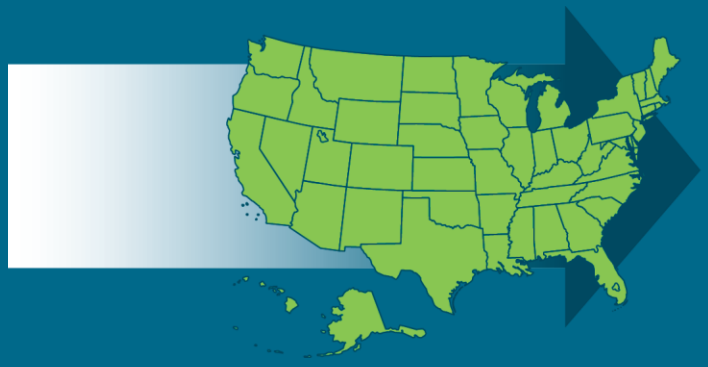


SEE Action

STATE & LOCAL ENERGY EFFICIENCY ACTION NETWORK

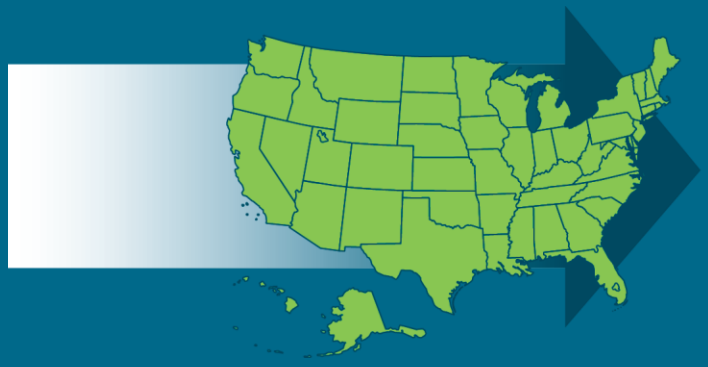
Evaluation, Measurement, and Verification (EM&V) of Residential Behavior-Based Energy Efficiency Programs:

Issues and Recommendations

November 13, 2012

Michael Li *U.S. Department of Energy*

Annika Todd *Lawrence Berkeley National Lab*



SEE Action

STATE & LOCAL ENERGY EFFICIENCY ACTION NETWORK

State and Local Energy Efficiency Action Network

Goal: achieve all cost-effective energy efficiency by 2020

- State- and local-government led initiative to take energy efficiency to scale, facilitated by U.S. DOE and U.S. EPA
- Network of 200+ professionals from state/local governments, business, industry, NGOs and others
- Best practice guides and technical assistance on EE policy and program design and implementation for:
  - State utility regulators and utilities
  - State and local policymakers

SEE Action
STATE ENERGY EFFICIENCY ACTION NETWORK

What SEE Action Offers

1. Decision-grade guidance documents based on state and local experience
   - Best practices & model policies
   - Successful approaches
   - Recommendations
   - What's working

2. Discussion forums to identify solutions to known barriers

3. Technical assistance from the best subject matter experts in the country

Sign up for news alerts at www.seeaction.energy.gov

Home | About | Working Groups | Information Resources | Technical Assistance

SEE Action

STATE & LOCAL ENERGY EFFICIENCY ACTION NETWORK

The State and Local Energy Efficiency Action Network (SEE Action) is a state- and local-led effort facilitated by the U.S. Department of Energy and the U.S. Environmental Protection Agency to take energy efficiency to scale and achieve all cost-effective energy efficiency by 2020. SEE Action offers information resources and technical assistance to state and local decision makers as they provide low-cost, reliable energy to their communities through energy efficiency.

SEE Action Working Groups »

- Existing Commercial Buildings
- Industrial Energy Efficiency and Combined Heat and Power
- Customer Information and Behavior
- Evaluation, Measurement and Verification
- Building Energy Codes
- Driving Ratepayer-Funded Efficiency through Regulatory Policies
- Financing Solutions
- Residential Retrofit

News »

» New SEE Action Resource: Evaluation, Measurement, and Verification (EM&V) of Residential Behavior-Based Energy Efficiency Programs
May 22, 2012

» New SEE Action Resource: Benchmarking and Disclosure: State and Local Policy Design Guide and Sample Policy Language
May 22, 2012

Events »

» EM&V Webinar: Documenting Avoided Emissions from Energy Efficiency Programs
August 8, 2012

Stay Updated »

» Receive the newsletter and announcements of upcoming events and new publications from the SEE Action Network

SIGN UP

Outline: EM&V of Behavior-Based EE Programs

- What is a behavior-based EE program?
- Why is evaluation of these programs hard?
- **How can we be confident that the energy savings are valid?**
- What are key guidelines on best practice methods (and why are RCTs the gold standard)?

Outline: EM&V of Behavior-Based EE Programs

- **What is a behavior-based EE program?**
- Why is evaluation of these programs hard?
- **How can we be confident that the energy savings are valid?**
- What are key guidelines on best practice methods (and why are RCTs the gold standard)?

What is a behavior-based EE program?

Behavior-based energy efficiency programs are those that utilize strategies intended to affect consumer energy use behaviors in order to achieve energy and/or peak demand savings. Programs typically include outreach, education, competition, rewards, benchmarking and/or feedback elements.

- Programs that affect the way that consumers use energy (without using traditional methods, such as rebates or time-based tariffs)
- Instead, use simple psychological levers or information to change behavior

What is a behavior-based EE program?

- Example 1: Comparing your energy use with your neighbors
- Example 2: Providing real-time information and feedback about energy use
- Example 3: Goal setting and reward points per kWh saved

What are the potential benefits and concerns of behavior-based programs?

- **Potential Benefits**
  - In theory, potentially cheap to implement and result in significant energy savings → **cost effective**
  - Currently, some **examples** of well designed, rigorously evaluated programs that show savings
  - As a result, increasingly being **adopted nationwide**

- **Potential Concerns**
  - These programs are **relatively new**
  - Evidence of energy savings in different types of programs, different situations, and program persistence is unclear
  - Potential for unsubstantiated claims (anecdotal evidence)

Why is rigorous evaluation crucially important?

→ It is very important to accurately evaluate the effectiveness of these programs

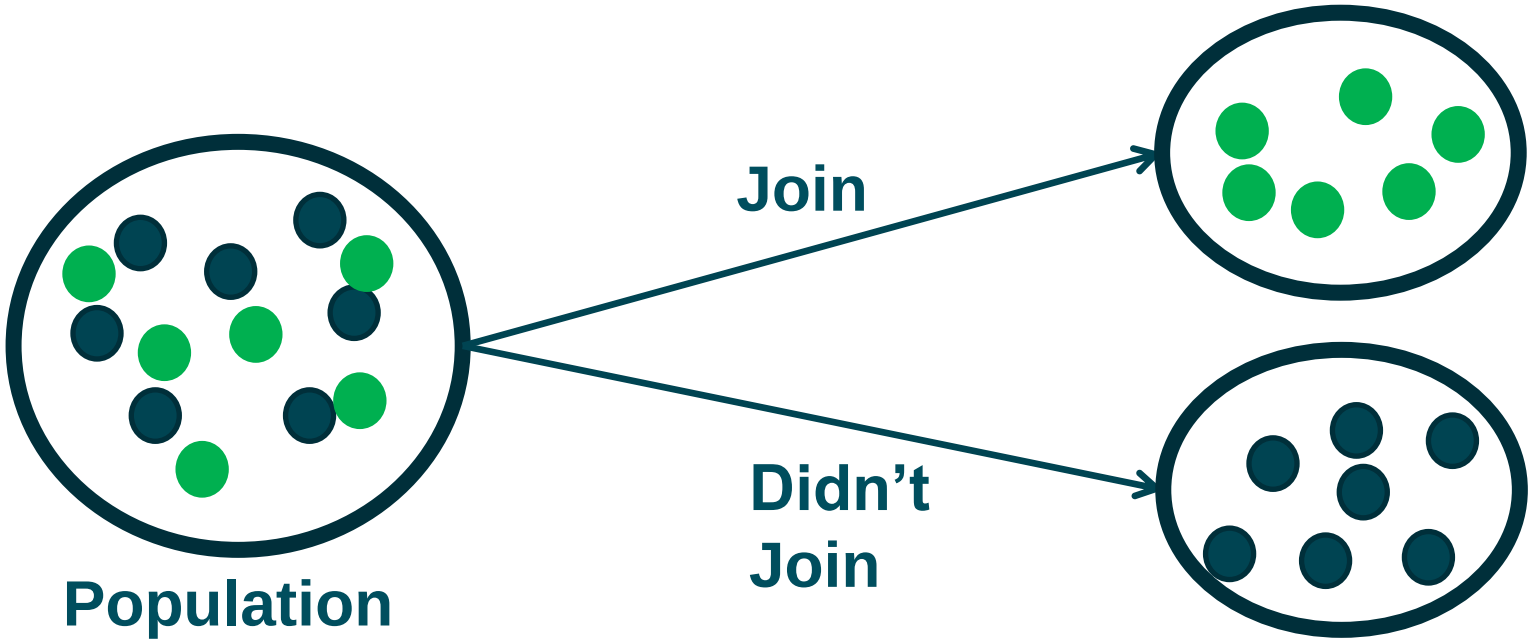
- For planning purposes - gain information about how well different types of programs work in different situations
- For validly claiming energy savings

Outline: EM&V of Behavior-Based EE Programs

- What is a behavior-based EE program?
- **Why is evaluation of these programs hard?**
- **How can we be confident that the energy savings are valid?**
- What are key guidelines on best practice methods (and why are RCTs the gold standard)?

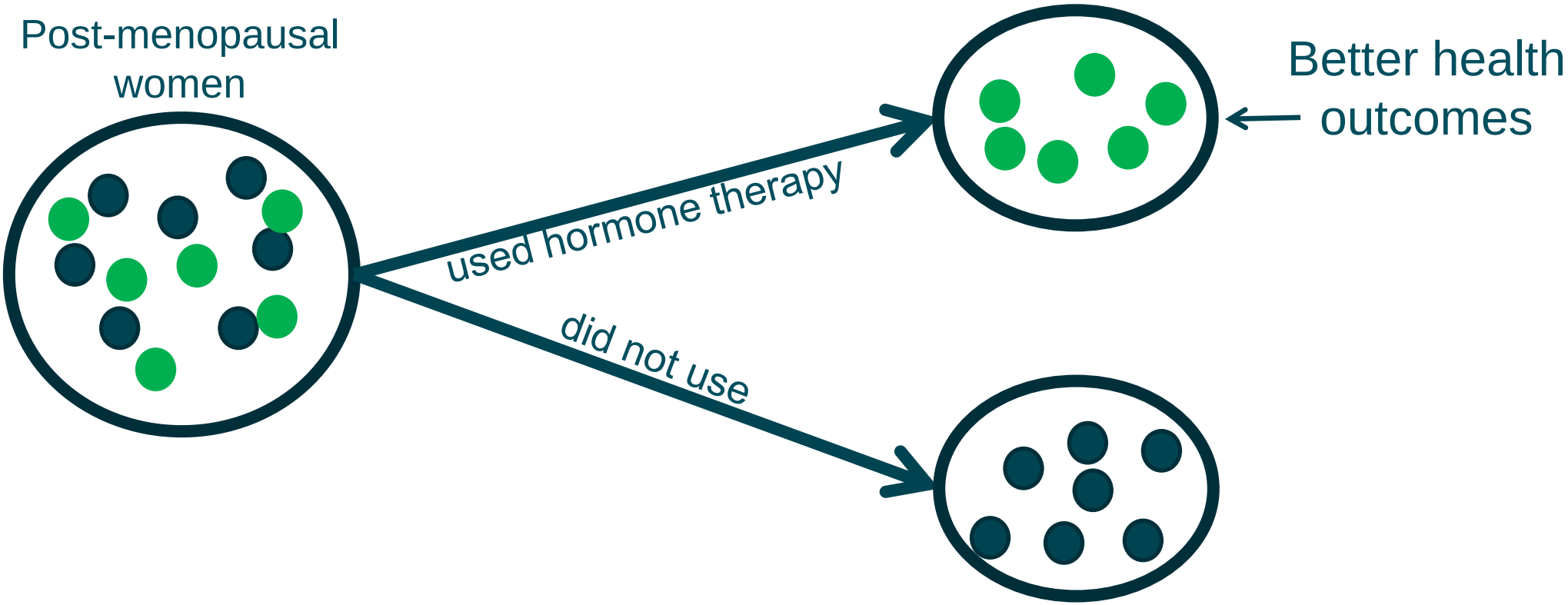
Why is evaluation of these programs hard?

- Strong problem of **"Selection Bias"**: households that join (e.g., opt-in, screened) are fundamentally different



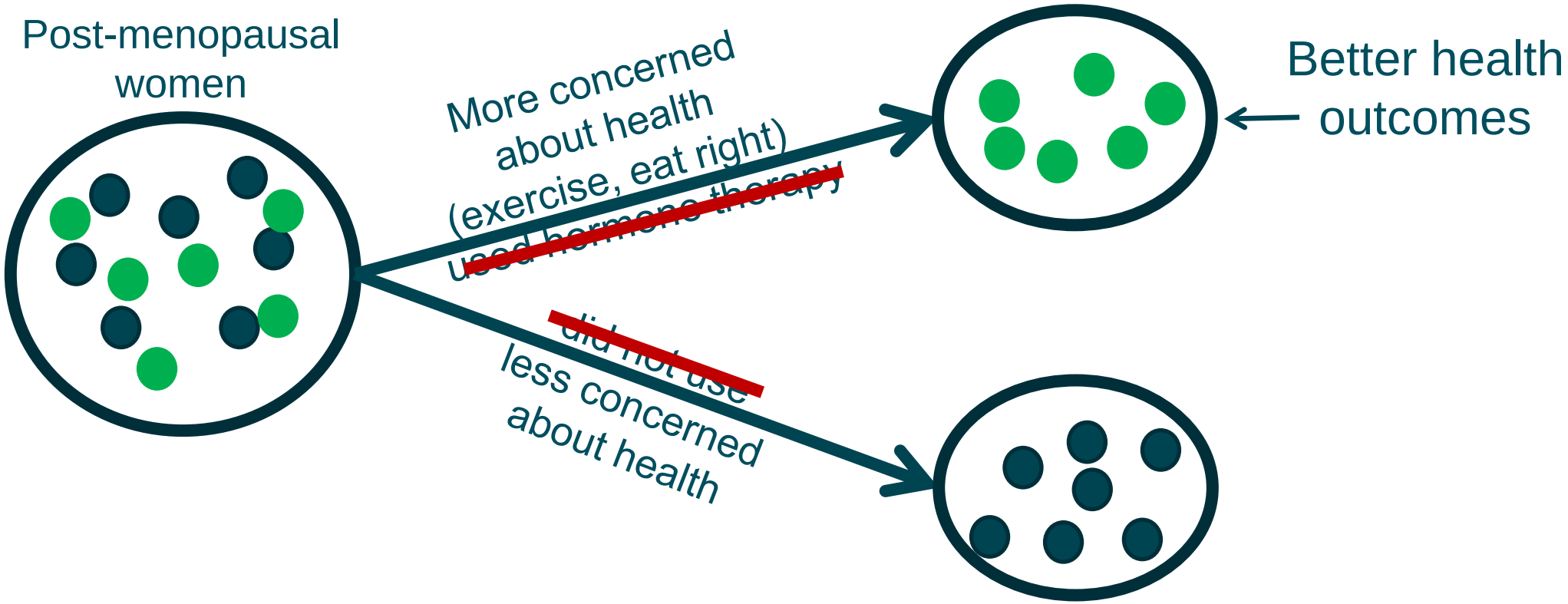
- Observed differences *might* be due to program, but might just be a difference between groups
- Selection bias can skew the results of the evaluation

Example: Post-menopausal hormone therapy



- 80's study: women who used hormone therapy had better health outcomes. As a result, doctors recommended it to all post-menopausal women.

Example: Post-menopausal hormone therapy



- 80's study: women who used hormone therapy had better health outcomes. As a result, doctors recommended it to all post-menopausal women.
- Rigorous RCT study: hormone therapy has ***negative*** impacts - what happened?
- **Selection bias** in the non-RCT study: women who chose to use hormone therapy were different types of women
- Better health outcomes were because the two groups were different, NOT because of hormone therapy

Why is evaluation of these programs hard?

- Behavior-based programs may be difficult to rigorously evaluate compared to other programs (e.g., appliance rebates):
  - **Savings are *relatively small*** (often 1-5%), so if an evaluation is biased (off by a few percentage points), could change conclusions about how effective the programs are
  - **Currently, less of a foundation for engineering estimates.** (What behaviors are people doing to save energy?)

Why is evaluation of these programs hard?

→ **Bad evaluation could lead to bad policy decisions**

- Implement programs that are not cost effective
- Screening out programs that may be cost effective

Outline: EM&V of Behavior-Based EE Programs

- What is a behavior-based EE program?
- Why is evaluation of these programs hard?
- **How can we be confident that the energy savings are valid?**
- What are key guidelines on best practice methods (and why are RCTs the gold standard)?

"EM&V for Residential Behavior-Based Energy Efficiency Programs: Issues and Recommendations"

- Provides guidance and best practices
  - For program design, analysis and evaluation methods

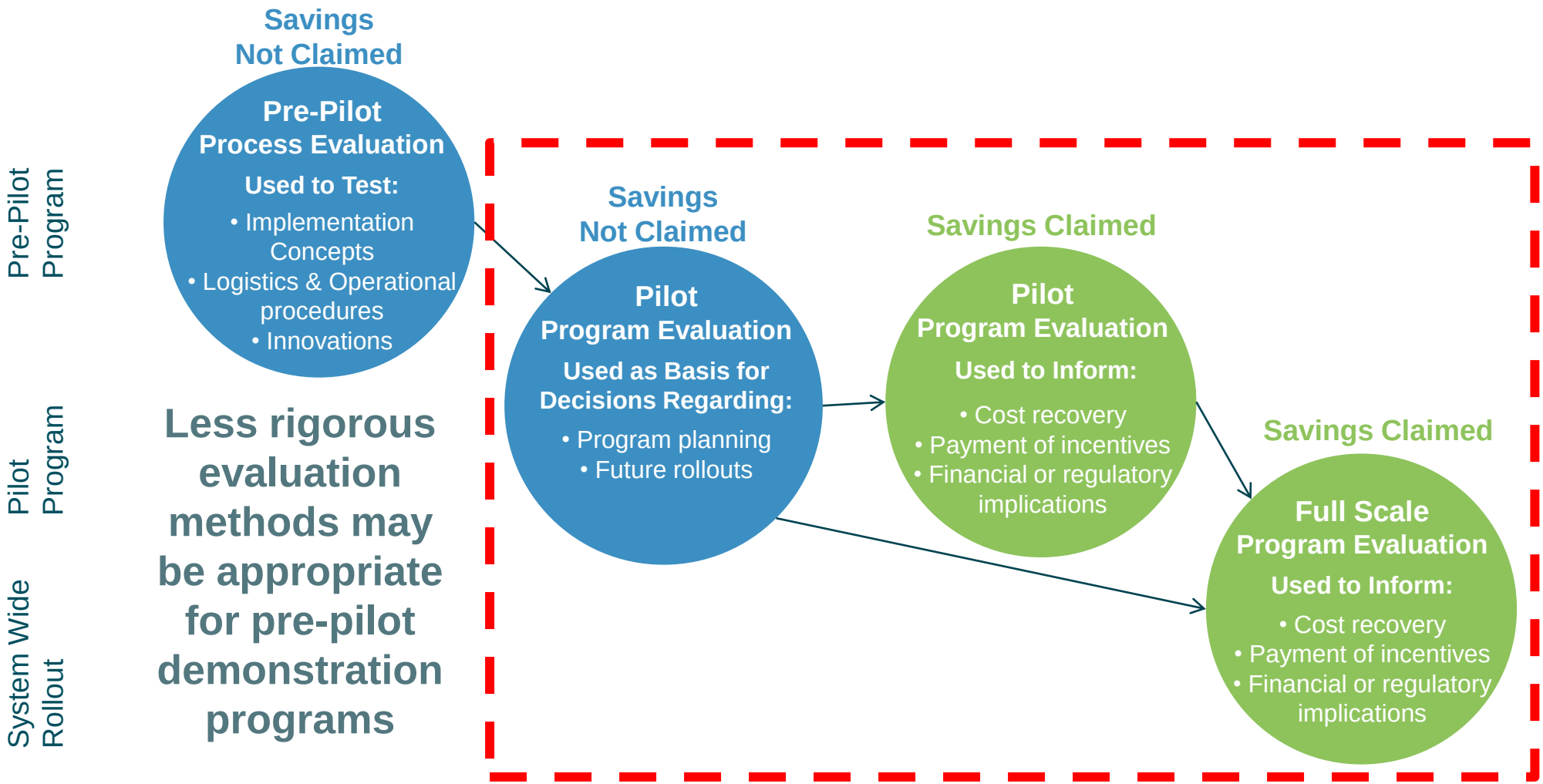
→ Ensure a **high degree of confidence** that estimated program energy savings impacts are **valid**

- Guidance is based on:
  - Consensus of researchers in many different fields and environments
  - Vetted by ~75 reviewers: technical, academics, program administrators, regulatory agencies, industry stakeholders

"EM&V for Residential Behavior-Based Energy Efficiency Programs: Issues and Recommendations"

- Target audiences:
  - Regulators, program administrators, evaluation professionals, stakeholders
  - Those responsible for overseeing and reviewing efficiency program designs and evaluations
- Experienced, sophisticated evaluators may already be familiar with these recommendations

Scope: Typical Program Life Cycle



Less rigorous evaluation methods may be appropriate for pre-pilot demonstration programs

Focused on pilot or full scale programs that are claiming savings or are used to make decisions about future rollouts

Outline: EM&V of Behavior-Based EE Programs

- What is a behavior-based EE program?
- Why is evaluation of these programs hard?
- **How can we be confident that the energy savings are valid?**
- **What are key guidelines on best practice methods (and why are RCTs the gold standard)?**

SEE Action
STATE & LOCAL ENERGY EFFICIENCY ACTION NETWORK

Key recommendation 1: use a randomized controlled trial (RCT)

Randomized controlled trial (RCT)

Regression discontinuity

Variation in adoption

Propensity score matching

Non-propensity score matching

Pre-post comparison

Key recommendation 1: use a randomized controlled trial (RCT)

Randomized controlled trial (RCT)

Regression discontinuity

- Primary recommendation – a program that is designed as a RCT results in:
  - Transparent, straightforward analysis
  - Robust, accurate, valid program impact estimates
  - **High degree of confidence in program evaluation**
  - RCTs are the gold standard

Key recommendation 1: use a randomized controlled trial (RCT)

Randomized controlled trial (RCT)

Regression discontinuity

- Why is designing a program as a (RCT) so important?
  - RCT means that households are assigned to the program randomly (as opposed to household choice or screening criteria)
  - Solves selection bias

Key recommendation 1: use a randomized controlled trial (RCT)

Randomized controlled trial (RCT)

Regression discontinuity

- RCTs have many different forms
- Can be used for Opt-in, Opt-out programs

Key recommendation 1: use a randomized controlled trial (RCT)

- If RCTs are not feasible, acceptable "quasi-experimental" methods
  - More opaque, complex analysis
  - Quasi-experimental methods try to correct for selection bias
  - Lower degree of confidence in validity of savings estimates

Key recommendation 2: avoiding potential conflicts of interest

- **Problem:** potential for a conflict of interest to arise regarding the validity of savings estimates
- **Recommendation:**

A third-party evaluator transparently defines and implements:

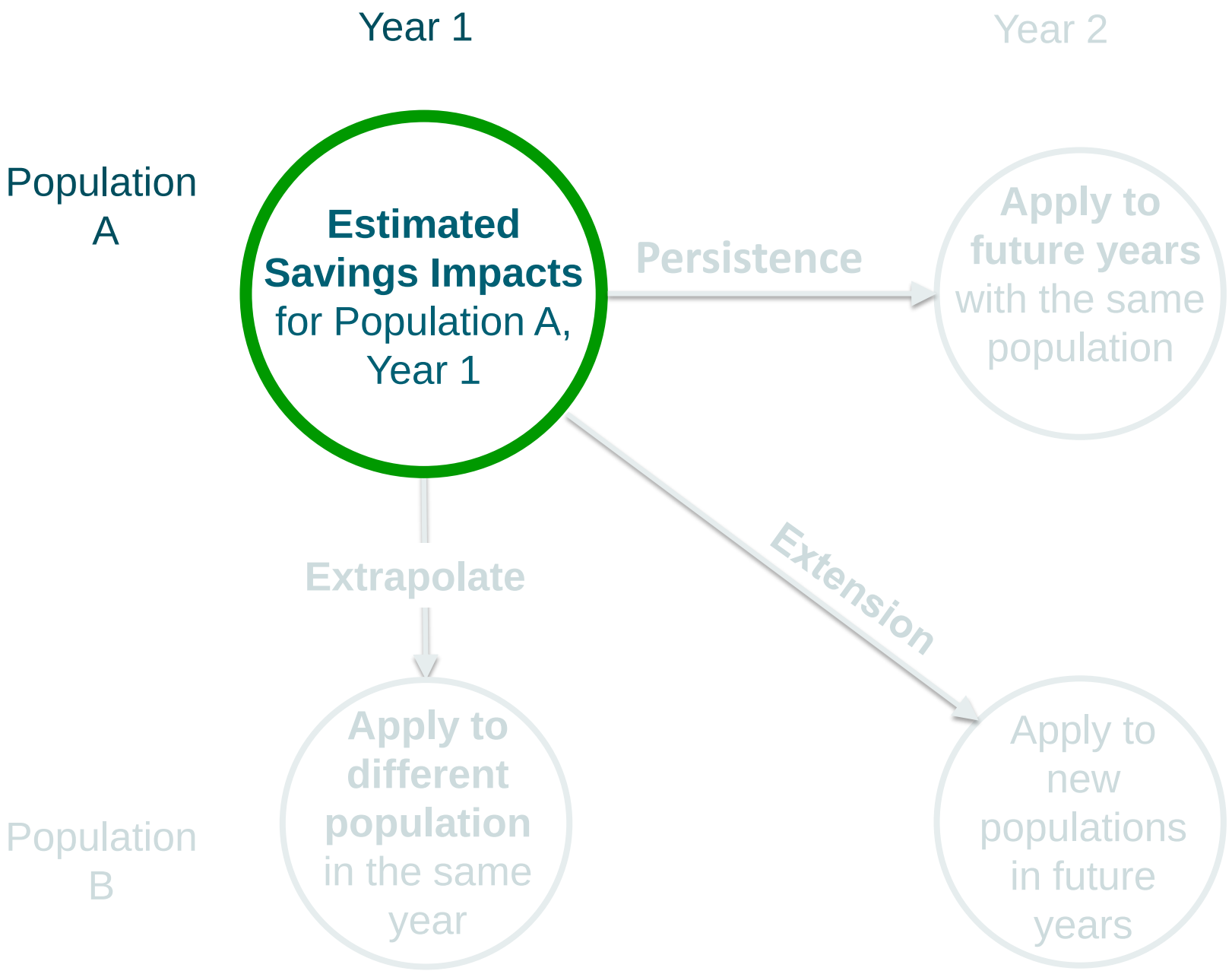
- Program evaluation
- Assignment of households to control and treatment groups
- Data selection and cleaning

Program implementer or sponsor implements any of the above

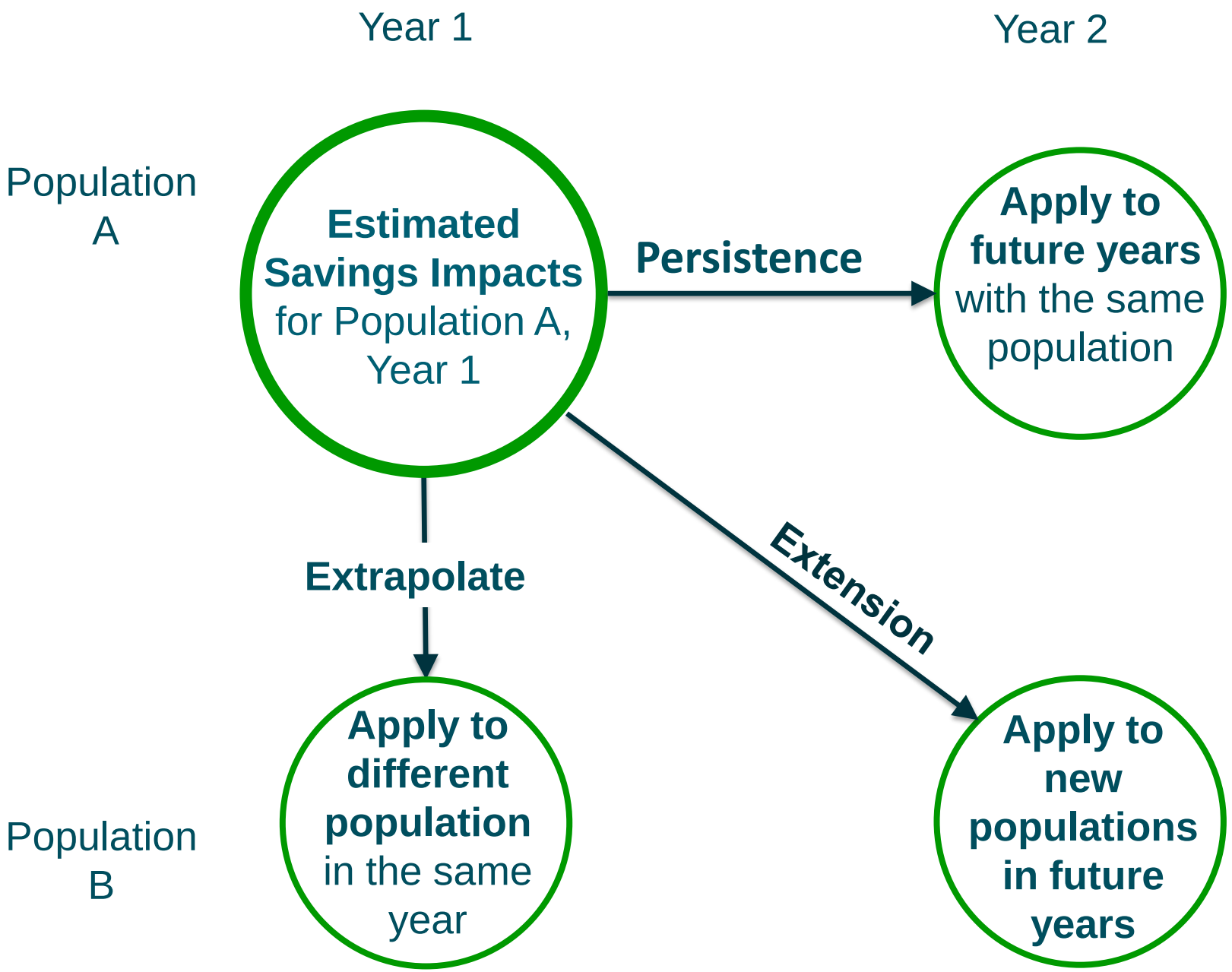
Key recommendation 3: accounting for potential double counting of savings

- **Problem**: the same savings may be claimed by two programs (e.g., a behavioral program & appliance rebate program both claim savings from appliances)
- **Recommendation:** estimate this "double counted savings" overlap to the extent possible by comparing control to treatment group
  - Easier for programs that can be *tracked* at the household level (e.g. installation of insulation by a contractor)
  - Should account for the measurement period (e.g., accounting for seasonal load impacts), and the effective useful lifetime of installed measures (when lifetime savings are reported)
  - Program costs should be appropriately allocated along with double counted saving

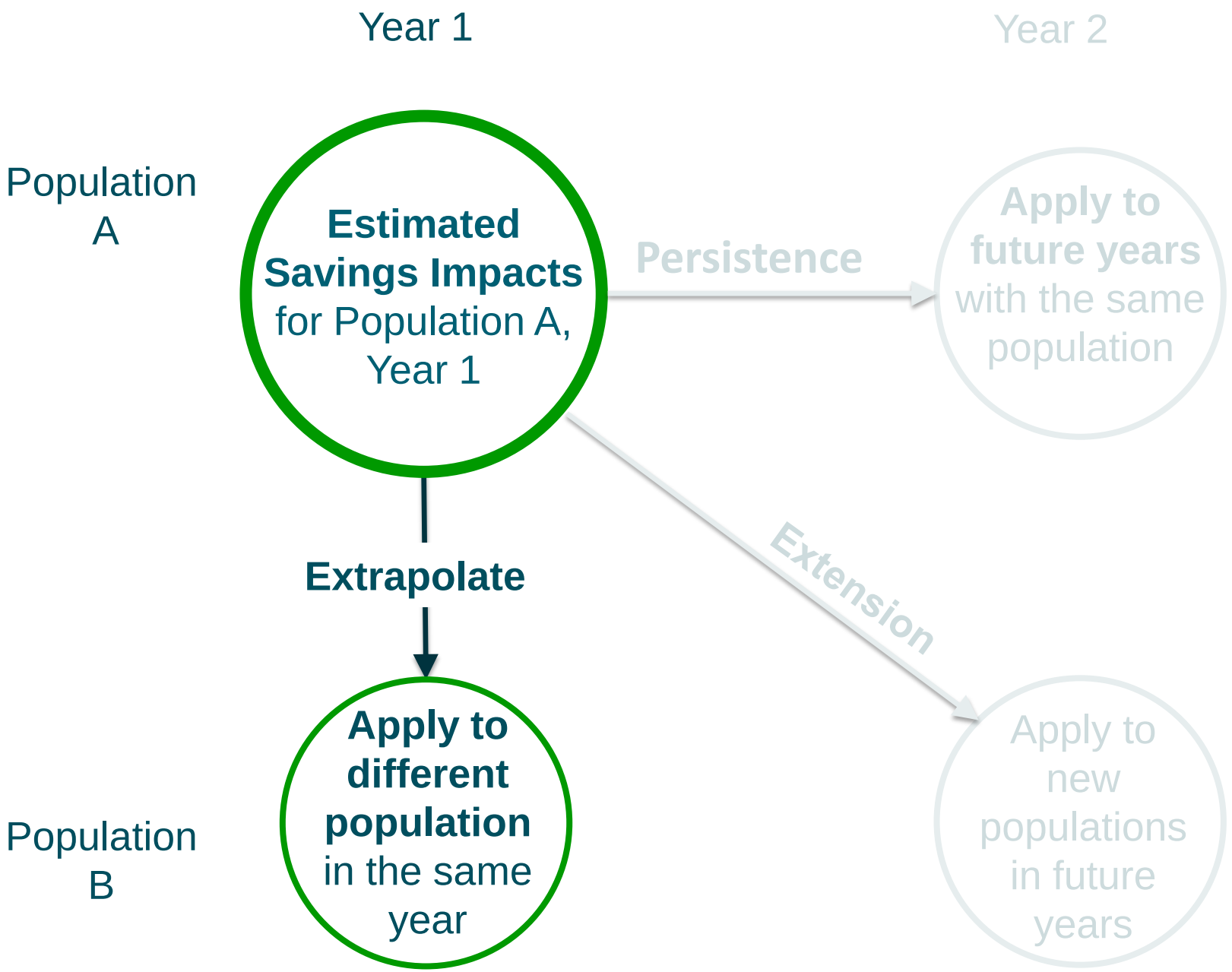
Key recommendations 1,2,3 address internal validity (for a given population, time frame)



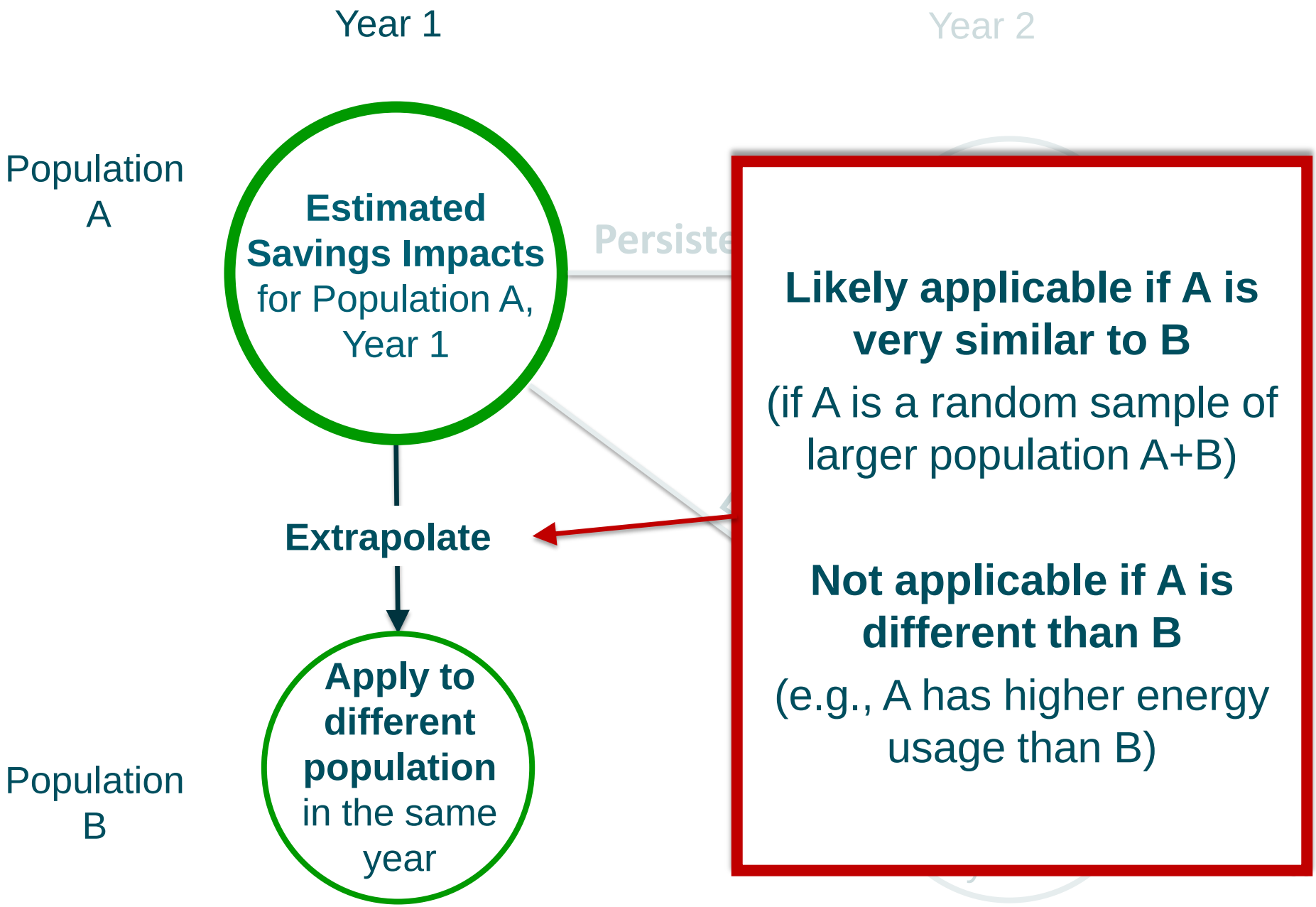
Recommendations for external validity: can the savings be applied to new situations?



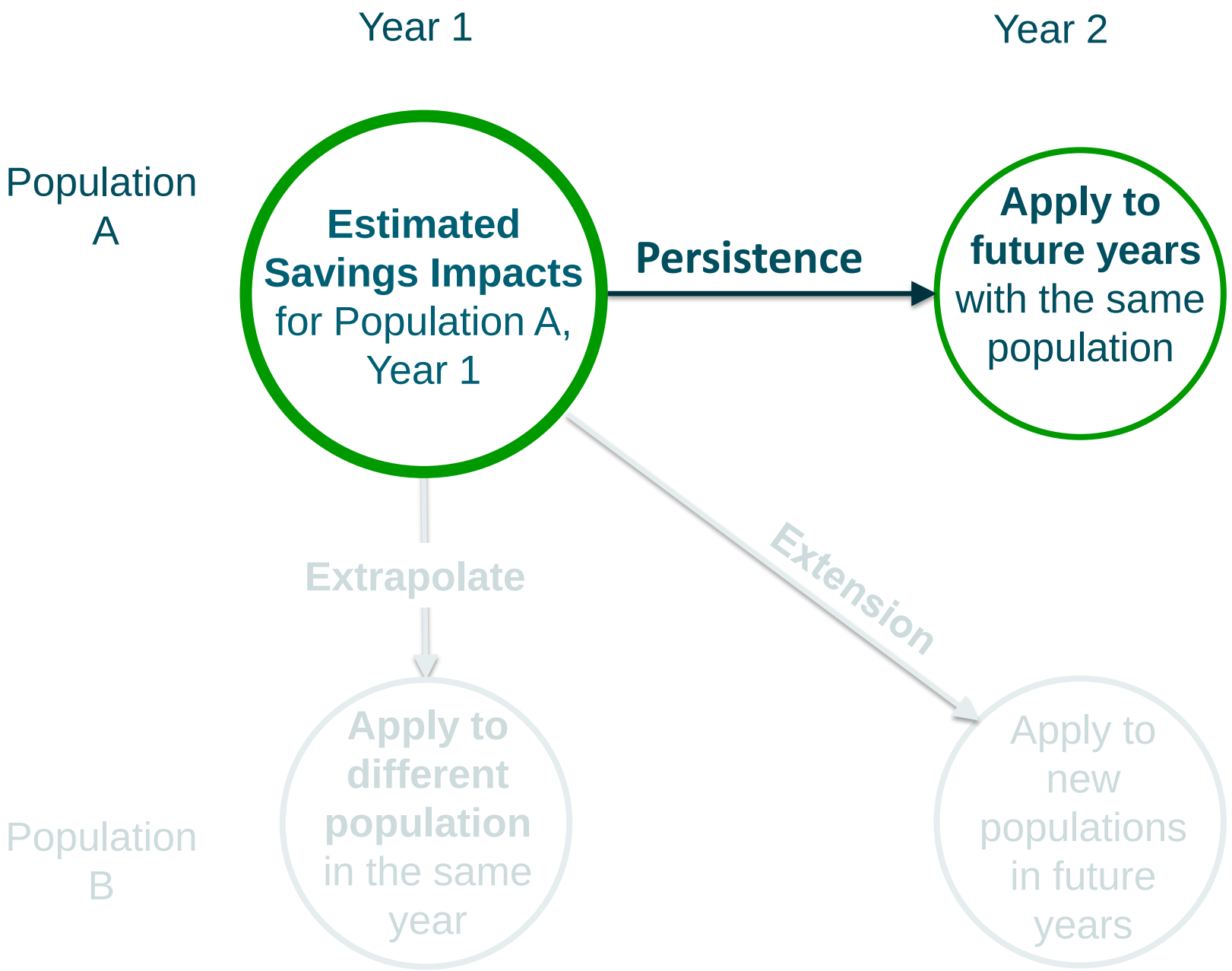
Are the savings applicable to different populations?



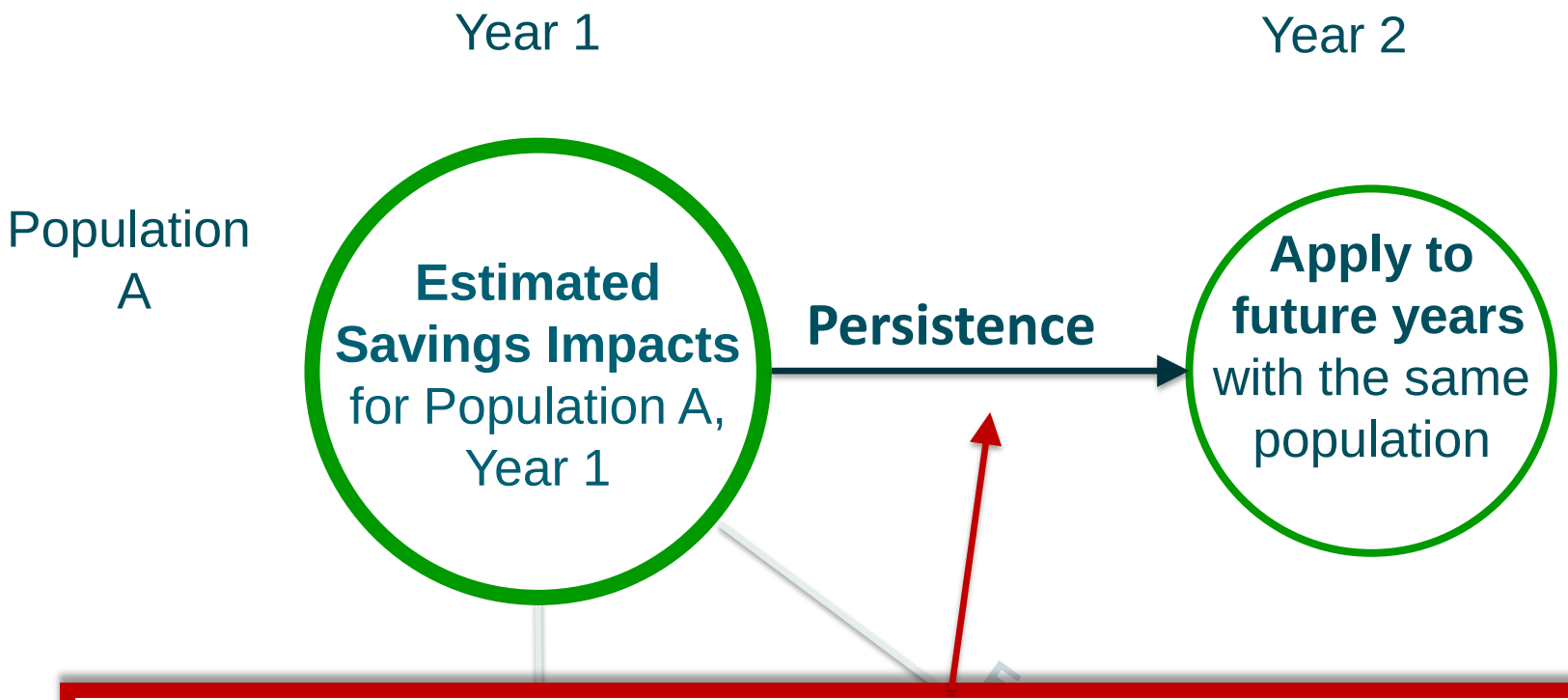
Are the savings applicable to different populations?



Do the savings persist over time if the program continues? If it stops?



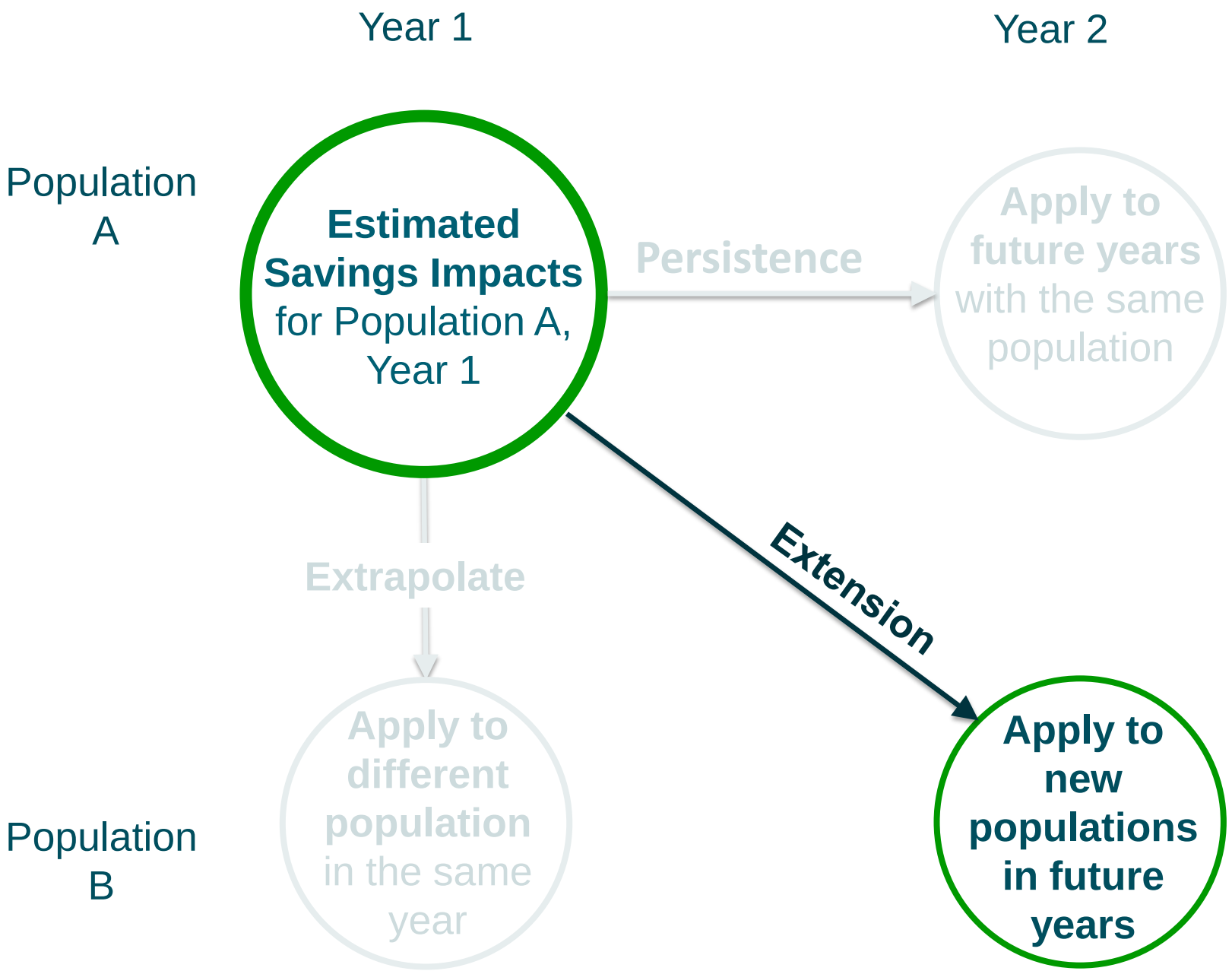
Do the savings persist over time if the program continues? If it stops?



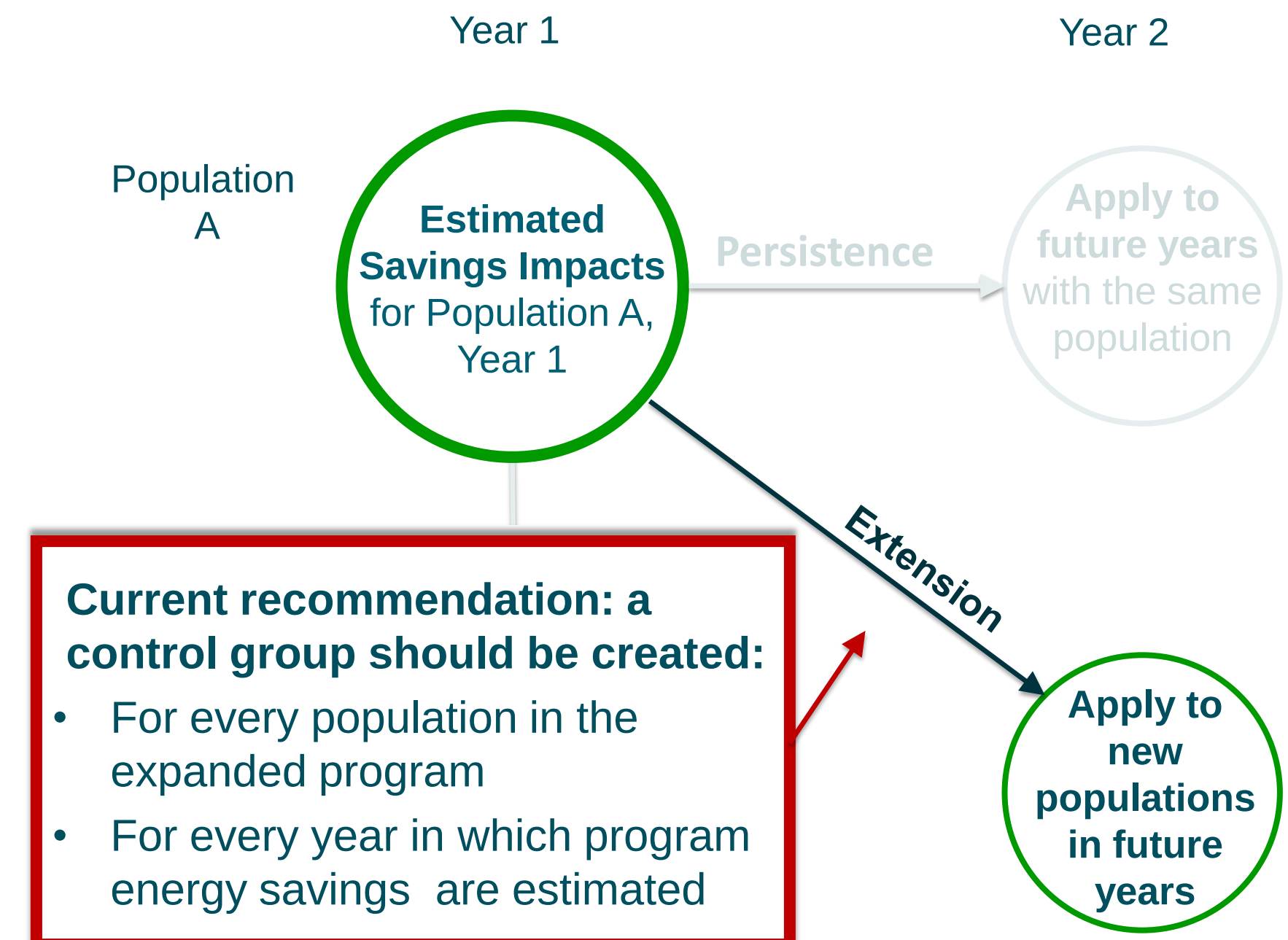
Until there is enough evidence on persistence in behavior-based programs, recommend:

- A control group is maintained for every year in which program impacts are estimated
- Evaluation is done each year initially, every few years after it has been running for several years

If the program is extended to a new population, is the initial savings impact valid?



If the program is extended to a new population, is the initial savings impact valid?



Current recommendation: a control group should be created:

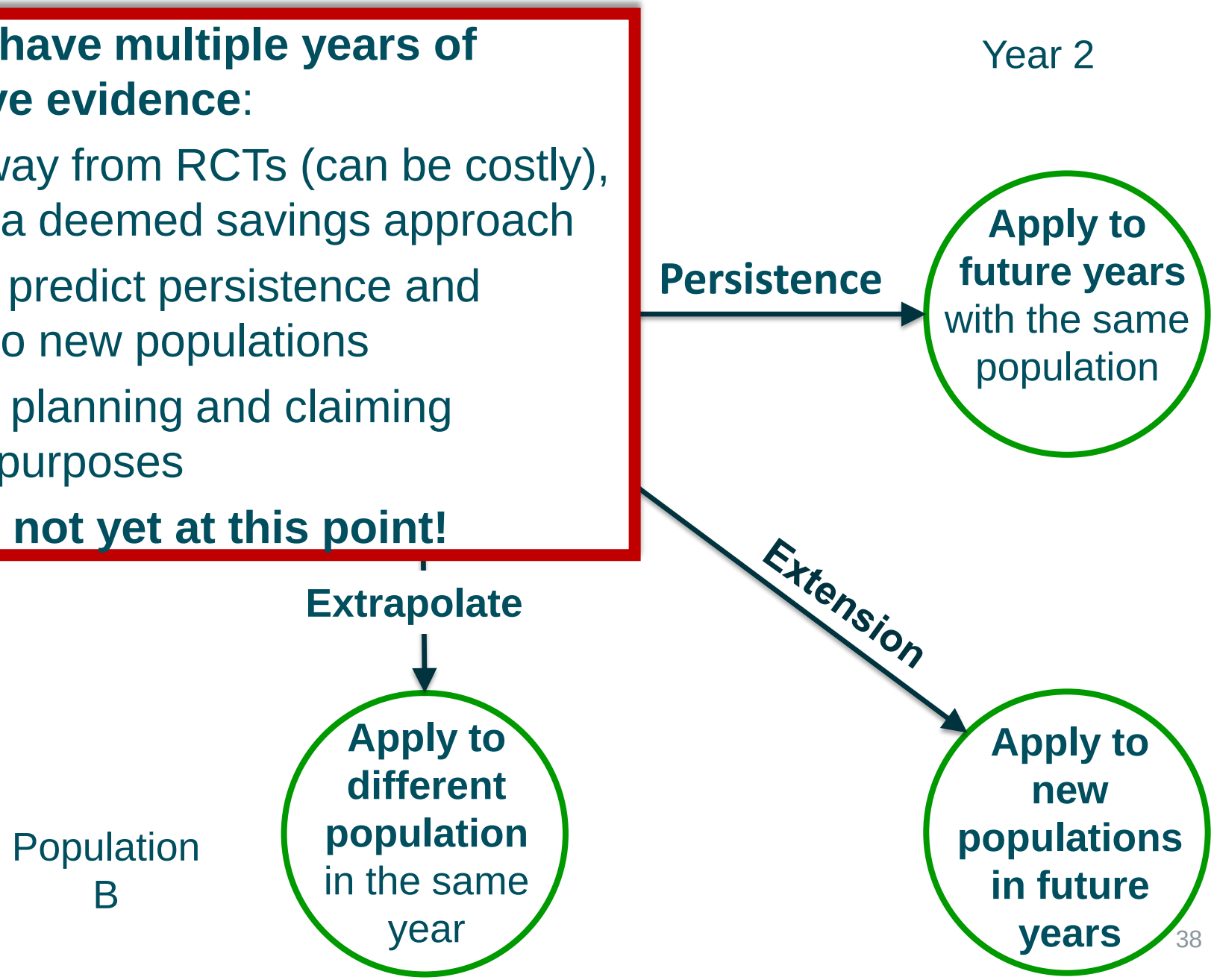
- For every population in the expanded program
- For every year in which program energy savings are estimated

In the future, can we move away from RCTs into a deemed savings approach?

Once we have multiple years of conclusive evidence:

- Move away from RCTs (can be costly), towards a deemed savings approach
- Credibly predict persistence and rollouts to new populations
- For both planning and claiming savings purposes

→ **We are not yet at this point!**



Conclusions & next steps

- **Main point:** if the recommended methods are used (gold standard is RCTs), then we can be confident that the program's energy savings are valid
- **This issue is timely**
  - Around 40 utilities are currently offering behavior-based EE programs, considering going system wide
- **New research provides insights into:**
  - Persistence of behavior-based programs
  - What behaviors are causing the savings

Questions?

- **Many guidelines and technical recommendations in the report:**
  - SEE Action website, www.seeaction.energy.gov
  - Lawrence Berkeley National Lab website: behavioranalytics.lbl.gov
- **LBNL can offer technical assistance** to state PUCs and energy offices for EM&V guidance and best practices for behavior-based EE programs

Mike Li: Michael.Li@hq.doe.gov

Annika Todd: atodd@lbl.gov

Additional Technical Recommendations

Additional internal validity recommendations

- **Problem**: how to ensure that the estimate of program impact savings is precise enough, not risky
- **Statistical significance recommendation**:
  - Define null hypothesis (the required threshold, e.g., cost effectiveness)
  - Estimate considered acceptable if statistically significant at 5% (i.e., 95% confidence)
  - 5% statistical significance *NOT* the same as 95/5

Additional internal validity recommendations

- **Historical data recommendation**: collect twelve months or more of historical data
  - Especially if program design is quasi-experimental

- **Analysis recommendation**: the model specification (econometric techniques, e.g., regressions) should:
  - Use panel data (many data points over time) vs. aggregated data
  - Not include interaction variables
  - If quasi-experimental, compare the *change* in energy usage vs. energy usage

Excluding Data from Households that Opt-out or Close Accounts

Data cleaning: which households to exclude

Only data from households that closed accounts are excluded*; households that opt-out of the treatment or control group are included in the analysis (although the program impact estimate may be transformed to represent the impact for households that did not opt-out, as long as it is transparently indicated).

Data from households that closed their accounts are included*

Households that opt-out are excluded from the analysis

**If there is a compelling reason to include households that closed their accounts and the analysis is undertaken correctly to deal with unbalanced data sets, then it may be advisable.*

Cluster Robust Standard Errors

Ensure that the standard errors are robust and account for clustering

Cluster Robust Standard Errors or Time Aggregated Data

Non-Cluster Robust Standard Errors with non-Time Aggregated Data

Equivalency Check

Validate that the control and treatment group are equivalent

An equivalency check is performed with energy use data as well as household characteristics

An equivalency check is performed with energy use data

An equivalency check is not performed