EVALUATION TECHNICAL ASSISTANCE BRIEF
for OAH & ACYF Teenage Pregnancy Prevention Grantees

Calculating Minimum Detectable Impacts in Teen Pregnancy Prevention Impact Evaluations

This brief provides an overview of how researchers can calculate the minimum detectable impacts (MDIs), which are related to power calculations, for Teen Pregnancy Prevention (TPP) evaluations. It describes a tool that evaluators can use for their own MDI calculations, and includes examples that highlight how to use the tool. A technical appendix provides more details on the formulae in the tool that inform MDI calculations.

One goal of a TPP impact evaluation is to show that the intervention being tested has a positive and statistically significant effect on student behavioral outcomes. During the design phase of a study, it is important to do a power calculation to determine the likelihood of the study being able to detect a statistically significant effect. One common way to do this is to estimate an MDI for the proposed study. An MDI is the smallest true impact, measured in the units of the outcome, for which it is likely that the estimated impact will be statistically significant.[1] For this brief, we define "likely" as having a probability greater than 80 percent, which we describe in more detail below.

By calculating an MDI, one can estimate how large a program's true impact must be in order for the proposed study design to be likely to detect it as statistically significant. For example, if a study will be powered to detect a 20 percentage point difference in sexual initiation rates (that is, the MDI is 20 percentage points), and previous research shows that the intervention has changed initiation rates by 8 percentage points, then the study is not sufficiently powered to detect the likely impacts. If a computed MDI is very large, the study will likely not yield statistically significant impacts, so researchers and funders should reconsider the study design or reassess the value of the impact evaluation.

Sections I through IV of this brief illustrate: (I) How to calculate MDIs, including the key parameters that inform them, (II) Examples of MDI calculations, (III) Methods of interpreting and evaluating MDIs, and (IV) How to present MDI calculations in a TPP proposal.

How to Calculate MDIs

This section focuses on how to calculate MDIs for two of the most common impact evaluation designs: (1) individual-level randomized designs, and (2) group- or cluster-level randomized designs.[2] In the first type of design, individuals are randomly assigned to treatment and control groups. In contrast, in a group randomized design, groups, as opposed to individuals, are assigned to treatment and control study conditions. For this reason, in a group randomized design all individuals in a group have the same treatment status.

An MDI is a function of two sets of parameters fixed by the evaluation: (1) the requirements of the commissioner or funder, and (2) the details of the evaluation design and the context in which the evaluation is taking place (such as the level of risk activity targeted by the intervention). On the following pages, we describe the components of these sets of parameters (Appendix A shows how these sets of parameters produce an MDI).

1. Parameters Fixed by the Evaluation Commissioner

Researchers should use the parameters fixed by the evaluation commissioner to produce a multiplicative factor to scale the standard error of the impact estimate (described on the following page) into an MDI. There are three parameters in this set:

- The **significance**, or probability of a *false positive*—incorrectly concluding that there is an impact when there is none. This is also called the probability of making a "Type I" error. A conventional significance level is 5 percent.

- The **power level**, or probability of not having a *false negative*—failing to detect an impact that truly exists. This is also defined as one minus the probability of a "Type II" error. An 80 percent power level is a common convention.

- The **type of hypothesis test**. Evaluators typically use a two-sided test because they are interested in whether the program has an impact regardless of whether the difference between the average outcomes for the treatment group were higher *or* lower than the average outcomes for the control group.

For the purpose of the TPP grant funding, all evaluations must assume a Type I error rate of 5 percent, a Type II error rate of 20 percent (for an 80 percent power level), and a two-sided hypothesis test.

2. Parameters Fixed by the Evaluation Design

The parameters fixed by the design are used to calculate the standard error (variance) of the impact estimate—the other component of the MDI (see Appendix A). When planning an impact evaluation, reducing the standard error of the impact estimate will shrink the MDI, making it more likely that a study will show a statistically significant impact. For individual-level designs, there are four parameters in this set:

- **The total number of individuals in the sample who contribute to the impact analysis.** The standard error of the impact changes inversely with the sample size. This is a key parameter because the evaluation costs typically increase with the sample size. Importantly, this number is *not* the number of individuals initially assigned to condition. Rather, this is the number of individuals who contribute to the final impact analysis, and thus, represents the final sample size after non-consent, program dropout, and follow-up nonresponse.

- **The proportion of individuals in the sample assigned to the treatment group.** The standard error of the impact increases when this proportion moves further from a 1:1 treatment-control group assignment ratio.

- **The variability (standard deviation or variance) of the outcome, which is a function of the prevalence rate of the outcome for dichotomous variables.** The standard error of the impact changes proportionally with the variance of the outcome. When the outcome is a dichotomous variable, the variability of the outcome can be directly computed by the prevalence rate of the outcome in the target population. For example, if the outcome of interest is teen pregnancy, then the prevalence rate of the outcome in the target population can be used to calculate the variability of the outcome for the MDI calculation.[3]

- **The proportion of the individual-level variance of the outcome related to a set of variables or covariates, if the impact estimation uses a regression model.** The variables could include factors such as demographic characteristics, baseline assessment of risk, and additional risk behaviors. When there is no correlation between the outcome and these variables, the proportion is equal to zero. When it is closer to one, much of the natural variation in the outcome can be accounted for by the covariates. Thus, the standard error of the impact is smaller when the set of control variables correlate closely with the outcome measure of interest.

Empirical estimates obtained by the Eval TA team (using National Longitudinal Survey of Youth 1997 data for youth ages 12 to 16) show that demographic characteristics, such as age, gender, or ethnicity, explain about 5 percent of the variance in the prevalence rate of sexual behaviors (such as pregnancy, risky sexual behavior, and so on). Combining demographics with baseline assessments of these sexual behaviors explains 10 to 20 percent of the total variance in sexual behavior outcomes. Finally, combining demographics and baseline assessments of sexual behaviors with baseline assessments of additional risk behaviors (such as drug and alcohol use, suspensions, and so on), explains 20 to 30 percent of the total variance in sexual behaviors.

In addition, for group-level designs, there are three other parameters:

- **The total number of groups.** For a fixed number of individuals in the sample, the standard error of the impact changes inversely with the number of groups. This is a key parameter because the evaluation costs typically increase with the number of individuals in the sample—that is, the total number of groups multiplied by the average number of individuals per group.

- **The intra-cluster or intra-class correlation coefficient (ICC), which is a measure of the degree to which outcomes of individuals within groups are correlated.** The ICC can range from zero to one. When it is zero, outcomes of individuals within groups are not correlated. When it is one, these outcomes are perfectly correlated—that is, the outcome has the same value for the entire group. In general, the standard error of the impact increases with the ICC, as the effective sample size shrinks from the total number of individuals in the sample (when ICC = 0) to the number of clusters in the sample (when ICC = 1).

For TPP outcomes, ICC empirical estimates (obtained by the TPP Eval TA team using *Add Health* data, a nationally representative sample of adolescents in grades 7 to 12) typically range from 0.01 to 0.04, which are regarded as liberal and conservative values for these types of outcomes, respectively. Larger values of the ICC (for instance, 0.10 or higher), dramatically increase the MDI, particularly if the total sample size is small (see Appendix A for more details).

- **The proportion of the group-level variance of the outcome related to a set of variables, if the impact estimation uses a regression model.** This parameter is similar to the individual-level parameter described above. The proportions of variance explained by various categories of baseline data (demographics alone, demographics plus baseline assessments of sexual behavior, demographics plus baseline assessments of sexual behavior and other risk behaviors) outlined above can also be used as empirical benchmarks for expected proportions of group-level variances. That is, if only demographic variables are included as covariates, then 5 percent of the between group variance in the outcome should be assumed for the MDI calculations. Similarly, if demographics are combined with a baseline assessment of sexual behavior, then 10 to 20 percent of between group variance can be assumed, and if they are supplemented with additional risk behaviors, then 20 to 30 percent of between group variance can be assumed for the MDI calculations.

Appendix A describes the formulae that combine these parameters into an MDI. Below is an overview of a tool for calculating the MDI, followed by examples of MDI calculations for the two basic designs.

Examples of MDI Calculations

Introduction

For this brief, the TPP Eval TA team developed a tool to facilitate the calculation of the MDI of a binary or continuous outcome for the designs described above.[4] The tool is available **here**.

The tool consists of an Excel workbook with three spreadsheets:

1. **Instructions for calculating the MDI (Yellow "Instructions" tab).** This spreadsheet describes the two sets of parameters required for the calculation: (1) parameters fixed by the evaluation commissioner, which are listed in the blue panel; and (2) parameters fixed by the evaluation design, which are listed in the orange panel. The green panel describes the MDI values returned by the spreadsheet.

2. **Example MDI calculation for an individual-level randomized design (Blue "RA-Individual (EXAMPLE)" tab).** In this spreadsheet, the parameters needed for the MDI calculation are selected from drop-down menus or entered into specific fields. These parameters follow the organization of the instructions spreadsheet described above. This spreadsheet is set for an individual-level randomized design. The data shown in this tab are outlined in example 1 to the right, but researchers can delete the sample data to use the tab for their own MDI calculations from individual-level randomized controlled trials (RCTs).

3. **Example MDI calculation for a group-level randomized design (Purple "RA-Group (EXAMPLE)" tab).** In this spreadsheet, the parameters needed for the MDI calculation are selected from drop-down menus or entered into specific fields. These parameters follow the organization of the instructions spreadsheet described above. This spreadsheet is set for a cluster-level randomized design, and therefore, includes more rows of data than the individual-level design example. The data shown in this tab are outlined in example 2 on the following page, but researchers can delete the sample data to use the tab for their own MDI calculations for cluster-level RCTs.

1) Example for an individual-level randomized design

Suppose a grant applicant is planning an individual-level randomized study of a novel TPP intervention for pregnant teens. The study will enroll teens continually over two years. Consent and baseline data collection will occur prior to random assignment. The study plans to collect demographic and sexual behavior data at baseline. It is expected that 200 teens will enroll each year, for a total sample size of 400; half will be assigned to the treatment group and half to the control group. The study team will collect follow-up survey data at the end of the program and expects that 75 percent of individuals will complete the follow-up assessment (due to a large incentive provided for completion). The outcome of interest in the study is incidence of repeat pregnancy, a binary outcome, and there are no prevalence estimates on this outcome in the target population.

Given this information, the following parameters are available for MDI calculation (see the blue tab in the tool):

Parameters Fixed by the Funder (Blue Panel):

- *Level of significance:* 0.05
- *Number of sides of test:* Two
- *Power:* 0.80

Parameters Fixed by the Study Design (Orange Panel):

1. *Total number of individuals in the sample contributing to the impact analysis:* 300, which reflects a 75 percent survey response rate of the 400 individuals originally assigned when enrolled in the study

2. *Level of randomization:* Individual

4. *Probability of assignment to the treatment group:* 0.50

5. *Type of outcome variables:* Binary

6. *Mean of the outcome variable:* Assumed to be 0.50 because its value is unknown and this value yields the most conservative estimate

7. *Standard deviation (SD) of the outcome:* Not applicable, since the outcome is binary, so the standard deviation can

be calculated from the prevalence rate of the outcome in item 6 above (this cell is blank in the worksheet).

9. *Proportion of the individual-level variance in the outcome explained by (baseline) covariates:* Assumed 15 percent, given that the study is collecting demographic and sexual behavior data at baseline.

The calculated MDI is reported in the green panel, second column from the right. In this example, the MDI is equal to 0.15, or 15 percentage points for this binary outcome, for the total sample of 300 teens. Given the small sample size, the MDI is relatively large. More specifically, the study can only detect changes of the incidence of repeat teen pregnancy of 15 percentage points or larger, which represents nearly 30 percent of the standard deviation of the outcome.[5] Whether a change in the outcome of this magnitude is feasible depends on the intervention's theory of change and extant evidence from similar studies of teen pregnancy prevention. As discussed in the previous section, the MDI could be reduced, and power increased, if:

- The sample size increases.

- The likely prevalence rate is different from 0.50.

- The proportion of the individual-level outcome explained by the covariates is greater than 15 percent (which would be possible by collecting data on additional risk behaviors).

2) Example for a group-level randomized design

Suppose a grant applicant is planning to conduct a cluster-level randomized study of a school-based TPP program implemented during 10th grade health classes. The applicant will randomly assign 10 schools to condition (half to the treatment group, half to the control group), and programming will occur for the full school year. The applicant states that there are approximately 80 10th grade students in each school, and that in previous evaluations it has received evaluation consent from approximately 50 percent of the sample. The study plans to collect demographic, sexual behavior, and other risk behavior data at baseline. For the follow-up assessment, the study team will offer incentives for participation and expects that 75 percent of individuals will complete the survey. The outcome of interest is incidence of risky sexual behavior (assumed to be 80 percent in this population, based on Youth Risk Behavior Survey [YRBS] data statistics calculated for the geographic area).

Parameters Fixed by the Funder (Blue Panel):

Same as those used in the previous example.

Parameters Fixed by the Study Design (Orange Panel):

1. *Total number of individuals in the sample contributing to the impact analysis:* 300. There were 10 schools of 80 students, or 800 students initially assigned to condition. The description assumes that consent will be obtained from 50 percent of the sample, reducing the sample size to 400. Follow-up data are expected to be obtained from 75 percent of the sample, so the final sample size for estimating impacts will be 300.

2. *Level of randomization:* Group

3. *Number of groups:* 10 schools will be assigned to condition

4. *Probability of assignment to the treatment group:* 0.50

5. *Type of outcome variables:* Binary

6. *Mean of the outcome variable:* 0.80, since the description of the study indicated that the YRBS data suggested this prevalence rate in the population

7. *SD of the outcome:* Not applicable, since the outcome is binary (blank in the worksheet)

8. *Intraclass correlation coefficient (ICC):* 0.04, which is a conservative value

9. *Proportion of the individual-level variance in the outcome explained by (baseline) covariates:* Assumed 25 percent, given that the study will collect demographic, sexual behavior, and other risk behavior data at baseline

10. *Proportion of the cluster-level variance in the outcome explained by (baseline) covariates:* Assumed 25 percent, given that the study will collect demographic, sexual behavior, and other risk behavior data at baseline

The calculated MDI is 0.19, or 19 percentage points, for the total sample of 300 teens distributed across 10 groups. Given the small sample size and the clustering of teens, the MDI is relatively large. More specifically, the study can only detect changes in the incidence of teen pregnancy of 19 percentage points or larger, which represents 47 percent of the SD of the outcome.

Note that the number of students contributing to the analytic sample in this design is *identical* to the number of students contributing to the analytic sample in the individual-level design. In addition, this study has an outcome that is more prevalent (80 percent versus 50 percent), and will collect additional baseline data on risk behaviors that are expected to explain variance in the outcome (25 percent versus 15 percent). Both of these differences in design are expected to improve the precision of the design (or reduce the MDI). However, the MDI calculated under this study design is *larger* than the MDI of the individual-level design because of the clustering of teens in schools (that is, the ICC is not zero). This is typical in a cluster-level intervention, which usually has a much *larger* MDI than an individual-level assignment design with similar sample sizes and design features. The MDI for this study could be reduced, and power increased, if:

- The total number of individuals or groups increases. Notably, the study could further reduce the MDI by increasing the number of groups without changing the total number of individuals (for example, randomly assigning 20 clusters of 40 students to condition, instead of randomly assigning 10 clusters of 80 students).

- The assumed prevalence rate is higher than 0.80.

- The ICC is smaller than 0.04, which is a challenging empirical question because it requires finding the data from other studies of teen pregnancy, on a similar sample, to estimate it.

- The proportion of the individual- and group-level variance of the outcome explained by the covariates is greater than 25 percent.

In sum, the MDIs for the two basic designs are high and, unless the intervention can generate changes in outcomes of that magnitude, the likelihood of finding a statistically significant impact is small.

Interpreting MDIs

Once an MDI has been calculated for a particular design, the onus of interpretation falls on the evaluator and/or program staff. After estimating the MDI, evaluation and program staff should consider two questions:

1. **Does the MDI seem reasonable or feasible for this intervention in this setting?** The first assessment of an MDI should be a "gut check" of the feasibility of obtaining a treatment/control difference as big as or bigger than the computed MDI. For example, if a computed MDI appears relatively large from a face validity perspective, and the evaluation is only testing a small difference in service offerings across treatment and control conditions, program staff and evaluators may be skeptical that the intervention will actually result in an impact as large as the computed MDI.

2. **Do the MDIs align with impacts observed from prior implementations of this intervention?** Some studies evaluate an intervention (or adaptation of an intervention) that already has an evidence base. In these situations, a natural benchmark for assessing the viability of a study is whether the existing evidence shows impacts that are as large as or larger than the MDI estimated for the current study.

If the MDI seems unrealistic based on perception or on an existing evidence base, program staff and evaluators should consider whether any potential changes to the evaluation design could make the MDI more feasible. This might involve changing the design to reduce the MDI (for example, by increasing the sample size or collecting additional baseline data), or increasing the dosage, duration, or intensity of the intervention to make the MDI more feasible to obtain.

Presenting MDI Calculations in a Proposal

A research proposal for a study that is aiming to demonstrate a positive and statistically significant effect on youth sexual risk behavior should present three key features:

1. **The computed MDI for the outcome of interest.** Of course, the most important feature of an MDI calculation is the actual MDI for the study.

2. **The assumptions and sources for assumptions that informed the MDI calculation.** As described above, certain key parameters determine the MDI for a research study. In order to convince a skeptical reader that the "ingredients" for the MDI are correct, proposals must articulate the key assumptions that serve as inputs to the calculation, and justify them based on the study design and framing. For example, at a minimum, a proposal should describe the number of individuals (and groups, as needed) that will be randomly assigned to conditions, the expected response rates, and the proportion of variance in the outcome explained by baseline covariates. In addition, as necessary, the sources for parameters should be included: for example if the MDI incorporates a prevalence rate for the outcome as an input to the calculation, the source of this prevalence rate should be described.

3. **The justification for the MDI as reasonable.** Finally, the proposal must convince the reader that the MDI is feasible for the given study to discover. The justification should be based on previous research that highlights impact estimates of a similar magnitude, an argument that a novel intervention might show substantively large impacts, and/or reasons why the MDI is an appropriate threshold for testing the intervention.

References

Dong, Nianbo, and Rebecca Maynard. "PowerUp!: A Tool for Calculating Minimum Detectable Effect Sizes for Experimental and Quasi-Experimental Design Studies." *Journal of Research on Educational Effectiveness*, vol. 6, no. 1, 2013, pp. 24–67. Excel workbook available at http://web.missouri.edu/~dongn/. Accessed July 2, 2014.

Spybrook, Jessaca, Howard Bloom, Richard Congdon, Carolyn Hill, Andres Martinez, and Stephen Raudenbush. *Optimal Design Plus Empirical Evidence: Documentation for the "Optimal Design Plus" Software, Version 3.0.*, 2011, Available at http://hlmsoft.net/od/. Accessed July 2, 2014.

Endnotes

[1] In contrast, a minimum detectable effect size (MDES) is the smallest true impact that is likely to be detected as statistically significant, measured in terms of effect size (standard deviations of the outcome). The MDES is particularly useful for evaluations with continuous outcomes, such as attitude or knowledge scales. For more info on MDES, please see the final section of Appendix A.

[2] The procedures outlined in this brief could also be used to compute power for quasi-experimental designs; however, they will provide an overly optimistic assertion of the MDIs. In a quasi-experimental design, background characteristics are typically correlated with treatment assignment (this is unlikely to occur in a randomized experiment), which increases the MDI. However, the magnitude of the correlation between background characteristics and the treatment assignment variable is very difficult to predict, and therefore, we have focused our efforts on the more tractable procedure of MDI estimation in experimental designs.

[3] If the prevalence rate of the outcome is unknown, researchers should conservatively assume that the prevalence rate is 50 percent in the study sample, as this will produce the largest MDI.

[4] Other tools available for calculating MDIs are *PowerUp!* (Dong and Maynard 2013) and *Optimal Design* (Spybrook et al. 2011). Because these tools require knowledge of the variance of the outcome for more complex evaluation designs, they are not discussed in this brief.

[5] Some researchers prefer to consider MDIs in terms of standard deviations of the outcome, which represents the smallest difference that can be detected in "effect size" units. See the last section of Appendix A for more details on this.

APPENDIX A:
Mathematical representation of the minimum detectable impact (MDI)

The MDI formula can be expressed as

(1) $MDI = Factor * SE(impact)$

where

Factor is a constant that is a function of the significance and the statistical power level, *SE(impact)* is the standard error of the impact estimate.

Below, we describe these two components in detail.

A. Factor

Factor becomes larger as the significance level decreases and the power level increases. Thus, the MDI rises when the evaluator seeks to reduce the chances of making Type I and Type II errors. *Factor* becomes smaller when the evaluator uses a one-sided test instead of a two-sided test with a fixed power level because the critical value of the test for a given significance level is smaller with a one-sided test. Mathematically, *Factor* is expressed by the following equation

(2) $Factor = (T_{df}^1(1-\alpha/2) + T_{df}^1(1-\beta))$

where

is the significance level (in this formula, it is divided by 2 because we want a two-sided test),

the desired power level,

Number of sides of the hypothesis test = the denominator below (we set this to be 2),

is the inverse of the *Student's t* distribution function with *degrees of freedom (df)* equal to the total sample size (individuals or clusters) minus 2.[1]

B. Standard Error for the Impact

The standard error of the impact, *SE(impact)*, varies according to the impact evaluation design and the parameters described in Section II. Generally, larger samples reduce *SE(impact)* and, thereby, the MDI, making the evaluation more powerful. Greater power is desirable because the evaluation is more likely to detect meaningful impacts, although greater power typically leads to higher costs resulting from a larger sample size and the associated costs of data collection.

Below we present the equations for the standard error of the impact for two basic types of designs: (1) individual-level randomized design, and (2) group-level randomized design.

1. Individual-level randomized design

The formula for the standard error of the impact of an individual-level randomized design is

(3a) $SE(impact_{IN}) = \sqrt{\dfrac{\sigma_\varepsilon^2 \left(1 - R_{(y_i,x)}^2\right)}{p(1-p)N}}$

where

N = total analytic sample size ***after attrition and survey nonresponse*** (that is, the sample on which you estimate impacts),

p = proportion of the total sample assigned to the treatment group,

is the variance of the outcome, which can be calculated as if the outcome is dichotomous—that is, the prevalence rate,[2]

the proportion of the individual-level variance of the outcome, *y*, explained by the individual-level variables *x*.

2. Group-level randomized design

The formula for the standard error of the impact of a group-level randomized design is

$$\textbf{(3b)} \quad SE(impact_G) = \sqrt{\frac{\sigma_\varepsilon^2}{p(1-p)N}\left[\frac{(1-\rho)\left(1-R^2_{(y_{WG},x)}\right)}{gm} + \frac{\rho\left(1-R^2_{(y_{BG},x)}\right)}{g}\right]}$$

where, in addition to the parameters defined above,

g = total number of groups,

m = the average number of individuals per group, which is calculated as N/g,

 intra-cluster correlation coefficient, which varies between 0 and 1(this parameter is a measure of the degree to which outcomes of individuals within groups are correlated),

 the proportion of the within-group (WG) variance of the outcome, y, explained by covariates x,

 the proportion of the between-group (BG) variance of the outcome, y, explained by covariates x.

Randomly assigning groups rather than individuals reduces *the effective sample size*—that is, the sample size that results from accounting for the correlation of outcomes within groups (remember, as approaches 1, the effective sample size of the study reduces to g instead of N). For this reason, the MDI of an individual-level design always is smaller than the MDI for a group-level design—and the "best-case" power for all designs.

C. Translating the Minimum Detectable Impact to a Minimum Detectable Effect Size

Finally, the standardized MDI— that is, the minimum detectable effect size (MDES), is equal to the MDI divided by the standard deviation of the *outcome, SD(outcome)*. The MDES formula can be expressed as

$$\textbf{(4)} \quad MDES = MDI / SD(outcome)$$

The MDES is particularly useful for evaluations with continuous outcomes, such as attitude or knowledge scales, which typically are expressed as the standard deviation of the outcome.

---

[1] This is a simplifying assumption and may not be accurate for studies using stratified designs or a large number of baseline covariates to improve precision of the impact analysis. The benefit of using this simplified calculation is that it alleviates user burden in calculating the number of degrees of freedom sacrificed for design and analytic approaches (see Dong and Maynard [2013] for a more formal calculation of the degrees of freedom for complex designs).

[2] If the goal of the MDI calculation is to estimate an MDI in standard deviation units, use $\sigma_\varepsilon^2 = 1$.