Mathematica
Progress Together

Pre-post outcome analyses

Building a foundation for a future impact evaluation

Webinar presented to TPP18 grantees 2/28/2020

Russell Cole, Ph.D.

Agenda

- **The evidence from a pre-post outcome study: a foundation for an impact evaluation**
- **Estimating and reporting pre-post differences: the basics**
- **Additional analyses to enhance pre-post findings**

Evidence from a pre-post outcome study

Mathematica

What is a pre-post outcome study?

- **A study that quantifies how participants' outcomes change over the course of a study**
  - Often, comparing how participants' outcomes change between program entry (baseline/pre) to program exit (follow-up/post)
- **Difference in outcomes from baseline/pre to follow-up/post represents individual change**
- **Aggregating this difference across all program participants quantifies how outcomes changed on average**
- ***Note: We will primarily focus on change analyses for two assessment points today***

Mathematica

How to interpret average pre-post outcome change?

- **Average change = how individual outcomes among program participants changed over time on average**
- **Average change ≠ the impact or effect of the program**
  - Without a counterfactual, we cannot disentangle change in outcomes caused by the program from naturally occurring outcome changes (for example, maturation, testing, or regression; Campbell and Stanley 1975)
- **Be sure to describe the findings (and their limitations) appropriately!**
  - "Between program entry and program exit, participant knowledge scores improved by 30 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."

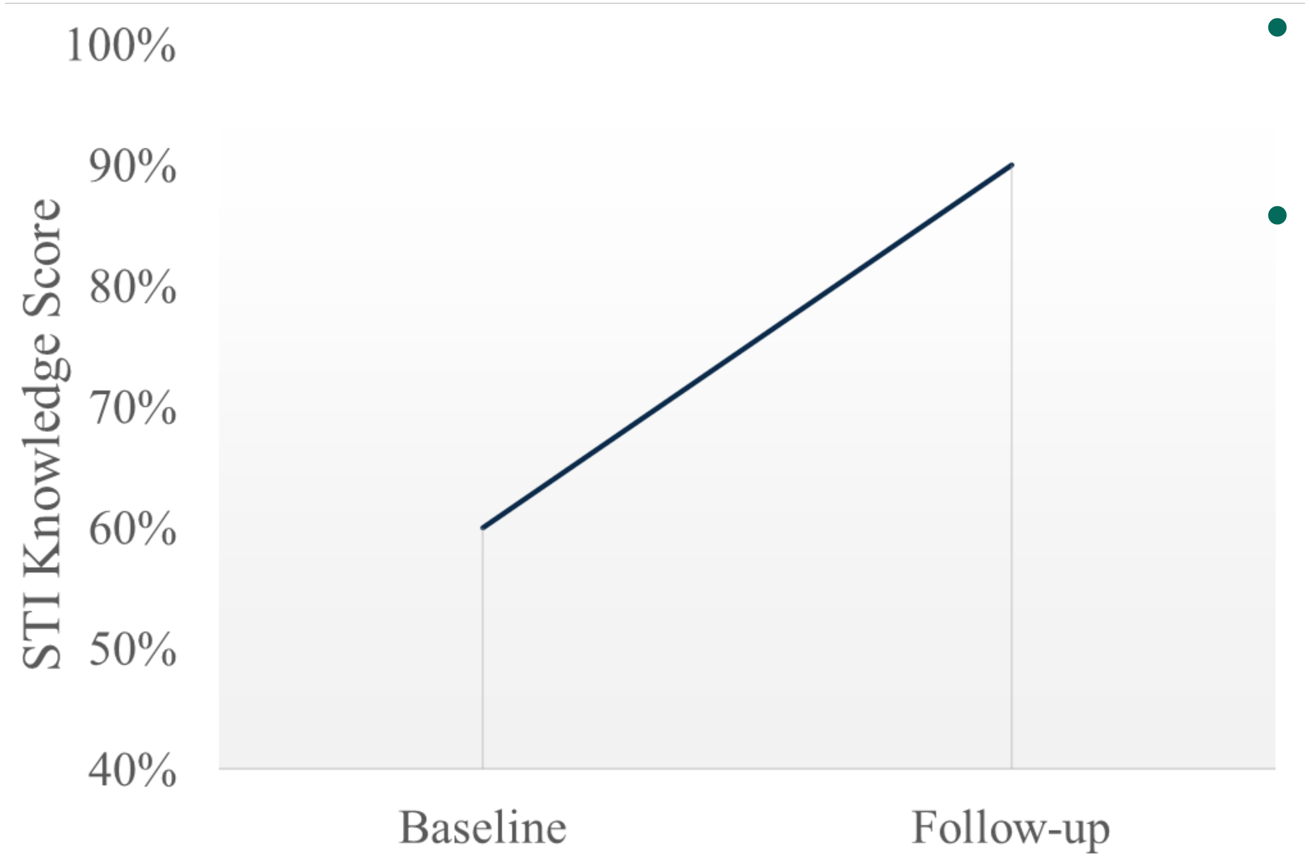
Mathematica

How can pre-post findings create the foundation for an impact study?

- **Are outcomes moving in the right direction?**
  - Logic model for an intervention presents a hypothesis about how participants' outcomes might change
  - Pre-post outcome findings provide data to test the hypothesis
  - At a minimum, proximal outcomes in the logic model should change over time!
- **Are the changes in outcomes *large*?**
  - In many cases, we might expect natural change in outcomes even in the absence of the program
  - Changes in outcomes might represent an upper-bound estimate of potential program impact for a subsequent impact evaluation
    - This information can inform statistical power analyses

Mathematica

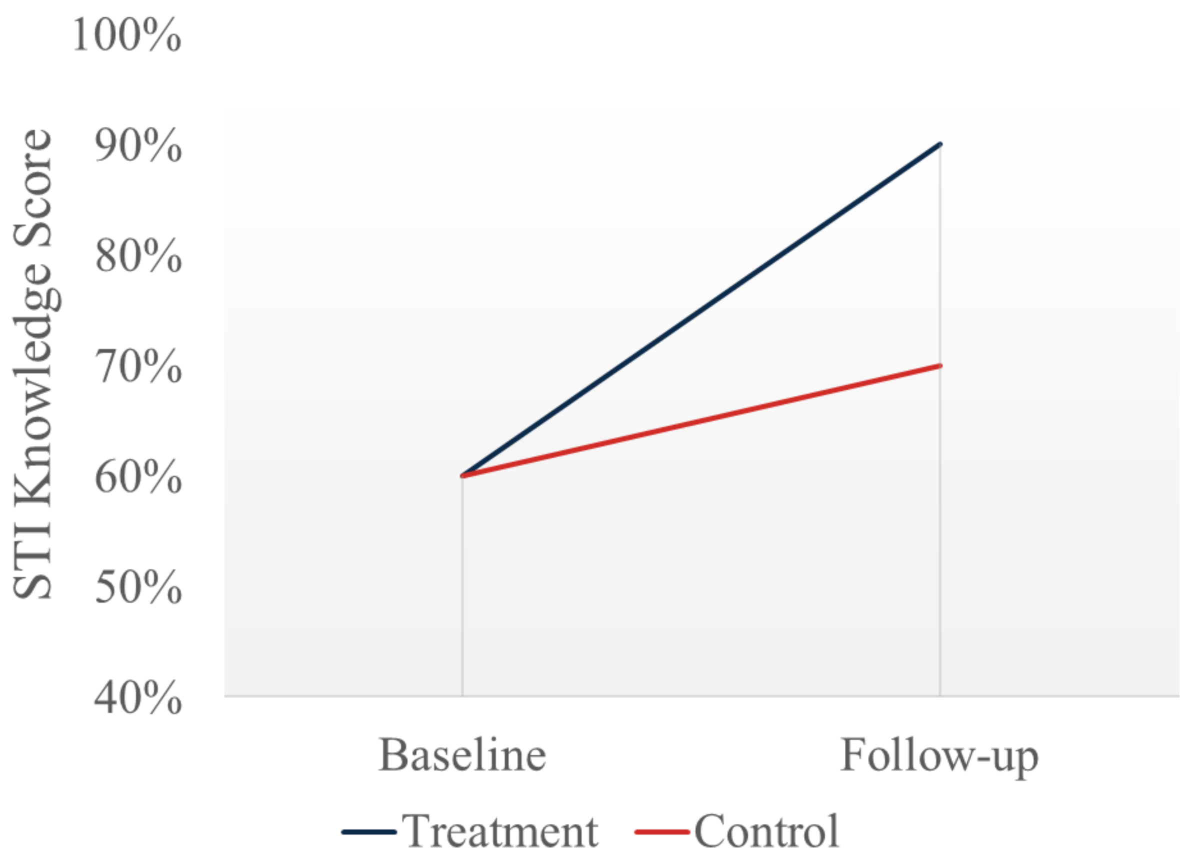
Example: Program participants improved sexually transmitted infection (STI) knowledge by 30 percentage points (30PP)!



- **Shows improvement in key outcome**
- **Provides upper-bound estimate for future impact study**
  - ❑ Don't expect to see program impacts of 30PP in a future impact evaluation…

Mathematica

Use pre-post impact estimate as upper bound for statistical power calculations



- **Natural improvement among comparison group suggests an impact evaluation will observe a smaller effect …**
  - ❑ Make sure you have sufficient power to detect impacts *smaller* than those observed in the pre-post analysis

Mathematica

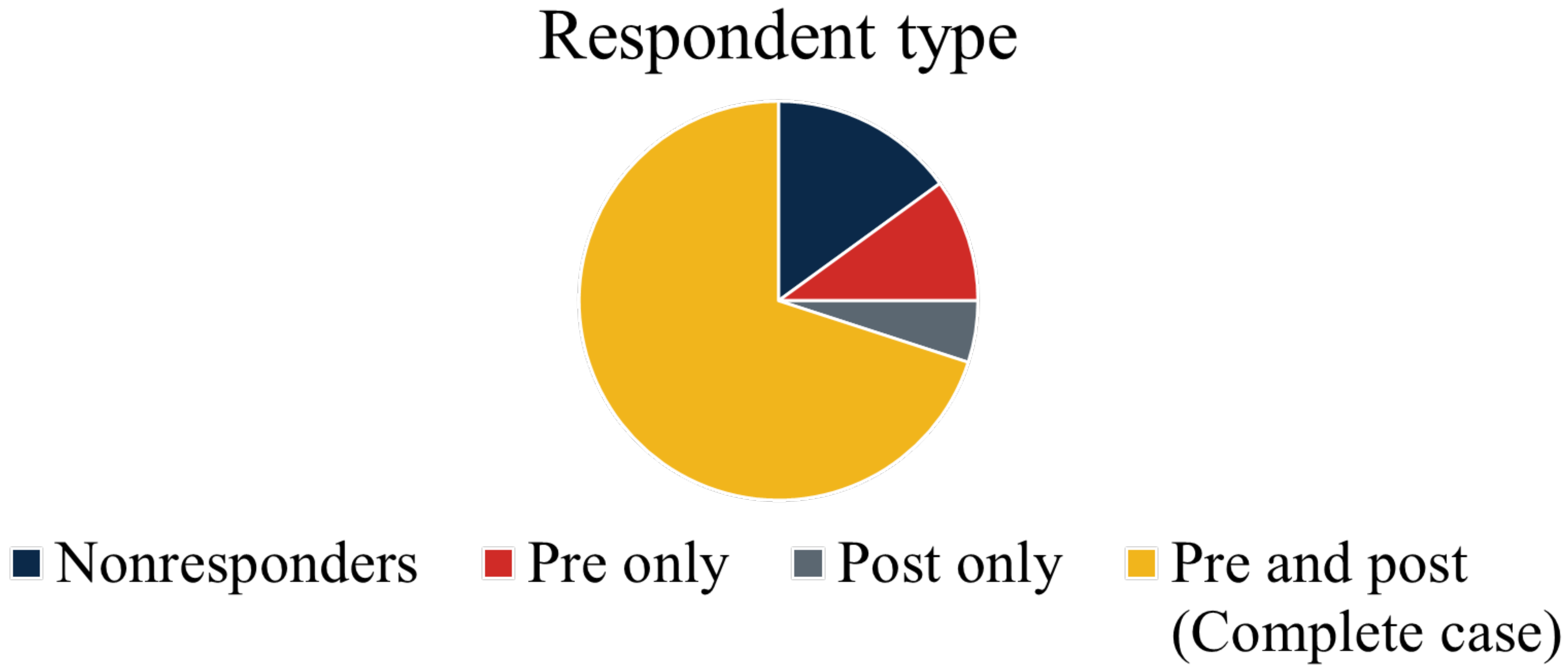
Estimating and reporting pre-post differences: The basics

Mathematica

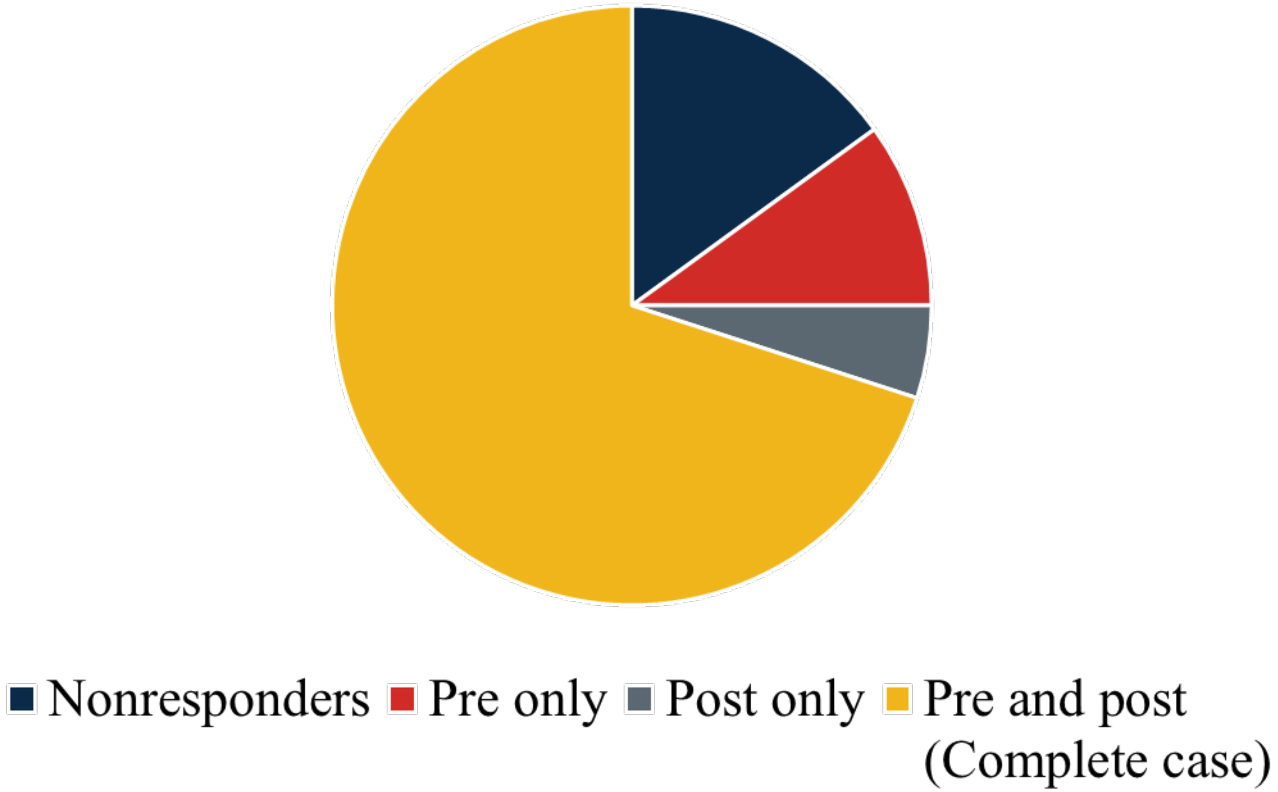
Goal: Describe how individual outcomes change over time

- **Choose the right types of outcomes**
- **Plan on conducting within-individual analyses**
  - Match pre and post outcomes for each individual
  - Eliminate individuals from analysis who are missing one or both assessments
  - Do this separately for *each outcome of interest*
- **Benefit of this approach**
  - Ease of interpretation: it eliminates compositional differences (biases) that can occur if analysis conducted with all available data
- **Limitation of this approach**

Different types of respondents observed in a pre-post study



Unpacking issue of composition as a source of bias in understanding individual change



Respondent type (prevalence rate)	Average score at baseline	Average score at follow-up
Nonresponders (15%)	-	-
Pre only (10%)	70	-
Post only (5%)	-	75
Pre and post (complete case = 70%)	60	80
Average of observed data	61.3	79.7

Mathematica

What to report among the complete case sample?

- **Pre and post means and standard deviations**
- **Difference in means**
  - In raw units
  - In standard deviation units (relative to the post-test period)
- ***p*-value of the difference**
  - Based on a paired t-test (or nonparametric analogue)

Explore and report heterogeneity in outcome change

- **Conduct analyses for proximal and distal outcomes in logic model**
  - Larger improvements in proximal outcomes (relative to distal outcomes) validates features of program logic model
- **Conduct analyses by subgroups of interest**
- **Conduct analyses by whether participant received the program as intended or not**
  - Potentially supplement the within-group analysis with a between-group analyses

Additional analyses to enhance pre-post findings

Mathematica

Skeptical readers will be unsatisfied with the basic presentation

- **Primary concern: The complete case sample does not adequately represent the full study sample**
- **We can address this limitation with additional analyses**
  - Response rate analyses
  - Nonresponse analyses
  - Use of nonresponse weights
  - Demonstrate value of nonresponse weights

Step 1: Response rate analysis

Response rate analysis

- **For each outcome of interest, categorize each individual as one of four types**
  - Nonresponder at both assessments
  - Pre only
  - Post only
  - Pre and post (complete case)
- **Report prevalence rates of types for each outcome**

Mathematica

Respondent type prevalence rates will vary due to item non-response

Respondent type	Intentions to remain abstinent	Recent sexual behavior
Nonresponders	20%	23%
Pre only	8%	9%
Post only	3%	4%
Pre and post (complete case sample)	69%	64%

Mathematica

Step 2: Nonresponse analysis

Nonresponse analysis approach

- **Identify factors that differentiate complete case respondents from the broader target population**
- **Likely limitation: Baseline survey is the key data source for describing characteristics of broader sample**
  - We won't have data on nonrespondents to baseline survey
  - Therefore, we won't be able to determine whether the complete case sample represents these baseline nonrespondents *on variables measured at baseline*
- **For today, we are going to focus on generalizing to individuals with key baseline measures**
  - We want baseline measures of outcome of interest (we want to know whether complete-case sample is a high-low risk group as defined by the baseline assessment)

Approach for assessing nonresponse bias (Step 1)

- **Assemble data for a list of variables to compare across complete case sample and target sample**
  - Demographics
  - Baseline assessment of outcome of interest
  - Site characteristics (if appropriate)
  - Other baseline variables in your data set that theory or literature suggests might predict survey response (for example, motivation, grit, persistence)
- **Create an indicator variable for whether an individual is in the complete case sample for that outcome**
  - For example, for an STI knowledge outcome, use an STI_CC indicator (= 1 if in the complete case sample, 0 otherwise)

Illustrative dataset

StudyID	Male	Hispanic	Age	GRIT	STI_Knowlege_Pre	STI_Knowlege_Post	STI_CC
101	1	0	15.5	8	78	79	1
102	1	0	15.7	7	77	82	1
103	1	1	16.1	8	45		0
104	0	1	15.8	9	56	54	1
105	0	1	15.9	10	65	67	1
106	1	1	16	8	91		0
107	0	0	16.1	8	25	52	1
108	1	1	16.5			95	0

Mathematica

Approach for assessing nonresponse bias (Step 2)

- **Regress complete case indicator on predictors of interest using logit or probit regression**
  - Cluster standard errors, as appropriate, for predictors measured at cluster level
- **Report raw and standardized beta coefficients, *p*-values from analysis**
- **Summarize key takeaways**
  - “Students receiving free or reduced-price lunch (FRL) were 2.4 times less likely than non-FRL students to be included in the complete case sample.”
  - “The complete case sample tended to represent a lower-risk sample; the baseline assessment of the outcome was the single strongest predictor of whether an individual was in the complete case sample.”
  - “Complete case sample members tended to be non-Hispanic, have high levels of self-reported motivation, and attended services in schools, rather than community settings.”

Illustrative SAS Code and output

```
Proc logistic data=mydata;
Model STI_CC (event = '1') = male hispanic age
STI_knowledge_pre GRIT / link=logit stb;
Run;
```

Analysis of Maximum Likelihood Estimates

Parameter	DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq	Standardized Estimate
Intercept	1	1.1096	2.8070	0.1562	0.6926	
male	1	0.1125	0.3780	0.0885	0.7660	0.0310
hispanic	1	0.7732	0.5226	2.1892	0.1390	0.1768
age	1	-0.0256	0.1812	0.0199	0.8877	-0.0143
STI_knowledge_pre	1	-0.00345	0.00771	0.2004	0.6544	-0.0473
grit	1	0.4623	0.0923	25.0890	<.0001	0.7487

Step 3: Calculate nonresponse weights

High-level summary of Step 3

- **The results from Step 2 might indicate that we are unsatisfied with our complete case pre-post results, knowing that the complete case respondents don't adequately represent the full study sample of interest**
- **We can calculate nonresponse *weights* using the same approach from Step 2, and incorporate these weights in our complete case analyses to make the complete case sample better represent the full study sample**

Nonresponse weights

- **Nonresponse weights are the inverse of the probability of being in the complete case sample**
  - Individuals who were very likely to be in the complete case sample and were a complete case sample member have a small nonresponse weight
  - And vice versa
- **Probability of being in the complete case sample for each individual can be output from logit or probit model in Step 2**

Illustrative SAS Code and output

```
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;
```

StudyID	STI_CC	P_STI_CC
101	1	0.67
102	1	0.87
103	0	0.42
104	1	0.89
105	1	0.91
106	0	0.51
107	1	0.49

Mathematica

Calculate nonresponse weights, and rescale

- **Nonresponse weight for complete case sample members**

  = 1 / Probability of being in the complete case sample

- **Rescale weights so that sum of weights = number of complete case respondents**

StudyID	STI_CC	P_STI_CC	STI_weight_raw *(= 1 / P_STI_CC)*	STI_weight_rescaled *(= **5** * STI_weight_raw / **6.91**)*
101	1	0.67	1.49	1.08
102	1	0.87	1.15	0.83
103	0	0.42	.	.
104	1	0.89	1.12	0.81
105	1	0.91	1.10	0.80
106	0	0.51		
107	1	0.49	2.04	1.48
TOTAL	**5**		**6.91**	**5**

Mathematica

Incorporate the nonresponse weight in a revised version of the pre-post analysis

- **Calculate the same statistics reported previously:**
  - Pre and post means and standard deviations
  - Difference in means
    - In raw units
    - In standard deviation units (relative to the post-test period)
  - *p*-value of the difference
    - Based on a paired t-test
- **However, incorporate the nonresponse weights when estimating the descriptive and inferential statistics**
- **Revised pre-post results make the complete case sample better represent the target sample**

Illustrative SAS Code and output

```
proc ttest data=psdata;
Where STI_CC = 1;
paired STI_knowledge_pre * STI_knowledge_post;
run;
```

The TTEST Procedure

Difference: STI_knowledge_pre - STI_knowledge_post

N	Mean	Std Dev	Std Err	Minimum	Maximum
365	10.0247	6.9974	0.3663	-9.0000	30.0000

```
proc ttest data=psdata;
Where STI_CC = 1;
paired STI_knowledge_pre * STI_knowledge_post;
weight STI_weight_rescaled;
run;
```

The TTEST Procedure

Difference: STI_knowledge_pre - STI_knowledge_post

Weight: STI_weight_rescaled

N	Mean	Std Dev	Std Err	Minimum	Maximum
365	9.9515	7.3428	0.3665	-9.0000	30.0000

Mathematica

Step 4 (Bonus!): Show that your nonresponse weights improved representability

Goal: Convince your audience that your nonresponse weights helped address the problem

- **Present how the nonresponse weights, when applied to the complete case sample, improve point estimates**
- **Calculate pre and post means for a given variable**
  1. Using all observed data (the “true” means)
  2. The complete case sample, without weights
  3. The complete case sample, after applying nonresponse weights
- **The findings from approach 3 should be closer to findings from 1 than findings from 2**
  - This conveys that the nonresponse weights helped recover the true population averages among the complete case sample

Mathematica

Illustrative example: STI knowledge scores with and without weights

	Average at pre-test	Average at post-test
1) Based on all observed data	61.3	79.7
2) Complete case sample average, without weights	60.0	80.0
3) Complete case sample average, after weighting	60.5	79.9

Mathematica

Conclusion

Summary of key takeaways

- **Pre-post findings do provide value**
  - Information on how program participants’ outcomes change over time
- **But they do *not* demonstrate the impact of a program**
- **They can establish a foundation of an argument for an impact evaluation**
  - Outcomes trending in the right direction (validating logic model)
  - Magnitude of outcome change as an upper-bound estimate for power calculations

Mathematica

Best practices for pre-post analyses

- **Don't be satisfied by solely doing a basic complete case pre-post analysis**
- **Supplement with**
  - Response rate calculations
  - Nonresponse analysis
  - Estimation of nonresponse weights and use of weights in pre-post analyses
  - (Bonus): Justify that the nonresponse weights improved the representativeness of the findings

Questions?

- **Russell Cole**
  - rcole@mathematica-mpr.com

Mathematica