code
- Hiring of experts and analysts to identify necessary customizations to model codes in order to make a code appropriate to a state's climate and economic conditions
- Outreach to community-level code bodies in order to demonstrate the benefits of code adoption in action and to create grass roots support for code adoption at the state level

For enhancement of code compliance, the SEP funding provided resources to improve voluntary compliance and to improve enforcement. This included:

[2] A Delphi process leverages the input of an expert panel through multiple rounds of interviewing. Between each round, the responses are aggregated and reported back to the group anonymously. The experts are then asked to revise their responses in light of the aggregated information. The revisions should drive down variability in responses and move the group towards a consensus. The process is useful for identifying quantitative answers through the subjective assessment of experts in an environment where directly measuring the value in question is prohibitively difficult.

[3] http://www.imt.org/uploads/resources/files/IMT_Report_Code_Compliance_Savings_Potential.pdf

- Commissioning "best practices" studies of building code enforcement practices at the county level
- Provision of code manuals for inspectors and others in industry
- Establishment of equipment rental programs for key energy code compliance testing equipment (e.g., blowers for blower door testing)
- General building science training
- Code-specific training for code enforcers
- Code-specific training for the public
- Training on compliance software
- Train-the-trainer programs for representatives at the county level who can serve as evangelists of code compliance strategies
- Online webinars for code compliance and code requirements
- Statewide code conferences to increase exposure to the code and improve knowledge of specific requirements

The range of activities is significant, and the activities vary from state to state. Some states were adopting a code for the first time. Some states were advancing their code to a new level. Some states already had the most advanced code available at time of study, but intended to improve compliance. Each state took a slightly different path.

Evaluation Methodology

The evaluation was tasked with estimating nationwide savings from all ARRA-funded SEP activities. The evaluation was subdivided into distinct evaluation efforts by the nature of the SEP activities; code-related activities were one of those subdivided efforts.

For the energy code impact evaluation, nationwide savings was estimated by taking a sample of seven states that best represented the nationwide efforts being evaluated. For each of those seven sampled states, a statewide estimate of savings was prepared. The statewide estimate represented two separate estimates of savings: one for residential buildings and one for commercial buildings, which are subject to different codes and which exhibit different levels of compliance in practice. For each of these sectors across all seven states the following high-level savings formula was used for gross savings:

Gross savings = (Previous code EUI – Current code EUI) × (Post-SEP compliance – Baseline compliance) × Construction activity

The following subsections walk through the components of this approach, which mirror some of the principles laid out by the Cadmus group in their report on code savings estimation authored for NEEP and IMT.[4] In this particular case, features of the work being evaluated – notably the impact of code acceleration –require that a few key points and assumptions be highlighted in order to understand this formula and how it was used.

First, all those who are failing to comply with the current code are assumed to be reverting to the *previously adopted* version of the code. For example, if a state jumped from IECC 2003 to IECC 2009, the previous code EUI is based off of IECC 2003. That is, wherever (in whole or in part) someone fails to comply with the current code, they are assumed at minimum to be fully complying with the previously adopted code.[5]

[4] The Cadmus Group et al. Attributing Building Energy Code Savings to Energy Efficiency Programs. February 2013.

[5] Of course, those who are not complying with the current code may not be complying with the previous code either, thus the approach may be underestimating savings. However, in practice, those who are coming into compliance with a new code or are improving their compliance, as a result of SEP activities, are those likely to have been following codes in the past.

Second, the compliance definition being used is "Weighted-Energy Compliance" or WEC. WEC aggregates requirement-specific compliance levels (e.g., compliance with the wall insulation provisions) using energy savings impact weights. That is, if a building complies with nine out of ten requirements, but the tenth requirement is responsible for half the savings expected to occur as a result of the code, then that building is considered to be only 50% compliant. WEC best represents the energy impacts of compliance, which suited the purposes of this evaluation well.

Third, the impact evaluation was required to report savings over a period of years (through 2040). For code activities, the impact of accelerated code adoption or improved compliance levels lingers for years, affecting year after year of construction. Thus, a different gross savings level was calculated for each year. In order to support this, code compliance was treated dynamically in both the Post-SEP case and in the baseline case (i.e., the counterfactual case). Constructing these estimations of compliance over time required forecasting compliance levels and estimating natural levels of improvement (i.e., those unrelated to SEP efforts).

Finally, where possible, hard data was sourced as an input to the above formula. For example, construction activity is documented on an historical basis and is projected by reputable entities on a forward-looking basis. Compliance, however, needed to be estimated through research performed by this study. The primary mechanism for gathering this data were interviews. The team interviewed experts in each state on state-specific conditions as part of a Delphi process. The following subsections walk through each element of the analysis in greater detail, including the interview approach.

Sampling

The overall SEP evaluation is based on a complex sample design. The data were aggregated to pre-defined program categories, of which Codes and Standards is one, using sample weights created from a multi-phased weighting process. The study began with a program definition stage, wherein program tracking data (from the PAGE information system) was acquired and managed for initial definition of the population of all programs in the evaluation periods. Collected program data included funding amounts, program administrator contact information, program milestone accomplishment tracking, and comments submitted to the system by state administrators. Using information gathered from PAGE, the contractor team identified all Codes and Standards programs, and selected a random sample to evaluate.

The evaluation team then entered the State Energy Office (SEO) data collection phase, wherein data was collected from program administrators. In this phase, the team determined which Codes and Standards programs were evaluable. During this phase of the study, the contractor team collected program-specific data from funding recipients and other program stakeholders for use in calculation of evaluated outcomes, as described below.

Interviews

The team interviewed a range of stakeholders of the energy code process in each state as part of a multi-stage Delphi process. These stakeholders included SEP program administrators, subcontractors employed by the program administrators, local building code officials, residential builders, commercial builders, architects, engineers, other construction contractors, trade groups, and building and efficiency organizations. These interviews gathered information on three key aspects of the study:

- **Code status and evolution before, during, and after the years of SEP code activities.** These questions helped us understand the state of the energy code prior to and after the SEP activities to discern if they were successful in moving the code forward. The questions also probed the natural evolution of code adoption in the state: had they been adopting every code? Every other code? Understanding the momentum in the political system helps to determine the extent to

Consequently, the overall underestimation of savings is likely to be small.

which the SEP activities truly accelerated code adoption. Finally, the interviewees shared information on state-specific code amendments.

- **Building practices (as pertaining to code compliance) for the years before and after SEP code activities.** Building practice questions were framed at the measure-level (e.g., wall insulation). Interviewees were asked what percentage of construction complied with a given requirement. The data collection instrument was capable of capturing these answers at the sector level (i.e., single family, multifamily, office, institutional, and retail), although not all interviewees could speak to this level of detail; at minimum, residential and commercial breakouts were required. The interviews probed for qualitative supporting information (e.g., differences between urban and rural construction, common examples of partial compliance, etc.) that helped the interviewee arrive at a quantifiable estimate as well as provided feedback for other interviewees as part of the Delphi process.
- **The impact of other non-SEP-related factors on building practices before, during, and after the years of SEP code activities.** The changes in building practices noted in the interviews could have been caused by factors other than SEP code activities. In order to appropriately understand the attribution of changes, it is important to understand how those other factors impacted building practice improvements. Examples of such other factors include utility programs, "green" consumer sentiment, non-SEP code trainings, natural code adoption cycles, etc. Interviewees were asked what other factors there were and how important those factors were to the changes observed in comparison to the SEP code activities.

Recruiting interviewees who were qualified to speak on these topics at length was a challenge. Consequently, interviews were limited to about ten per state, spanning both private and public sector roles. The team developed a credibility weighting system in order to weight their responses against their peers. It was a subjective system with scores of 1, 2, or 3 provided. Scores were based on: credentials and experience; internal consistency (do their answers make coherent sense with themselves?); and external consistency (do their answers make sense in light of what other experts are saying?).

After being interviewed, all interviewees were afforded a chance to review the responses of others (anonymized) and a chance to revise their own responses in light of what others said. The information provided back to interviewees were not only the quantitative estimates of compliance with each measure, but also the qualitative reasons, which may prompt updated thinking in the eyes of interviewees. This Delphi process was intended to bring about some consensus viewpoints. In the end, measure-specific compliance responses were fairly consistent across interviewees, particularly the relative compliance levels (differences across measures) and the change in compliance from one period to the next (differences pre- and post-SEP code activities).

Estimating Energy Consumption and Savings at Different Code Levels

The Pacific Northwest National Laboratory has developed a tool that estimates energy code savings on a statewide basis called the Utility Savings Estimator (hereafter, the "PNNL tool").[6] The PNNL tool was too rigidly geared towards estimating savings on a prospective basis (e.g., if we change the code this way, how much will we save?) to be used for this study's purposes, but it had many components that were of use for the study.

Notably, the tool provides a modeled EUI for each code version (e.g., IECC 2009) for each state. These EUI estimates are climate-specific and are broken into components. For residential, the components are electric and fuel heating, electric and fuel cooling, and electric and fuel other (6 total) and are provided on a per-housing-unit basis. For commercial, the components are electric and fuel HVAC and electric and fuel other (4 total) and are provided on a per-square foot basis. Through expert interviews and through a

[6] Tool produced by Olga Livingston, Rosemarie Bartlett, Doug Elliott of Pacific Northwest National Laboratory.

review of historical documentation, ERS determined the baseline code (e.g., IECC 2003) and the code that was adopted as a result of the program.

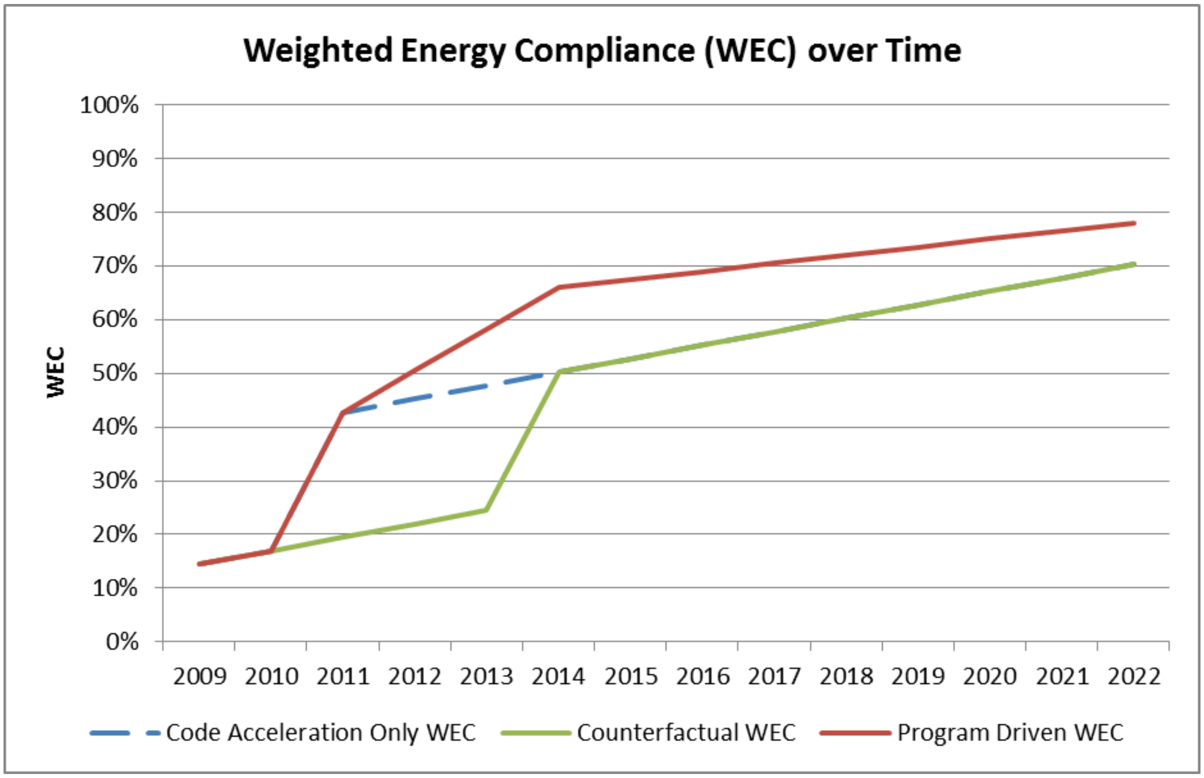
The PNNL EUIs do not reflect adjustments for amendments. As such, ERS determined through expert interviews and through a review of historical documentation which amendments had been attached to the model codes. ERS then used engineering judgment to estimate the impact that the amendments would have. For the most part, amendments were relatively modest in scope, which minimized the impact of any adjustments.

These modeled and, in some cases, lightly adjusted EUIs enabled the team to estimate a rough energy consumption level for each code on either a per-unit or per-square footage basis. The difference between the baseline and post-SEP codes was used as the savings potential of moving from one code to another. This represented an upper bound on the savings for the analysis. As a reminder, the baseline code represents the level of energy performance expected if someone does not comply at all with the practices dictated by the new, SEP-promoted code. Hence, if the state moved from 0% compliance with the new code to 100% compliance, the state would be able to claim gross savings equivalent to the difference in the PNNL EUIs. The next section articulates the practice of estimating the true improvement in building practices.

Building Practice Deltas

In the team's savings model, code adoption and compliance are treated dynamically – that is, they change over time. In order to estimate savings, a real-life scenario (i.e. what happened because of SEP; the "post-SEP" case) is constructed and is compared to a counterfactual (i.e., as if the SEP code activities did not happen; the "baseline" case). The counterfactual scenarios must address when a code would have been adopted (for those states that did accelerate code adoption via SEP code activities) as well as the trajectory of building practices under a regime that did not benefit from SEP code activities (e.g., code acceleration, training, equipment rental, etc.). The post-SEP case must simply represent the change in code and the trajectory of building practices as they changed in real life over a period of years while also presenting a credible forecast for how those building practices will continue to improve over time in the future.

The team determined the year of code adoption in reality by observing the historical record and verifying it with interviewees. To determine the counterfactual, baseline year of adoption (i.e., the year the code would have been adopted in the absence of the program), interviewees were asked to describe the political climate in the state and estimate explicitly the year that the code would have been adopted without the program. For example, one state was on an "every other cycle" adoption trajectory for codes and typically adopted on a two-year delay, which suggests that a new code would have been adopted in 2014, as compared to reality (i.e., the post-SEP case) where it was adopted in 2010. Thus, for the counterfactual, baseline case, the entire trajectory of building practices is shifted by four years. The following figure illustrates the concept of post-SEP and counterfactual, baseline building practice (i.e., compliance) trajectories.



Compliance values – the percentage of construction that adhered to a specific requirement in their standard building practices – were estimated using interview data. Interviewees were asked to estimate the percentage of projects in the state that adhered to the adopted code *prior to it being adopted*[7] and again how many were adhering to it in current practices. This was done on the measure level for a set of prescriptive practices outlined in the code. The former question established a baseline level of compliance: the naturally occurring market adoption (NOMAD) of technologies and measures in the code. The latter question showed how compliance improved following the adoption of the code and the provision of program training and other SEP activities.

A single weighted-energy compliance (WEC) value was estimated by taking the measure-specific compliance estimates and weighting them by the relative importance of each measure. The team used a modified weighting system based on PNNL's Score + Store compliance tracking system.

The pre- and post-WEC values established two anchor points by which to determine the trajectory of compliance over time in the program-induced scenario and the counterfactual, baseline scenario.

- For the program-induced scenario (reality), the pre-WEC was assumed to have been achieved simultaneous to the code adoption year. The post-WEC was assumed to align to the period of questioning (late 2013) and is thus reflected in the 2014 savings analysis. Compliance prior to the pre-WEC anchor is estimated to reduce linearly by 2.5% per year as you go back in time; this represents the learning rate for NOMAD of a less-informed populace per the PNNL tool's assumption (i.e., the natural rate at which a population will learn and adopt the code provision on their own, described as the incremental percentage of the overall population, which learns the code each year). Compliance between the anchor points is assumed as follows: half of the difference between the pre- and post-WEC is achieved immediately as a consequence of a change in code requirements and the remaining half is achieved linearly over the period of years between the two anchors as a consequence of training and learning. Compliance following the post-WEC is estimated

[7] For consistency of language, we call this compliance all though it does not technically qualify as compliance because the code was not yet in place.

to grow at 1.5% per year as you go forward in time; this represents the learning rate of a more-informed populace per the PNNL tool's assumption (fewer ignorant people leaves less opportunity for learning).

- For the counterfactual, baseline scenario, the pre-WEC is assumed in the same year as above. Compliance in the years before the counterfactual year of code adoption is determined using the 2.5% NOMAD growth rate and the pre-WEC anchor. In the counterfactual year of adoption, the compliance level is assumed to jump an amount equal to that in the program-induced scenario (i.e., nominally half of the difference between pre- and post-WEC values). In the years following adoption, the compliance rate is assumed to grow at a rate of 2.5% (since fewer people understand the code than in the program-induced scenario) until it reaches the level of compliance observed in the program-induced scenario, at which point it matches the program-induced growth rate (i.e., 1.5%). This allows the two lines to converge over time, reflecting the team's belief that the program's effect on code outcomes will diminish over time.
- An additional compliance trajectory is also required to estimate the split in savings between accelerated code adoption and the training component of the program (for the purposes of attribution). This compliance trajectory reflects the estimated WEC for a scenario including accelerated code adoption, but excluding training. It follows the program-induced trajectory through the year of code adoption, but following that year it simply grows at the 2.5% learning rate until it catches up to the program-induced scenario.

These compliance assumptions are best observed graphically in the figure above.

As previously mentioned, all non-compliance (i.e., 100% - WEC) is assumed to imply consumption at the old-code EUI level. All compliance is assumed to imply consumption at the new-code EUI level. Thus, savings is the difference in WEC between the program-induced scenario and the counterfactual baseline multiplied by the difference in EUI between those two codes multiplied again by construction volumes (i.e., the original formula above).

Scaling Energy Savings by Construction Activity

Construction volumes were a combination of census data for historic volumes and PNNL projections. For residential, estimates were in units constructed per year and for commercial it was square footage per year. Simply, these volumes are multiplied by the per-unit or per-square foot estimates of savings (net of compliance differences) as described above. These values represent the gross savings values.

Annualizing Savings

Savings is calculated yearly, thus implying a stream of savings per year. Moreover, the savings of a single building are assumed to last for 20 years, reflecting the fact that these measures impact core building systems (mostly envelope) and will have long lifetimes. Hence, there are new first-year savings being generated many years into the future, for as long as the impacts of the program are estimated to exist. In addition, though, there are lifecycle savings layered upon one another as the building stock that is influenced by this SEP activity is maintained out into the future.

Attribution

Attribution, or net-to-gross (NTG), was addressed in a variety of ways. Note that it must inherently be subdivided between code acceleration activities and code compliance enhancement activities. These are addressed below.

First, one portion of code acceleration NTG is handled inherently in the gross estimation methodology. Asking interviewees to estimate the counterfactual year of adoption is a NTG question. It is applied in the gross section out of necessity, but it gets at NTG analysis aspects.

An additional layer of NTG factors were applied based on four interview questions that explicitly addressed the other influences that might have led to the observed changes in compliance or code.

One question specifically addressed code acceleration. It asked interviewees to rate, on a scale of 1 to 10, the influence that the program had on the code adoption process. A qualitative component of the question probed why the interviewees responded the way they did. Answers were weighted by the same credibility weightings described in the preceding section and then averaged. The average value was then normalized to 7.5 (i.e., 7.5 and above were considered 100% NTG). This was done because there was an observed underestimation of program influence as judged by the qualitative answers. Interviewees scores often were lower than 10, but simultaneously came with qualitative answers that did not identify alternative influences and/or praised the program as an absolutely essential component of the code acceleration process. Having normalized to a 7.5 scale, the percentage NTG was then applied to the portion of gross savings associated with code acceleration, as shown in the "Net" tab of the analysis spreadsheet.

An additional form of NTG analysis was applied for known jurisdictions that were expected to have adopted the code anyway in absence of the program due to historic records of advanced code adoption. For those jurisdictions, code acceleration savings were removed proportional to their population as a percentage of the state's population. This method is applied for those years during which the jurisdiction is expected to have adopted the code.

Three questions, phrased in different ways, addressed the influence of trainings on the changes in compliance. One question asked how likely, on a scale of 1-10, code officials or builders would have been to improve their skills absent the training. The other two asked about other influences on compliance improvements, asking interviewees to identify other sources of influence (e.g., rising energy costs) and then rate their influence relative to the trainings. ERS translated the latter questions into a 1 to 10 scale by assigning 10 to emphatic declarations that the program was the sole influence, 7.5 to those answers saying it had more influence, 5 to those answers saying alternate factors had equal influence, 2.5 to those saying the program had less influence than alternate factors, and 0 to those saying the program had no influence at all. Answers were weighted by the same credibility weightings described in the preceding section and then averaged. The average value was then normalized to 7.5 (i.e., 7.5 and above were considered 100% NTG). This was done because there was an observed underestimation of program influence as judged by the qualitative answers. Interviewees scores often were lower than 10, but simultaneously came with qualitative answers that did not identify alternative influences and/or praised the program as an absolutely essential component of the code acceleration process. Having normalized to a 7.5 scale, the percentage NTG was then applied to the gross savings from code training only

The NTG ratios were applied to their respective savings streams. Subsequently, code acceleration and training savings are then recombined to form the full net savings stream.

One final layer of NTG adjustment that is applied is the elimination of first-year savings for buildings constructed more than 10 years after the completion of the program in 2012. It is team's judgment and assumption that program influence after 10 years is too speculative to be determined definitively. Notably, this judgment applies to first-year savings only; savings from buildings constructed in, for example, 2020 will continue on for 20 years, per the described persistence estimation. However, buildings constructed after 2022 are considered outside the window of program influence. This is an essentially arbitrary judgment (e.g., it could have been 5 or 15 years instead), but is based on the idea that code practices are inherently political and thus subject to the whims of the political process. It is exceedingly likely that between now and then some political event – either supporting or impeding code adoption or compliance – will occur at either the state or federal level that will override any influence from this program. Consequently, first year savings streams all end at 2022.