Illinois EE Stakeholder Advisory Group
Market Transformation Savings Working Group
Small Group Meeting

Wednesday, October 26, 2022
9:00 – 10:30 am
Teleconference

Attendees and Meeting Notes

Meeting Materials

- SAG Market Transformation Savings Working Group Webpage
- Posted on the October 26 meeting page:
  - October 26, 2022 MT Small Group Agenda
  - SAG Facilitator Presentation: Introduction to Market Transformation Small Group
  - Northwest Energy Efficiency Alliance Presentation: Assessing Influence of Market Transformation Programs Through Theory Based Evaluation
  - Apex Analytics Presentation: Theory-based Evaluation for Market Transformation

Attendees (by webinar)

Celia Johnson, SAG Facilitator
Greg Ehrendreich, Midwest Energy Efficiency Alliance (MEEA) – Meeting Support
Allen Dusault, Franklin Energy
Chris Neme, Energy Futures Group, representing NRDC
Dan Violette, Apex Analytics
Hannah Collins, Leidos
Jane Colby. Apex Analytics
Jason Christensen, Cadmus Group
Jayden Wilson, Opinion Dynamics
Jeff Erickson, Guidehouse
Jeff Harris, Northwest Energy Efficiency Alliance (NEEA)
Jim Fay, ComEd
John Lavallee, Leidos
Kegan Daugherty, Resource Innovations
Michael Frischmann, Ecometric Consulting
Nicholas Crowder, Ameren Illinois
Patricia Plympton, Guidehouse
Rachel Marty, Guidehouse
Randy Opdyke, Nicor Gas
Rick Tonielli, ComEd
Rita Siong, Resource Innovations
Ryan Wall, Guidehouse
Shannon Kahl, ILLUME
Stu Slote, Guidehouse
Ted Weaver, First Tracks Consulting, representing Nicor Gas
Thomas Manjarres, Peoples Gas & North Shore Gas
Vincent Gutierrez, ComEd
Wayne Leonard, Guidehouse

Meeting Notes

Opening and Introductions

Celia Johnson, SAG Facilitator

Purpose of Meeting: To introduce topics for discussion by the Market Transformation Small Group.

Introduction to Market Transformation Small Group

Celia Johnson, SAG Facilitator

- Background and Purpose of MT Small Group:
    - In 2022, participants in SAG MT Savings Working Group approved edits to TRM Attachment C (Framework for Market Transformation Savings in Illinois).
    - The final meeting to discuss Attachment C edits was in August 2022. ComEd suggested an "MT Small Group" meet in the TRM off-season to address a few open MT questions.
    - Purpose of MT Small Group is to resolve open MT questions before the next IL-TRM update process begins in spring 2023.
    - Participation is open to all, unless a topic may present a financial conflict of interest, or when consensus resolution is needed.
    - Goal is to reach agreement by end of March 2023. Small Group meetings will be held monthly.
- MT Small Group Topics
    - 1. Logic Model Definition
        - Define "preponderance of evidence" terminology, what it means and how it will be applied for MT initiatives; discuss theory-based evaluation
    - 2. MT Energy Savings Framework
        - Discussing re-opening criteria for the Framework, and what happens to savings when framework is revised – credit for future savings, banking, etc.
    - 3. Discuss MT Policy Questions
        - When MT framework was developed in 2019 and added to TRM Version 9.0, there were remaining policy questions to address. The SAG MT Savings Working Group reached "final draft" resolution of these policy questions in mid-2020. However, participants were not ready to finalize the policies since MT initiatives were in very early stages at that time.
        - MT Small Group to discuss:
            - Whether participants are comfortable with current final draft policy resolution, or whether additional edits are needed
            - In what forum does this resolution belong? Options include TRM Attachment C, IL EE Policy Manual, or SAG website

Logic Model Definition

Jim Fay, ComEd; Dan Violette, Apex Analytics; Jeff Harris, NEEA

Introduction by Jim Fay, ComEd

- What burden of proof should be applied to meet "preponderance of evidence" test?
- What is currently in TRM Attachment C: Introduced quotes from the Attachment – methodological differences between program approaches, qualitative attribution

- When we started, we wanted to revise Attachment C to provide more detail to explain what is meant by "preponderance of evidence" – so implementers and evaluators could have certainty.
- ComEd requested several veterans of MT and evaluation of MT programs to present the philosophy of theory-based evaluation and how it applies to MT programs.

Presentation by Jeff Harris, Northwest Energy Efficiency Alliance (NEEA)

- This presentation will provide more questions than answers. From experience of 25 years of MT at NEEA, there are challenges. Title slide uses the word "influence" as opposed to "attribution" – NEEA has landed on influence deliberately. Proof of attribution as noted in Attachment C is a very difficult and high bar and comes with challenges. Influence is a different level of "proof" – easier to identify, but harder to quantify that influence. May sound like semantics but this is where NEEA has landed.
- Big Picture
    - Context is always important. Literature on theory-based evaluation has the same concepts repeated.
    - 1. Traditional EE programs are having a harder time delivering cost-effective net energy savings (e.g., lighting standards). Remaining opportunities are challenging to implement and measure.
    - 2. MT offers an alternative way to achieve results – by focusing on making changes to the marketplace that will be sustained and deliver permanent savings – align EE with market purposes to carry it forward.
    - 3. If MT is a desirable alternative, we need to ask ourselves what is actually needed to support MT in Illinois?
- The amount of effort and rigor put into trying to resolve evidentiary standards should be proportional to the outcomes – if it is desirable for the state, then we may make choices from a policy standpoint that can raise or lower the bar for evidence.
- MT as we practice in the Northwest, and as co-author of Attachment C, made an effort to embed this theory.
    - Trying to change the trajectory of adoption.
    - A new, potentially adoptive service or measure, across the maximum breadth of the market possible for that innovation. Only able to measure what happens in the market after we're done.
    - The counterfactual exists as an alternative question of what would have happened if we do nothing. But once we start, we have changed that trajectory. Discussing how much of the change above the counterfactual is due to the MT program vs other factors.
    - There is a theory that underlies the whole statement around MT being the acceleration of market adoption – there is a diffusion of innovation process undergirding all of it. That diffusion of innovation includes the concept that as you progress on the timeline you work with different groups of adopters (early, middle, late, laggards).
    - This theory is critical to developing the program logic and ultimately the evaluation logic - depending on where you are on the curve at what point in time.
    - We have embraced the concept of logic models as the mechanism to describe what we are trying to accomplish and what program elements will result in those accomplishments. How can we take interim measurements that are intended to validate the program logic? Logic model at start may be different than the middle or end of the diffusion process.

    - If you are looking only at the technology perspective at the left side of the curve – make sure the tech works, is delivering what it promises. Effort and logic model reflect getting product into the market. Past that and in the early market adoption phase, we're thinking more about marketing objectives and communication to a wider audience. As we get to the late majority, we may need more adoption if we change the applicable federal equipment standard, then it becomes a regulatory and policy problem with a different logic model.
    - A good logic model will envision the endpoint, but the interim steps may change over time. As market progresses, the market progress indicators (MPIs) and the evidence for them may need to change as well. The evidentiary standard then depends on where you are on the graph.
- Key Elements of MT Program Design
    - MT logic has to be tailored to the specific EE opportunity – seems obvious, but it turns out to be harder than it looks. E.g., ENERGY STAR Retail Products Platform (ESRPP) targeting retailer behavior relative to a set of innovations, but the actual diffusion of innovation is the individual products. Changes the stocking, marketing, pricing and promotion practices – for example bottom mount refrigerators that qualify under ES emerging technologies award criteria – very specific opportunity. Not just retail for appliances, not just retail for refrigerators, but a very specific configuration and efficiency level.
    - The MT theory for which we are looking for evidence is built off of a market study and baseline level that documents the barriers and opportunities and gets revised as we go up the adoption curve and learn more. The logic model has to document the components – intervention strategies with outputs, and the outcomes in near, mid and long-term. The logic model has a time sequence, and the market progress indicators are the observable component.
    - We believe in our practice that we evaluate indicators on a regular basis (annually). Use those market progress indicators to change the program design to sync to the market. Use a continuous improvement approach.
- TRM Attachment C says all market adoption that is not influenced by ratepayer funded activities should be captured in the natural market baseline. There is an assumption that everything above that natural market baseline is attributable to programs (MT or RA). No further adjustment for free-riders is then needed. That's important. There might be reasons for further adjustment (attribution factors) pursued for other reasons unspecified. For example, building in some conservatism.
- Important to distinguish between market logic, the natural baseline, and the policy drivers that might ask you to change something.
- "Preponderance of Evidence" is primarily a term from legal proceedings – greater than 50% of evidence suggests that the hypothesis is more likely to be true than not. That is a perspective we could use. How we define that…could say 50% of all MPI's pass the test, each MPI could have some quantification; or 50% of the MPIs have to point in the right direction. This could go either way. That's something we need to land on.
    - Consider, what is the Default Position – the inference in Attachment C is that the natural market baseline has absorbed the risk of being wrong. If in fact the natural market baseline can be changed by an attribution factor, that creates some significant uncertainty for everyone because we could do the best job we can with baselines and MPI measurement but leave the door open for ex post modification for "X, Y, Z" policy reasons. What would be the appropriate drivers to add an attribution factor?

- Finally, is utility involvement in MT an important goal for IL? Is CPAS credit the only tool or are there others?

  [Jim Fay] You said the logic model will change over time as a function of the customer mix and the strategies, but you also said the evidentiary standard changes over time?

  [Jeff Harris] As the MPIs change, the evidence that supports those MPIs may change – the evidentiary standard doesn't necessarily change, but at different times along the curve we may have different challenges with measurement. The PoE standard may still apply but the data may be harder to be confident with.

  [Chris Neme] The way I was thinking about this, when you started talking about 50%, if our best guess is that the level of savings we are estimating is right, then we go with that. But that's different from 51% of MPIs saying we are on track and 49% saying we are off track. Those things seem different. If our best guess is that we get 100 units of savings, I accept that the error bands might be different than an RA rebate program, that's different than 49% of MPIs saying we're achieving nothing and 51% saying we are achieving everything. Is there a spectrum of achievement – how does it really work in progress to decide the size of the savings claim?

  [Jeff Harris] Your confusion means I successfully raised some questions about how difficult this is. The PoE as a legal standard may not be appropriate – the very question you are asking. The challenge of creating a quantified trigger point – a level that says everything is okay or a range that shows we are going in the right direction.

  [Chris Neme] I'm open minded about the idea of 51% of evidence, or the preponderance at that threshold suggests we are going to get the savings, not opposed to that basis for claiming savings. Concerned about how we try to apply and operationalize the standard. Percent of MPIs or best guesses where they are going individually or collectively.

  [Jeff Harris] Good questions. Time is also a really important factor in all of this. For most MT innovations, the adoption curve will take a decade or more. A single year slice out of that is "blind men and the elephant" exercise – looking at just that part of the curve and its challenges that might not be relevant 2 years later or might build to eventual achievement of the MPI in 3 years, but you can't tell from it today if you are going to get there. This is fundamentally different component of MT program assessment and evaluation than it is for a RA framework.

  [Chris Neme] We have to be careful as we try to make a functional approach, that we don't oversimplify things. We could have ten MPIs but not all are of equal importance and their relative importance can change over time. If the least important half points in the right direction and the most important point the other way, that could be a problem too.

[Ted Weaver] First, agree with Chris that the framing of 51% question is provocative. Implies that the evaluators job is to find an ex ante number and the PoE says it is right or not. Some evaluators have done that, but I don't think that's how we do it in IL or most jurisdictions these days. Now evaluator says "this is the number I think is right" – the standard is do we think by consensus that the number is about right. Second, thanks for calling out the attribution issue. The framework in IL is we define a measurement protocol (e.g. TRM algorithm and NTG adjustment for RA) – and there are some analogues in MT that don't map perfectly. But both frameworks get you to the same kind of answer. We need to come up with applications of the measurement protocol that we understand and should be analogous to the RA side – when things change, we understand how those changes will be applied proactively or retrospectively. We should apply appropriately. I think they are all definable and things in the Policy Manual discussion we've been advocating for. Third, raising the CPAS question, we've been talking about that dimension in the chat. Gas utilities don't measure with CPAS and even on electric side you have to turn it into an annual number as well.

Presentation by Dan Violette, Apex Analytics

- We've been asked to discuss theory-based evaluation and some of the issues Jim brought up. The context has already been established. There was a quote in the 2022 ACEEE summer study paper by Val Jensen (ICF) – existing program designs have a bias toward programs that are easily evaluated. In the early days of EE programs, every utility did commercial lighting. Easy to calculate the savings, specific protocol. They wanted to implement because the procedure was in place and it was less risky. We have to overcome that for MT programs.
- Many important decisions and resource decisions fall in the category where we can't do this quasi-experimental design to answer all the important questions. We have to figure out what is cost-effective and the best way to evaluate them.
- Theory-based evaluation (TBE) tends to focus on theories of change. The point is that you have to have a logic model first. Then it is the story of what happens in the arrows between the boxes and what happens there. The theory of change can be tested.
- Example of ESRPP model (see presentation). The question is how successfully we got there. What metrics give you confidence that you have overcome those barriers? Performance indicators can be tracked and see if you hit the targets you have set. A lot of people think of the logic models and program theory and don't understand the difference between the two.
    - NYSERDA has done a good job with MPIs. One of the things that they did, they looked at EE skeptics in the market – overcoming the skepticism was an important factor, so a measure of increasing confidence, and another indicator that was looked at in the "non-quality" of delivery and market activities. Mistakes get made, phantom installation, drive-by audits. Quality of program delivery is important then.
    - Tracking MPIs is important and is the core of TBE. The question then is whether that is enough – if you specify good indicators is that all you need to do?
- What does "preponderance of evidence" mean? Burden of proof.
    - >50% probability of hypothesis being true
    - Not all indicators are the same (key indicators, other indicators) – need to be careful about that.

    - The hypothesis is more likely to be true than not – the definitions often use the term probability which implies a quantitative relationship. That can be a trap. Not advocating for a lot of quantitative methods – need something that is more common sense.
    - What are we trying to prove is more likely than false? How do you frame the question? Some questions are harder to get 50% probability of being true.
    - The PoE claim is somewhat unsatisfying. We have to put context around it. In almost any evaluation of MT, you are going to exceed that 50% criteria. By focusing there, it allows critics to say "you aren't doing enough." But if you do a good market evaluation like NEEA and NYSERDA, they do much better than 50% probability. That anchor number can hurt us in a way. It's a number to be exceeded and we expect to exceed it by a large margin. Thinking about the legal context – e.g. civil case; jury has to be a super-majority or unanimous depending on the state. Must exceed 50% and convince a super-majority that we are achieving that.
    - A little leery of legal comparisons – don't want to get trapped by those constraints. But when you think about super-majority of reasonable people accepting the criteria – that's often overlooked in the context.
    - We need to recognize that resource acquisition uses judgment much more often than experimental evidence.
    - We have to make mid-course corrections, to help manage program risk. This is an interesting question – a new program maybe you should collect MPIs more frequently to make sure program is on track, or the position that you have to wait a year to have enough evidence. But then you could be off-track or change the slope of the curve. We have decided that roll-out should be different for evaluation than a mature program. This is something you can argue.
    - Drawing on NYSERDA work, we often looked at small sample verifications. No control groups or large groups of participants and non-participants. But could validate 10-12 installations – a "high validity study" – a lot of proof on a small sample. Then the MPIs can allow you to extrapolate that and show it works.
    - Structured expert panel – analogous to the jury – could be assembled.
    - Looking at TBE compared to current approaches that have been used in Illinois
        - Lots of aspects of attribution depend on judgmental decisions – the word logic appears throughout Attachment C. Also in the NTG TRM attachment.
        - Lots of decisions made about what NTG ratio is being used. A lot of the program theory concepts are already there. And there are surveys of market actors. Given the use of logic models in NTG, you could say that the basic elements of program theory evaluation aren't new to Illinois. Could see it as an extension of what is already being done, rather than a revolution.
    - Expert judgement – referenced in Attachment C. Need to think about the kind of decision being made. Some questions are not appropriate for experimental design. Expert judgement is used in this context. It often holds up well in a regulatory context. Random control trials have to be stitched together with other studies from other places and times, and the expert judgement figures out how they work together.
    - There are examples in the literature – DOE UMP NTG chapter, there is some experience with it. We should have more applications that we have seen. Surprised there hasn't been more.
- Example of the Wind Power America program for DOE

    - Not enough for experimental design, so used range estimates – expert lower and upper bounds. This allowed experts to help us dimension the uncertainty in different outcomes. Some outcomes were more uncertain than others. I wonder about the lack of use of expert judgement and the support for structure frameworks, maybe it's a communication issue.
- IRP supply-side decisions use a lot of expert judgement, so we shouldn't be apologizing for doing it here.
- Range estimation is important for the credibility of these studies.
    - Consider factors & the level of impact of them at the beginning – to get them thinking about what creates the lower and upper bounds of the range. Also the importance of the factors. It helps in program design and in future research.
    - Start by specifying the lower bound, then the upper bound, then a likelihood estimate. E.g., divide into bins (upper, middle, lower thirds) – which bin is likely to have the true value in it? Have started to move away from trying to get point estimates. Need reasonable questions they can respond to. Range and sub-range estimates are the way to go. You tend to get a smooth distribution with a half dozen experts.
- Why not successful? Not expensive. Maybe it is the case of not having enough experience with these methods. Lots of industries use Delphi panel approaches. We ought to develop protocols so we can do it more. We have engineers and economists and some of them may be uncomfortable with some of this type of approach, whereas policy folks may be more comfortable.
- Some examples where EE/DSM decision makers used results of these studies
    - NEEA Sunset Study
    - Iowa Utilities Study
    - Ontario Study
- Success factors
    - Thoughtful selection of experts. How you organize the results. How do you deal with outlier experts.
    - Don't ask questions that experts simply cannot answer – too much resolution. Use ranges of estimates. Get feedback and re-estimate. Don't use point impacts – use "fuzzy" Delphi.
    - At the end, ask the experts if they felt the process was credible and represented their views.
    - Conclusions – don't let evaluation approaches overly influence the decision to implement – don't get so wrapped up that you don't undertake new programs. Establish a context that supports good decision making. MPIs will be the core of the evaluation effort, but other things such as small sample studies and structured expert judgment around the edges.

[Ted Weaver] In Illinois we have a framework for expert panels that is very functional for the TRM and NTG processes. The experts are the entire SAG and our approach gets to decisions that have worked for 12 years. We can leverage that. Also, while this can be complicated, we are ready to do complicated things. We have a transparent framework where everyone can understand the math and the risks. We can come up with it for MT as well.

Closing and Next Steps

Celia Johnson, SAG Facilitator

Next MT Small Group meeting: Wednesday, November 30. Planned topics:

- Follow-up on October meeting – summary of next steps (Jim Fay, ComEd)
- Introduce Topic 2 – MT Savings Framework (Jim Fay, ComEd)
- Introduce Final Draft Market Transformation Policy Resolution from 2020 (SAG Facilitator)
- Goal is to identify volunteers to draft IL-TRM Attachment C edits and/or policy options for the group to discuss at the December meeting