

# OPA EVALUATION TA



## Pre-Post Outcome Study How to Guide

### What is a pre-post outcome study?

A pre-post outcome study quantifies how participants' outcomes change over the course of a study. For example, a pre-post outcome study can document the change in outcomes between a period before programming (pre) and a follow-up period after programming (post). Researchers can aggregate these differences across many people to represent how outcomes changed on average, which can help inform the planning for an impact study.

*A pre-post outcome study can build a foundation for a future impact evaluation.*

This **How to Guide** is intended for a research audience. It provides detailed instructions on how to prepare data and conduct analyses that support a pre-post outcome study.

### What types of evidence or information does a pre-post outcome study provide?

A high-quality pre-post outcome study provides evidence about how the outcomes of youth who experienced a program changed over time. For example, a youth who participated in a teen pregnancy prevention program might show improvement in their knowledge of sexually transmitted infections (STIs) by the end of the program relative to the start of the program.

### What types of evidence or information does a pre-post outcome study not provide?

A high-quality pre-post outcome study cannot provide evidence that the program itself caused the changed outcomes. A key limitation of a pre-post outcome study is that it lacks a counterfactual, or a comparison group. As a result, we cannot know for certain how outcomes might have changed in the absence of the program. In many cases, we would expect improvement in participants' outcomes even without an intervention. For example, youth's knowledge of STIs might improve over time as a function of natural maturation and exposure to information outside of the program. Therefore, we cannot assume that favorable changes in participants' outcomes we see in a pre-post study are caused solely by the program.



## How can findings from a pre-post study be presented?

Suppose researchers found that program participants improved their knowledge of STIs by 20 percentage points since the start of the program. Participants' knowledge might have improved, however, because of natural maturation and regular education. The researchers cannot assume the program caused favorable changes. The study report might say the following:

*“Between program entry and exit, participant knowledge scores improved by 20 percentage points. This analysis assesses individual change over time without a counterfactual—it is not appropriate to assert that the program was solely responsible for the observed improvement in outcomes.”*

## How can a pre-post study build evidence and help inform a future impact evaluation?

A high quality pre-post outcome study can build evidence and inform a future impact evaluation in two main ways.

1. First, a pre-post outcome study can help **validate a program's logic model**. Most outcomes shown in a program's logic model can be expected to change or improve during a typical pre-post outcome study. Improvement in key outcomes shown in the program logic model can be considered preliminary evidence suggesting that a program might be ready for an impact evaluation because the program has shown some promise in having key outcomes improve. In addition, the proximal outcomes that are well-aligned with the intervention are likely to show greater change or improvement relative to distal outcomes. If a pre-post outcome study demonstrates relatively larger changes in proximal outcomes than distal outcomes, this helps to validate the information in a program's logic model and builds the argument that the program is ready for a future impact evaluation.
2. Second, a pre-post outcome study might **provide an optimistic, best-case estimate of a program's impact**, which can help in future impact evaluation planning. For example, assume a pre-post study shows that participants' knowledge increased by 0.25 standard deviation units. Researchers could consider this 0.25 standard deviation improvement to be an optimistic, best-case

target for a possible impact study; this is because many outcomes will naturally improve in the comparison group (and thus, the impact measured across conditions will be less than 0.25 standard deviations).

## What outcomes should I look at in a pre-post study?

Choose outcomes informed by the program's logic model and the study's duration. Ideally, the study includes shorter-term and longer-term outcomes to track key components of the logic model. It is also preferable to select outcomes that are measured in other pre-post outcome studies to help interpret findings in context of the wider body of research.

In general, outcome measures that would stay relatively unchanged without the program lead to findings that are easier to interpret as evidence for the program. Examples of this are shorter-term outcomes that are tied to the program logic, such as knowledge and attitudes toward topics specifically covered by the program.

An example of an outcome measure that is harder to interpret as evidence for the program is sexual initiation. Because it is defined cumulatively, sexual initiation rates will increase as youth age. A pre-post outcome study over a long period would likely show that sexual initiation has increased substantially whether or not the program actually changed behavior.

## How should I estimate pre-post differences?

A rigorous approach to measuring the pre-post change for each outcome starts with a complete-case analysis, which calculates the average change for individuals (cases) with complete pre and post outcome data. If there are multiple outcomes, complete-case analyses must be done separately for each outcome.

In a statistical software package, the dataset would be structured so that each row is one individual and that two variables (columns) contain the measure at pre and post. If these variables were named *pre* and *post*, then the paired *t*-test would be performed in Stata® with the command “`ttest pre == post`” or in SAS® with the “`proc ttest`” procedure and “`paired`” statement.

The steps for complete-case analysis and statistical testing for the pre-post change are listed below.

1. For each individual, match the pre and post values and then take the difference. People who are missing a pre or post value (or both) will be missing a difference.
2. Calculate the pre-post change as the average of the individual differences.
3. Use a paired *t*-test to determine whether this pre-post change is statistically significant. This test takes into account that the pre and post data are from the same set of individuals.

We recommend this approach over the alternative of taking the difference of the averages of all available data observed at pre and post, which allows different groups of people to contribute to the pre and post averages.

### What should I report for each pre-post change in outcome?

For each outcome, it is good practice to report the pre and post means and standard deviations, differences in pre and post means, *p*-values from appropriate statistical tests, and sample sizes. An example table for a complete-case analysis can be found below.

Presenting the standardized pre-post difference could also help readers who want to compare findings across outcomes or studies. Standardization involves dividing the difference by the outcome's post standard deviation.

### What subgroups can I examine?

Examining changes in outcomes for subgroups of study participants can provide additional information about the logic model and sample considerations for a future impact evaluation. Consider the program model, its implementation, or related research when selecting subgroups. A pre-post outcome study could include subgroups based on (if applicable): demographic characteristics such as sex, race, and age; features of the study sites such as urbanicity or program enrollment cohort; or features of individuals' program participation such as dosage or completion.

### How can I know whether the findings from the study sample also apply to the study population, and is calculating response rates or using nonresponse weights necessary?

If pre and post data are not available for all participants, then there are valid concerns about how much the observed pre-post change results apply to the full sample. You cannot fully address these concerns, but you can do the following three analyses separately or in combination to respond to these concerns.

- **Response rate analysis.** If the rates of participants with data at both pre and post (complete-case data) are high, the data are more likely to be representative of the full sample. A response rate analysis measures the rates of individuals in four categories based on whether they have data at (1) both pre and post, (2) only pre, (3) only post, or (4) neither pre nor post. Report the prevalence of these categories for each outcome. An example table can be found below.

#### Example table for reporting complete-case findings from a pre-post outcome study:

Outcome	Pre		Post		Difference	<i>p</i> -value	Sample size
	Mean	Standard deviation	Mean	Standard deviation			
Measure 1	#	#	#	#	#	#	#
Measure 2	#	#	#	#	#	#	#

#### Example table for reporting a response rate analysis:

Outcome	Data at both pre and post	Data at pre only	Data at post only	No data (pre or post)
Measure 1	%	%	%	%
Measure 2	%	%	%	%

- **Nonresponse analysis.** The differences between participants with complete-case data and those with only pre data can suggest whether the pre-post change is likely an underestimate or overestimate of change for the full sample. For a nonresponse analysis, identify baseline factors that could predict whether a participant has complete-case data, such as the baseline outcome measure, demographics, and site characteristics if appropriate. Then, create a binary indicator variable defining whether an individual has complete-case data. Regress this variable on the identified factors using logit or probit regression and clustering standard errors as appropriate. Notably, consider clustering standard errors if predictor variables are measured at a cluster level (for example, site-level characteristics). Report regression coefficients, standard errors, transformed or standardized estimates if they are easier for readers to interpret, and *p*-values from appropriate statistical tests. For example, exponentiated coefficients from logistic models are easier for many readers to interpret than the coefficients themselves because exponentiation gives an odds ratio. Standardized coefficients might be easier for readers to interpret for factors that do not have an intuitive scale, such as raw test scores. An example table can be found below.
- **Use nonresponse weights in the pre-post analysis.** Findings that apply to the full sample of program participants can be more policy relevant than findings that apply to the complete-case sample. For example, policymakers might be more interested in the change in outcomes for everyone who enrolled in a sexual education course rather than the subset who completed the program and had complete pre-post data. If you've completed the nonresponse analysis we described previously, then you can weight individuals in the complete-case analysis to better represent the full sample. Those who were less likely to be in the complete-case sample would each receive a higher nonresponse weight because they represent more people who did not provide complete

Statistical packages often have ways to change the presentation of coefficients. For the example of logistic models, Stata®'s "logistic" regression command displays odds ratios, and the "logit" regression command displays the raw coefficients. In SAS®, the "proc logistic" procedure provides raw coefficients, and the "stb" option adds standardized coefficients representing the predicted change from a 1-standard deviation increase in the predictive factors, holding all other factors fixed.

data, and vice versa. The following is an approach to applying nonresponse weights:

1. **Calculate raw probabilities of being a complete case sample member.** Using the same logit or probit model estimated for the nonresponse analysis, calculate the predicted values of the dependent variable for every sample member. Save that as a new variable: *p\_CC*.
2. **Calculate raw nonresponse weights.** Take the inverse or reciprocal of the predicted values, so a predicted value of 0.8 would become  $1/0.8 = 1.25$ . Save that as a new variable: *weight\_raw* =  $1/p_{CC}$ .
3. **Calculate a rescaling variable.** Divide the complete-case sample size by the sum of the raw nonresponse weights for complete-case sample members from step 2. Save that as a new variable: *rescale*.
4. **Calculate final nonresponse weights.** Multiply each raw nonresponse weight from step 2 (*weight\_raw*) by the scaling variable from step 3 (*rescale*). The results are nonresponse weights for complete-case sample members that sum to the complete-case sample size. Save this as a new variable: *weight\_rescaled*.
5. **Use the weights in all pre-post outcome analyses.** Incorporate the weights when estimating any descriptive and inferential statistics to provide findings for the full study sample.

**Example table for reporting a nonresponse analysis:**

Factors in the regression model	Coefficient	Standard error	Odds Ratio	Standardized coefficient	<i>p</i> -value
Factor 1	#	#	#	#	#
Factor 2	#	#	#	#	#

You can show the benefits of the nonresponse weights by reporting means of variables at pre and post for the full sample and the complete-case sample, with and without weights. The variables can include outcome measures and other characteristics such as demographics and site characteristics (if applicable).

The weighted complete-case sample will typically have means that are closer to the full sample means than the unweighted complete-case sample had. This supports the argument that the weights have made the complete-case sample more similar to the full sample and offers validation of the utility of the weights for improving the representativeness of the findings. An example table comparing the sample means can be found below.

## How should I interpret the pre-post findings after incorporating nonresponse weights?

The same statistics should be reported from the pre-post analysis after using the nonresponse weights. Each statistic (mean, prevalence rate, standard deviation, *p*-value, and so on) will be different after the use of the nonresponse weights relative to the analysis that treated all complete-case sample members as contributing equal weight.

Whereas the original unweighted statistics provide information on pre-post changes among the complete case sample, the new, weighted statistics provide updated information for the full sample and thus might represent a

In Stata®, you could perform the process for steps 1 to 4 with the following commands for a logit regression where *factors* is the list of baseline variables that could predict nonresponse and everyone has baseline data:

```
. generate complete_case = 1 if !missing(post)
. replace complete_case = 0 if missing(post)
. logit complete_case factors
. predict p_CC
. generate weight_raw = 1/p_CC
. summarize complete_case
. generate rescale_num = r(mean)*r(N)
. summarize weight_raw if complete_case == 1
. generate rescale_den = r(mean)*r(N)
. generate rescale = rescale_num / rescale_den
. generate weight_rescaled = weight_raw * rescale
```

For step 5, the weighted pre-post change could be calculated as the following:

```
. generate change = post - pre
. mean change [pweight = weight_rescaled]
```

The appendix contains a more detailed description of this process in SAS®.

more policy-relevant set of results. As a result, the weighted data findings should be the benchmark findings to be reported when describing pre-post outcome changes.

### Example table for reporting variable means with and without nonresponse weights:

Variable	Pre Means			Post Means		
	Full sample	Weighted complete-case sample	Complete-case sample	Full sample	Weighted complete-case sample	Complete-case sample
Outcome 1	#	#	#	#	#	#
Outcome 2	#	#	#	#	#	#
Factor 1	#	#	#	#	#	#
Factor 2	#	#	#	#	#	#

#### Pre-post findings:

- ✓ Provide value in describing how participants' outcomes change over time.
- ✗ Do not demonstrate the impact of a program.
- ✓ Can serve to establish a foundation for a future impact study.

## Best practices for pre-post outcome analyses

✔ **Don't be satisfied by only doing a basic complete-case pre-post analysis.** The youth from whom you have pre- and post-test data might differ in important ways from the full sample of youth that were offered the program.

### Plan on supplementing your pre-post analyses with the following:

- ✔
- Response rate calculations
  - Nonresponse analysis
  - Estimation of nonresponse weights and use of weights in pre-post analyses
  - Justification that the nonresponse rates improved the representativeness of the findings

## Appendix: SAS code for example of estimating and using nonresponse weights

In this example, we are interested in measuring change in knowledge of sexually transmitted infections from pre (*STI\_knowledge\_pre*) to post (*STI\_knowledge\_post*) using a data set called "mydata." This example is also described in the OPA webinar on pre-post outcome studies.

StudyID	Male	Hispanic	Age	GRIT	STI_Knowledge_Pre	STI_Knowledge_Post
101	1	0	15.5	8	78	79
102	1	0	15.7	7	77	82
103	1	1	16.1	8	45	.
104	0	1	15.8	9	56	54
105	0	1	15.9	10	65	67
106	1	1	16	8	91	.
107	0	0	16.1	8	25	52

The code that follows creates the complete-case sample indicator *STI\_CC* and predicts probabilities for *STI\_CC* based on sex, ethnicity, age, baseline knowledge of sexually transmitted infections, and a baseline measure of grit.

Data mydata;

```
set mydata;  
STI_CC = 1;  
if cmiss(STI_knowledge_post) = 1 then STI_CC = 0;
```

Run;

Proc logistic data=mydata;

```
model STI_CC (event = '1') = male hispanic age STI_knowledge_pre GRIT / link=logit stb;  
output out=psdata predicted=P_STI_CC;
```

Run;

Data mydata;

```
set psdata;  
if STI_CC = 1 then STI_weight_raw = 1/P_STI_CC;
```

Run;

```
Proc sql;
  create table mydata_weights as select
    StudyID,
    STI_CC,
    P_STI_CC,
    STI_weight_raw,
    sum(STI_CC)/sum(STI_weight_raw*STI_CC) as rescale,
    (sum(STI_CC)/sum(STI_weight_raw*STI_CC))*STI_weight_raw as STI_weight_rescaled
  from mydata;
quit;
```

```
Proc print data=mydata_weights;
  sum STI_CC STI_weight_raw STI_weight_rescaled;
Run;
```

The final output produced by the last “proc print” statement follows:

Obs	StudyID	STI_CC	P_STI_CC	STI_weight_raw	rescale	STI_weight_rescaled
1	101	1	0.67	1.49	0.72	1.08
2	102	1	0.87	1.15	0.72	0.83
3	103	0	0.42	.	0.72	.
4	104	1	0.89	1.12	0.72	0.81
5	105	1	0.91	1.10	0.72	0.80
6	106	0	0.51	.	0.72	.
7	107	1	0.49	2.04	0.72	1.48
		5		6.91		5

## Suggested citation

This report is in the public domain. Permission to reproduce is not necessary. Suggested citation: Lee, J., and Cole, R. (2020). *Pre-Post Outcome Study How to Guide*, Washington, DC: Office of Population Affairs, Office of the Assistant Secretary for Health, U.S. Department of Health and Human Services.

