National TAACCCT Evaluation: Understanding Sample Sizes Needed for Impact Analysis

Tuesday, April 1, 2014

2:00 p.m. – 3:00 p.m.

Welcome and Tech

- Please mute your phones upon entry
- This webinar is being recorded
- WebEx Technical Support: 1 -866-229-3239

The Chat Box can be found in the bottom right of your screen. Please include your questions here.

Agenda

- Overview and Introductions
- Statistical Power and Minimum Detectable Effects
- Power Analysis in Practice: TAACCCT Examples
- Q&A

Burt Barnow
George Washington University

Framework of Preferred Evaluation Methodologies for TAACCCT Impact/Outcomes Analysis

- Random Assignment (Experimental Design) preferred
  - High proportion of TAA-eligible participants precludes RCT (TAA eligibles cannot be randomly assigned)
- Comparison Group Cohort (Quasi-Experimental Design)
  - Plans to enroll a moderate to high number of participants preferred

(Considerations for sample size outlined in the Round 2 SGA in Appendix H: Framework for Evaluation Methodologies, p. 53)

Planned Evaluation Methods of Round 2 Grantees

Planned Impact/Outcomes Evaluation Design for Round 2 Grantees	# of Third-Party Evaluations Using Method
Experimental Design with Random Assignment	**7**
Non-Experimental Design	**68**
Regression Discontinuity Design	2
Propensity Score Matching	21
Non-Specific, but Plan to Use a Comparison Cohort Design	56
Descriptive Outcome Study	**4**
All Round 2 Grantees	**79**

A Review of MDE Basics

Statistical Power and MDE Issues

- In hypothesis testing, it is important to avoid both type I and type 2 errors
  - Type 1 error is when we reject null hypothesis (hypothesis that there is no impact) when it is true
  - Type 2 error is when we do not reject null hypothesis when it is false, that is we are unable to reject the hypothesis that there is no impact when the true effect is not zero

Statistical Power and MDE Issues

- Typically, we worry about type 1 errors more and set the maximum type 1 error, referred to as α, at .05
- With small samples, as we have in many TAACCCT sites, there is a high risk of making a type 2 error—not being able to reject the hypothesis of no effect even though there is an impact

Statistical Power and Acceptable Error Rates

- Power is defined as the probability that, for a specified value of the true impact, we will reject the null hypothesis when it is false
- Power depends on true impact, sample size, and size of α selected
- There are no absolute rules on levels of type 1 and type 2 errors that are acceptable
- Typically we want small type 1 errors: .01, .05, or .10
- For power, we usually want power to be .70 or .80

Minimum Detectable Effect

- A concept developed by Howard Bloom to help evaluators select the most appropriate sample size and to see what size impact we can likely obtain statistically significant estimates for when there is an impact
- MDE is defined as the smallest true impact that would usually be found to be statistically significantly different from zero with specified level of significance and power
- MDE takes power calculations and puts them in a relevant form

Minimum Detectable Effect

- Orr (1999) and other sources have formulas for computing MDE.
- The MDE formulas can be used to determine the MDE for a given sample size *or* to determine the sample size needed to produce a small enough MDE.
- The most common multiplier, referred to as "Z" on the next page, is 2.80 for a 2-tail test and 80 percent power; see Table 1 in MDRC paper for alternative values of Z

Formula for MDE for Binary Outcome

- $MDE = Z(\pi*(1-\pi))^{.5}((1-R^2)/(nP(1-P))^{.5}$, where

- Z = a multiplier which converts the standard error of an impact estimator to its corresponding minimum detectable effect based on power, type 1 error, type 2 error, and 1 or 2 tail test (See MDRC paper for table with Z values on the SharePoint)
- π = the proportion of the study population with a successful outcome,
- R^2 = the explanatory power of the impact regression,
- P = the proportion of sample members randomly assigned to the program group, and
- n = the total number of sample members.

Trachtenberg School
of Public Policy
& Public Administration
THE GEORGE WASHINGTON UNIVERSITY

Formula for MDE for Continuous Outcome

- With random assignment:
  - $MDE = Z\sigma((1-R^2)/((nP(1-P))^{.5}$.
- With nonrandom assignment
  - $MDE = Z\sigma((1-R^2)/((nP(1-P)\ (1-R^2{}_A))^{.5}$,
  - Where $R^2{}_A$ = the R^2 obtained when program status is regressed on background characteristics

Trachtenberg School
of Public Policy
& Public Administration
THE GEORGE WASHINGTON UNIVERSITY

Applying MDE: A Hypothetical Example

- Suppose we are evaluating a training program where we have a comparison group developed through propensity score matching
- Assume we have 2,900 people in the treatment and comparison groups combined, with half in the treatment group and half in the comparison group
- We expect to have wage record data for the entire sample and we will set α at .05, power at .80, and use a two-tail test.

Applying MDE: A Real Example (continued)

- We need an assumption for the R^2 of the regression of earnings on the explanatory variables and the standard deviation of earnings
- We assume the R^2 is .20 based on earnings regressions in the literature, and for σ, we assume that it is $4,899 based on figures reported in similar evaluations
- Using the spreadsheet posted on SharePoint, we then find MDE = $455.70
- This means if the impact of the program on earnings is at least $455.70, we have a reasonable likelihood of obtaining a statistically significant impact estimate with the sample of 2,900

Things To Reduce the MDE

- 50-50 T/C split (not 2/3 and 1/3)
- Include more explanatory variables to increase the R^2
- Increase the size of the sample
- You can also reduce the MDE by increasing α, decreasing power, and using 1-tail tests, but these decisions should be made in advance for scientific reasons, *not* to reduce the sample needed

What Happens to the MDE When We Change Some Parameters?

- Sample size is only 1,450: MDE = $644
- Two-thirds of sample is in treatment group: MDE = $483
- We have no explanatory variables ($R^2 = 0$): MDE = $509
- We set type 1 error to .10, power to .70, and use a one-tail test: MDE = $293
- We end up with a sample of only 400: MDE = $1,227
- Note that small sample size is a big problem—many training programs have smaller impacts than this

Summary of Results from Previous Slide with Parameter Changes

	Sigma	R SQUARE	P	N	MDE
Base case	4899	0.2	0.5	2900	$455.7
1/2 sample	4899	0.2	0.5	1450	$644.4
2/3 in T group	4899	0.2	0.666667	2900	$483.3
$R^2 = 0$	4899	0	0.5	2900	$509.4
Change α, β, 1 tail	4899	0.2	0.5	2900	$292.9
Small sample	4899	0.2	0.5	400	$1,226.9

Power Analysis: TAACCCT Examples

Webinar Presentation: April 1st, 2014

Greg Cumpton

Research Associate

Ray Marshall Center

The University of Texas

RMC RAY MARSHALL CENTER for the Study of Human Resources
THE UNIVERSITY OF TEXAS AT AUSTIN Lyndon B. Johnson School of Public Affairs

Practical Considerations

- **Power calculations should be performed twice :**
  - **<u>At the beginning of the project</u>: to make certain that if the implementation and evaluation plans are executed as written the analysis will yield useable results.**
  - **<u>Prior to performing the final analysis of results</u>: to ensure that changes in sample sizes or other facts-on-the-ground (i.e., data access) are considered prior to determining program effects.**

RAY MARSHALL CENTER
for the Study of Human Resources
THE UNIVERSITY OF TEXAS AT AUSTIN
Lyndon B. Johnson School of Public Affairs

Practical Considerations

- How you create these two power analyses differ :
  - At the beginning of the project,
    - Assume that grant details will occur as stated,
    - Determine your preferred method of analysis (experimental or non-experimental),
    - Pull information needed for calculations from the literature,
    - Play with changeable aspects of the power analysis to inform responses to different potential scenarios, and
    - Use results to help communicate with grantees.

Note: Decisions regarding α, power, and 1-tail vs. 2-tail tests should be made now for scientific reasons, and should not be changed in the final round of power analysis.

Practical Considerations

- How you create these two power analyses differ :
  - Prior to performing the final analysis of results do another power analysis.
    - Use actual data from the project,
    - Pull information needed for calculations from your data (R^2 and σ),
    - Use results to be certain measured effects are beyond the minimal detectable effect .

RMC RAY MARSHALL CENTER for the Study of Human Resources
THE UNIVERSITY OF TEXAS AT AUSTIN Lyndon B. Johnson School of Public Affairs

Power Analysis Sample Size

- Throughout the project, monitor numbers of participants to ensure targets are being met or exceeded.
- Using propensity score matching and regression discontinuity will likely reduce your sample size from both the control and treatment groups. This sample loss can range from 5-15% and should be considered in your initial power calculations.

RMC RAY MARSHALL CENTER for the Study of Human Resources
THE UNIVERSITY OF TEXAS AT AUSTIN
Lyndon B. Johnson School of Public Affairs

Power Analysis Examples

- The Training for Regional Energy in North Dakota (TREND) project seeks to provide training and education for jobs in the industries benefiting from the Bakken Formation oil boom.
- The Retraining the Gulf Coast Workforce through IT Pathways Consortium (Consortium) intends to re-train workers for the IT workforce in local manufacturing and health care industries.

RMC RAY MARSHALL CENTER for the Study of Human Resources
THE UNIVERSITY OF TEXAS AT AUSTIN Lyndon B. Johnson School of Public Affairs

Power Analysis Examples

- Use the grant proposal to determine the binary outcomes,
- Determine the MDE of each of the binary and continuous variable outcomes using appropriate formulas,
  - Establish best-case worst-case scenarios:
    - Best case: Sample size is as projected, loss due to PSM or RD is 5%, R^2 is .40.
    - Worst case: Sample size is reduced by 10%, loss due to PSM is a further 15%, R^2 is .10.

RAY MARSHALL CENTER
for the Study of Human Resources
THE UNIVERSITY OF TEXAS AT AUSTIN
Lyndon B. Johnson School of Public Affairs

TREND Projected Outcomes

Outcome Measure	Number of Participants	Share of Participants
Participants served:	1527	100%
Total completing a funded program of study:	939	61%
Total number of participants completing credit hours:	1280	84%
Total number of participants earning credentials:	955	63%
Total number of participants employed after study completion	1189	78%
Total number of participants who received a wage increase:	424	28%

- **The TREND grant proposal already included projections of the share of successful participants.**

RMC RAY MARSHALL CENTER for the Study of Human Resources
THE UNIVERSITY OF TEXAS AT AUSTIN Lyndon B. Johnson School of Public Affairs

TREND Power Analysis

Outcome Measure	MDE Worst-case Scenario	MDE Best-case Scenario
Completing a funded program of study:	4%	3%
Participants completing credit hours:	3%	3%
Participants earning credentials:	4%	3%
Participants employed after study completion:	4%	3%
Participants who received a wage increase:	2%	2%
Measurable wage increase:	507.7	394.5

- **Considering the incredibly high demand for workers with even partially complete training in North Dakota, it seems reasonable that measured effects will exceed these MDEs.**

RMC RAY MARSHALL CENTER for the Study of Human Resources
THE UNIVERSITY OF TEXAS AT AUSTIN Lyndon B. Johnson School of Public Affairs

Gulf Coast Projected Outcomes

Outcome Measure	Number of Participants	Share of Participants
Participants served:	1954	100%
Total completing a funded program of study:	1660	85%
Total number of participants completing credit hours:	1174	60%
Total number of participants earning credentials:	2988	
Total number of participants employed after study completion	1475	75%
Total number of participants who received a wage increase:	459	23%

- **The Gulf Coast Consortium proposal included outcome projections.**

RAY MARSHALL CENTER for the Study of Human Resources
THE UNIVERSITY OF TEXAS AT AUSTIN Lyndon B. Johnson School of Public Affairs

Gulf Coast Power Analysis

Outcome Measure	MDE Worst-case Scenario	MDE Best-case Scenario
Completing a funded program of study:	3%	2%
Participants completing credit hours:	3%	3%
Participants earning credentials:		
Participants employed after study completion:	3%	3%
Participants who received a wage increase:	2%	1%
Measurable wage increase:	448.7	348.7

- **The slightly larger number of participants provides a lower MDE than with TREND.**

RAY MARSHALL CENTER for the Study of Human Resources
THE UNIVERSITY OF TEXAS AT AUSTIN Lyndon B. Johnson School of Public Affairs

Power Analysis Next Steps

- Once you’ve calculated the MDE with at least two reasonable scenarios, communicate with the TAACCCT grantees and ask:
  - If they currently foresee any changes in the sample size,
  - Whether they expect their outcomes to be the same,
  - Whether the expected wage increase amounts are close to the expected wage increases for participants (they should comfortably exceed those you’ve calculated).
- If they already think sample sizes or the outcome numbers change, make these changes to the power analysis to determine two new MDE scenarios.

RMC RAY MARSHALL CENTER for the Study of Human Resources
THE UNIVERSITY OF TEXAS AT AUSTIN
Lyndon B. Johnson School of Public Affairs

Thank you for your time

Webinar Presentation: April 1st, 2014

Greg Cumpton

Research Associate

Ray Marshall Center

The University of Texas

RAY MARSHALL CENTER
for the Study of Human Resources

THE UNIVERSITY OF TEXAS AT AUSTIN
Lyndon B. Johnson School of Public Affairs

QUESTION & ANSWER

Chat Box – Bottom Right of your screen

JOBS FOR THE FUTURE

CONTACT INFORMATION

Greg Cumpton
Research Associate
Ray Marshall Center
gcumpton@raymarshallcenter.org

Burt Barnow
Professor
George Washington University
barnow@gwu.edu

Randall Wilson
Senior Project Manager
Jobs for the Future
rwilson@jff.org

Dudney Sylla
Project Manager
Jobs for the Future
dsylla@jff.org

THANK YOU FOR YOUR PARTICIPATION

TAACCCTEVAL@URBAN.ORG

JOBS FOR THE FUTURE

TEL 617.728.4446 FAX 617.728.4857 info@jff.org

88 Broad Street, 8th Floor, Boston, MA 02110

122 C Street, NW, Suite 650, Washington, DC 20001

WWW.JFF.ORG