The Retail Products Portfolio (RPP): Overview of the Evaluation Plan

Presentation for the California Technical Forum

March 20, 2015

Today's Agenda

- Market Transformation and the RPP Program
- Evaluation Objectives
- Evaluation Approach
- Impact Evaluation Methods
- Integration of Data Streams
- Appendix

Market Transformation

Resource acquisition:

- Objective: Generate energy savings to meet energy efficiency goals and/or alleviate grid constraints
- Time horizon: The here and now
- Tools: Primarily incentives

Market transformation:

- Objective: Change the structure and functioning of how a market operates to create more demand for energy efficient products thus generating long-term, sustainable energy savings
- Time horizon: The long term
- Tools: Incentives + Retailer/Manufacturer/Distributor Engagement + Marketing + Outreach/Education

Source: NEEA's Definition of Market Transformation

Market Transformation Program Evaluation Challenges

- Emphasis on theory-driven evaluation
- Measurement of short-term performance metrics
- Baseline mid-term performance and long-term performance and MT metrics
- Coordination with national-level evaluation efforts
- Assessment of possible comparison group designs to estimate broader market-level trends and effects
- Need to lengthen time horizon of cost-benefit analysis

Objectives of the Evaluation

Evaluation has both impact & process components to:

- Validate program theory & logic
- Identify, operationalize, collect, & analyze key short & mid-term program performance metrics
- Identify, operationalize, collect & track key mid- & long-term market effects metrics to develop baselines for comparison in future years
- Suggest frequency at which metrics should be collected
- Estimate gross & net energy & demand savings
- Continue to test various methods to evaluate the RPP Program
- Conduct secondary literature review of similar evaluation metrics for MT programs
- Provide the foundation upon which future evaluation methods & analyses are built to support defensible conclusions regarding the efficacy of the RPP Program
- Provide input, insight, & direction to how an RPP Program might be evaluated in various jurisdictions

Evaluation Approach: Theory-Driven

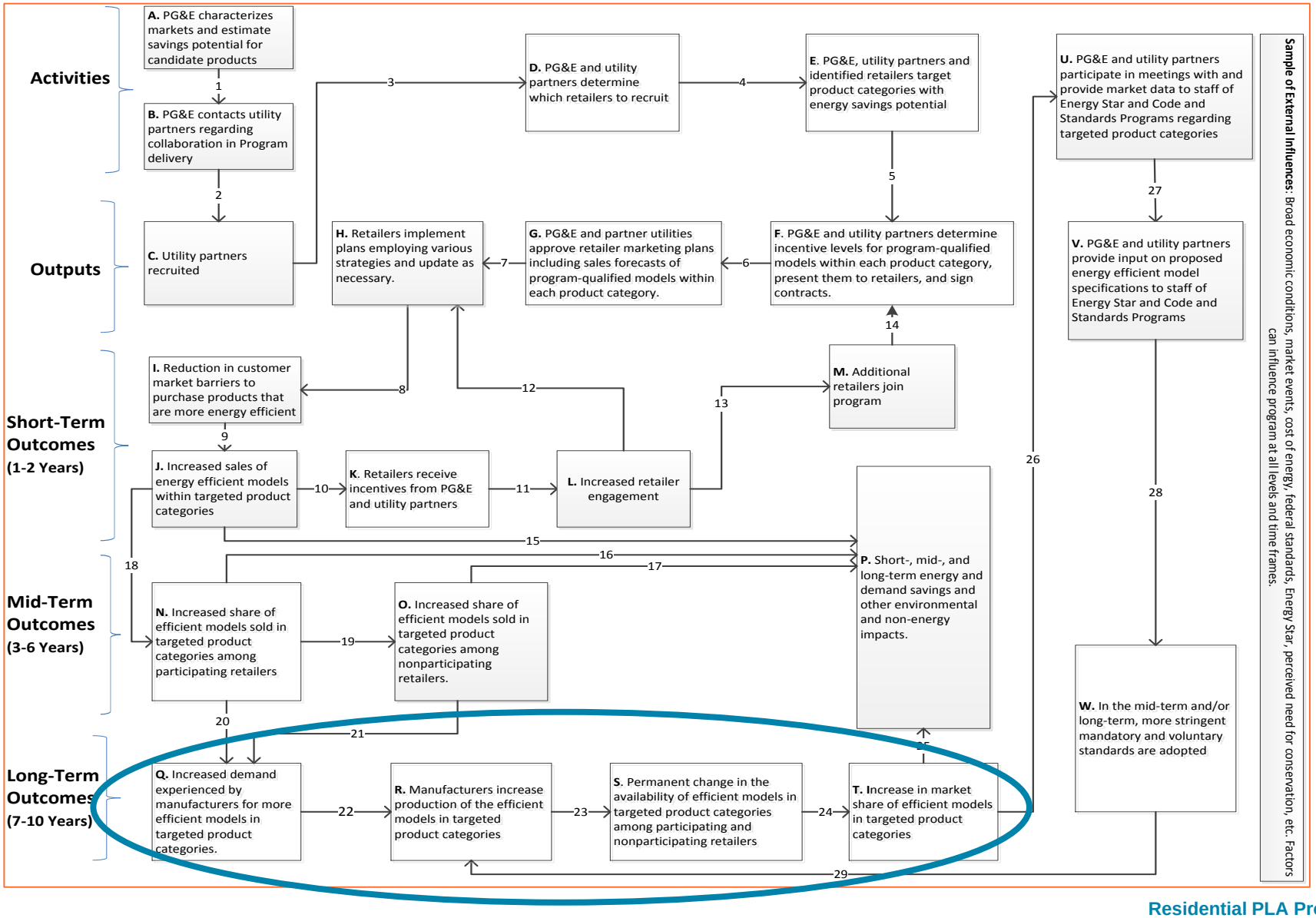
For this evaluation, theory-driven evaluation denotes:

"... any evaluation strategy or approach that explicitly integrates and uses stakeholder, social science, some combination of, or other types of theories in conceptualizing, designing, conducting, interpreting, and applying an evaluation (Coryn et al., 2011)."

Principles of a theory-driven evaluation

- Theory-driven evaluators should formulate a plausible program theory
- Theory-driven evaluators should formulate and prioritize evaluation questions around a program theory
- Program theory should be used to guide planning, design, and execution of the evaluation under consideration of relevant contingencies
- Theory-driven evaluators should measure constructs postulated in program theory
- Theory-driven evaluators should identify breakdowns, side effects, determine program effectiveness (or efficacy), and explain cause-and-effect associations between theoretical constructs

Market Transformation and RPP



Evaluation Approach: Level of Rigor

- Measures expected to represent 2% or more of total IOU portfolio savings referred to as a high impact measures (HIMs)
- Evaluations aimed at estimating savings for HIMs, due to relative importance to portfolio, are typically designed at higher level of methodological rigor
- Achieving HIM status in 2015 is unlikely since there are only four retailers and five product categories
- As more retailers engage with the Program and number of product categories is expanded, the chances of achieving HIM status in the future increase
- Evaluation plan for the RPP Program set at the enhanced level of rigor, representing a “no-regrets” evaluation design

Analysis Methods: Quasi-Experimental

- **True experimental design is most rigorous of all research designs**
- **Would involve random assignment of stores across the state to participate in RPP, which is NOT POSSIBLE because retailers self-select into the program**
- **Variety of *quasi*-experimental designs are proposed that are able to control for the effects of history:**
  - Using a non-equivalent control groups (comparison groups)
    - Non-participating retailers
    - Participating retailer stores outside of CA
    - Other points of comparison such as commercially available market-level data
  - Using pre-post participant data only

Analysis Methods: Non-Experimental

- In 1993, recognition that methods involving comparison groups were not always feasible was first formalized in energy efficiency field in CA in the *Protocols and Procedures for the Verification of Costs, Benefits, and Shareholder Earnings from Demand-Side Management Programs*
- For budgetary, timing, statistical, and research design issues, quasi-experimental designs and analyses are sometimes supplemented with non-experimental approaches such as self-report
- Self-report can include both quantitative and qualitative information
- Self-report can include surveys and in-depth interviews
- Self-report approaches have widespread acceptance as tools for assessing net-to-gross ratios

Self-Report

- **Various market actors will be asked to report their assessments of changes in program-qualified sales as attributable to the program including:**
  - Retailer purchasing staff
  - Retailer marketing staff
  - Retailer store staff
  - Manufacturers
  - Codes & standards representatives
- **Actors will also be probed in terms of:**
  - Likely behaviors absent the Program intervention
  - The influence of the Program on their behaviors
  - The extent to which the Program was successfully implemented
- **Program staff documentation of key initiatives**
- **Results will not be used on their own to make any claims of attribution, but as other evidence to support a comprehensive assessment within the theory-driven evaluation framework**

Theory-Driven Evaluation: Bringing It All Together

- The results of evaluation efforts will be combined.
- If the predicted steps between program activities, outputs and outcomes can be confirmed in implementation, then this matching of the theory to observed outcomes lends a strong argument for causality.
- At various points throughout the life of the RPP Program, the preponderance of evidence will be systematically examined to assess the extent to which the program is on track to achieve its ultimate objectives.
- To be credible, the preponderance of evidence approach must be rule-based, transparent, and repeatable.
- For each causal hypothesis, an objective case for causality must be constructed based on all quantitative and qualitative data analyzed.

Theory-Driven Evaluation: Bringing It All Together (cont.)

- Each case must include, at a minimum, a summary of the data collected and analysis approach and results, an assessment of the validity and reliability of the data, and a discussion of the level of uncertainty surrounding the conclusions.
- Once the evidence for each hypothesis is presented, an assessment should be conducted as to whether all of the evidence warrants the conclusion that the entire network of cause-and-effect relationships depicted in the logic model have combined to achieve the ultimate objectives of the RPP Program.
- This assessment must be organized around a set of rules about how to combine these data into a coherent, internally consistent story about the efficacy of the program.

Questions and Discussion

Appendix

Evaluation Approach: Error Minimization

Total Error

Total Survey Error

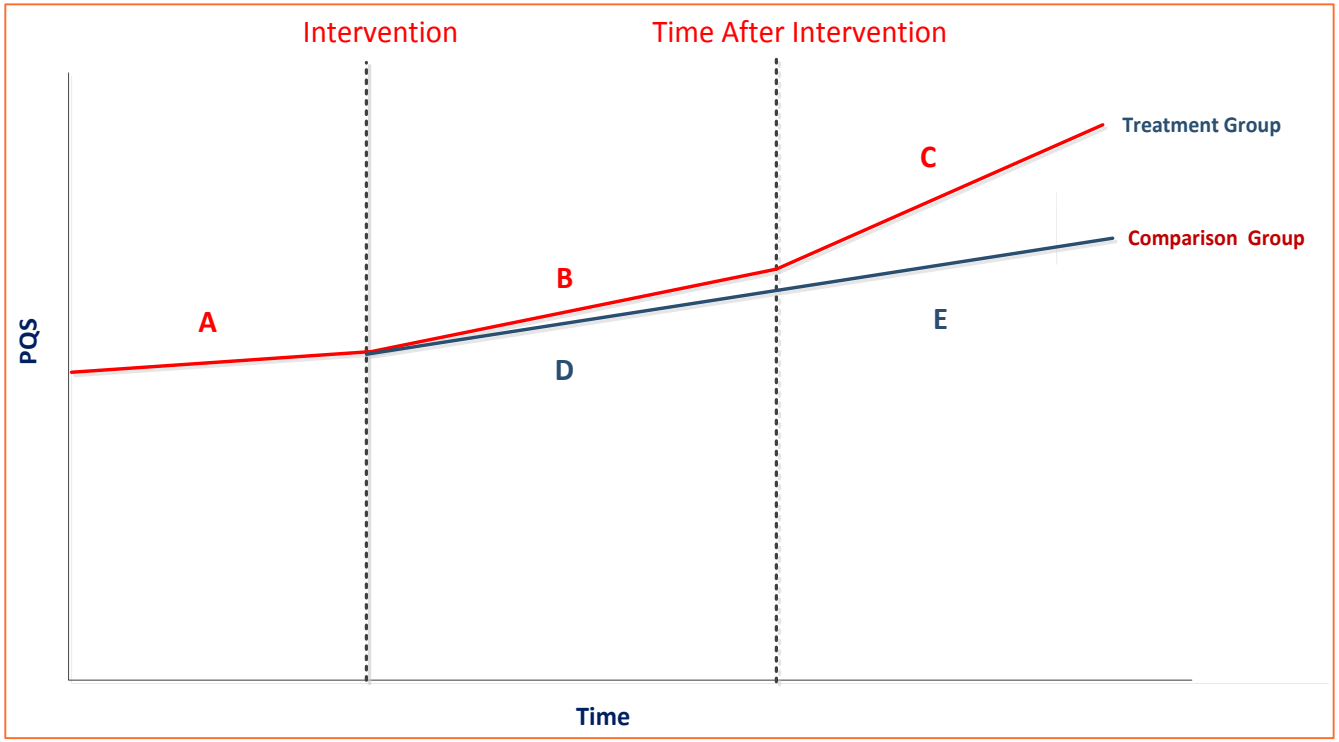
Sampling

Non-Sampling
- Measurement errors
- Non-response errors
- Frame errors
- Processing errors

Non-Survey Error
- Modeler error (statistical and engineering)
- Low Internal and/or external validity
- Choosing an inappropriate baseline
- Low statistical conclusion validity

Comparison Group, Option 1

- **Considered using non-participating CA retailers as comparison group**
  - Deemed infeasible since they are very likely unwilling to provide the detailed sales data necessary for our analysis
- **Exploring use of participating retailer stores outside of CA as a comparison group**
  - Will do retailer-level segmented regression analysis



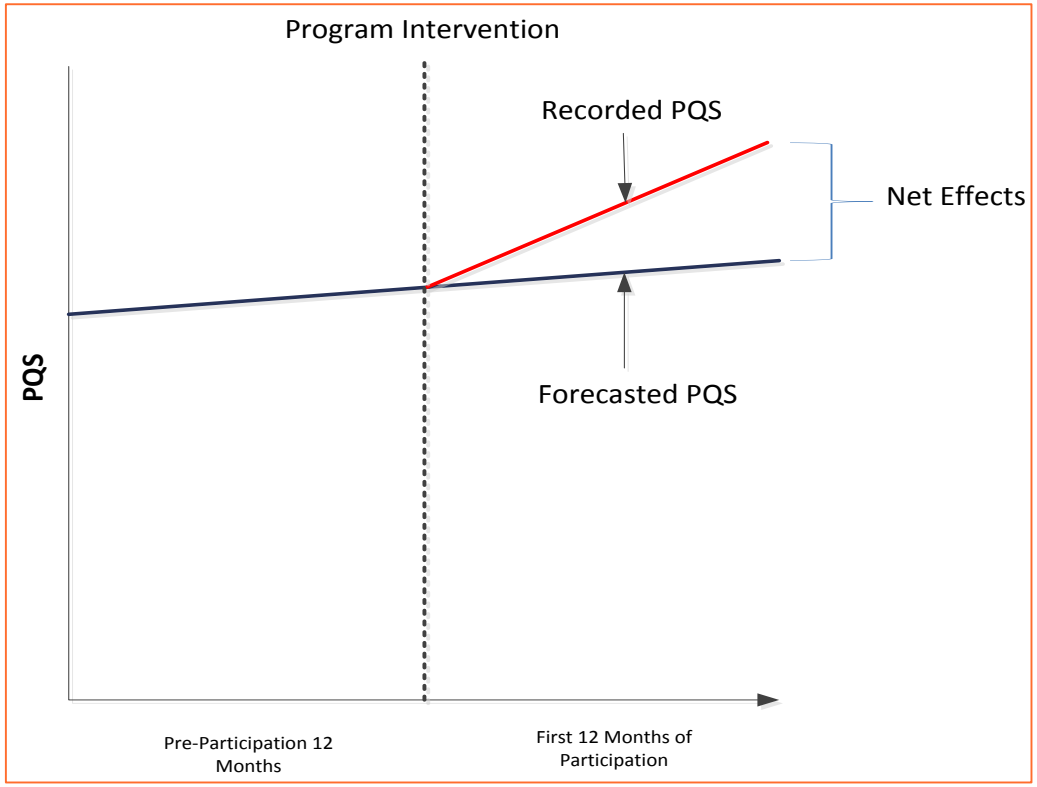
Comparison Group, Option 2

If suitable comparison group cannot be formed, other points of comparison such as commercially available market-level data for relevant metrics such as sales of program-qualified products or the PQS will be explored:

- NPD quarterly market share data
- Association of Home Appliance Manufacturers (AHAM) quarterly market share data

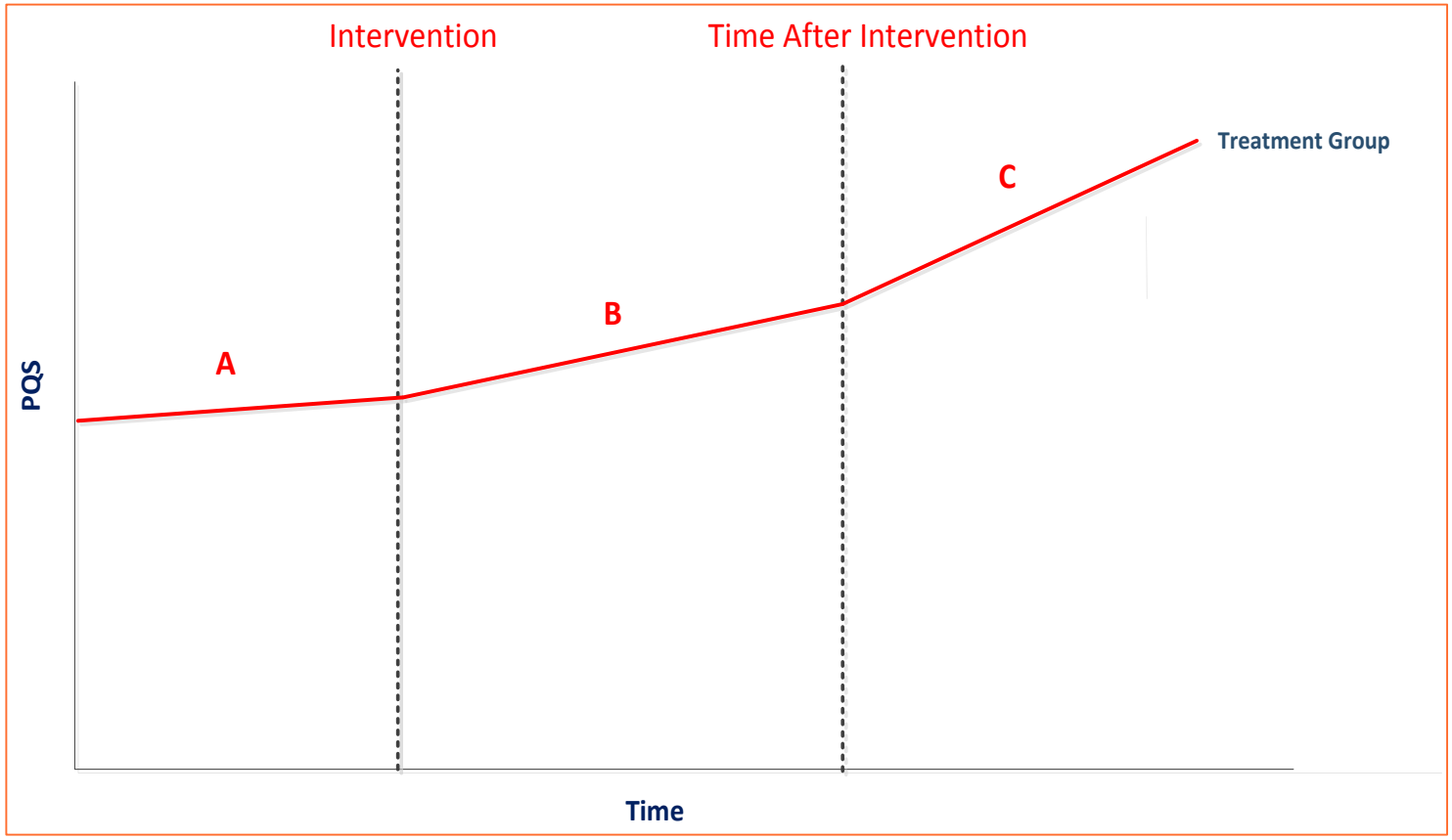
Participant Only, Option 1

- Use 12-month historical data for participating retailers to forecast a counterfactual PQS and sales volume baseline
- Forecast compared to actual program-period sales data
- Difference between the program-period data and the forecasted baseline is net effect of program



Participant Only, Option 2

- Interrupted time-series (segmented or piecewise) regression of pre-post participant sales data
- Determine if and how slopes in different time segments differ



Participant Only, Option 1 Equations

PQS/sales volume baseline forecast developed using statistical models for each product subcategory:

$Yt = \beta 0 + \beta 1 Timet + \varepsilon t$

where:

- Yt = PQS/sales volume in month t
- $Time_t$ = The number of months from the start of the series
- B0 = The intercept
- B1 = Change in Y_t for a one month change in time

Could also use an exponential smoothing technique, a procedure for continually revising a forecast in light of more recent experience:

FT+1=AXT+(1-A)FT

where:

- FT+1= Forecasted value for the next period
- A = Smoothing constant ($0 \leq A \leq 1$)
- XT= Actual value of series in period T
- FT= Average experience of series smoothed to period T or forecasted value for last period

Participant Only, Option 2 Equations

General form of segmented regression model to be used is:

$Y_t=\beta_0+\beta_1 Time_t+\beta_2 Intervention_t+\beta_3 Time\_After\_Intervention_t+\varepsilon_t$

where:

- Y_t = PQS/sales volume in month t
- $Time_t$ = Indicates the number of months from the start of the series
- $Intervention_t$ = A dummy variable taking the values 0 in the pre-intervention segment and 1 in the post-ntervention segment
- $Time\_After\_Intervention_t$ = 0 in the pre-intervention segment and counts the months in the post-intervention segment at time t
- β_0 = Estimates the base level of the outcome at the beginning of the series
- β_1 = Estimates the base trend, i.e. the change in outcome per month in the pre-intervention segment
- β_2 = Estimates the change in level in the post-intervention segment
- β_3 = Estimates the change in trend in the post-intervention segment
- ε_t = Estimates the error

Once models are estimated, they are evaluated without and with program:

Without Intervention:

$Y_t=\beta_0+\beta_1 Time_t$

With Intervention:

$Y_t=\beta_0+\beta_1 Time_t+\beta_2 Intervention_t+\beta_3 Time\_After\_Intervention_t$

Participant Only, Option 2 Equations

The difference between these equations is the following, which represents net reduction in PQS/sales volume:

$$Y_t = \beta_2 Intervention_t + \beta_3 Time\_After\_Intervention_t$$

Basic model can be extended to add more time variables to assess incremental change at various time points after program implementation