December 2012 • Brief 2

Estimating Program Impacts for a Subgroup Defined by Post-Intervention Behavior: Why is it a Problem? What is the Solution?

The impact of teenage pregnancy prevention programs on an outcome like contraceptive use for sexually active youth is often of interest to researchers and policy-makers. This brief describes a serious pitfall in examining impacts among sexually active youth and provides strategies for structuring analyses in ways that avoid the problem.

The strongest evidence of program effectiveness comes from evaluations based on an experimental design. The use of random assignment to form the treatment and control groups ensures that the two are equivalent on observed and unobserved characteristics; any differences between the two groups before the program begins are guaranteed to be random. The only systematic (nonrandom) difference between the treatment and control groups is the offer of a particular program to the treatment group. This is why evidence from experimental evaluations is considered strong—we can be confident that differences in outcomes between the treatment and control groups are caused by the program being evaluated. In other words, estimates of program impacts are unbiased.

Anything that introduces systematic differences between the treatment and control groups, other than the offer of the program, undermines the strength of the experiment and the ability to draw unbiased conclusions about program effectiveness. For example, researchers typically recognize that sample attrition has the potential to create differences between groups and cause biased estimates of program effectiveness. Noncompliance with assignment status is another commonly recognized source of bias. Less recognized is the fact that certain analytic decisions made by researchers can also create systematic compositional differences between the treatment and control groups, thereby biasing impact estimates even in studies with a perfectly executed random assignment design.

One example of such an analytic decision (and the focus of this brief) is to estimate program impacts for a subgroup of the sample that is defined by characteristics or behaviors observed post–random assignment that may themselves be affected by the program. For example, in evaluations of teenage pregnancy prevention programs, it is tempting to estimate program impacts on contraceptive use among the subgroup of youth who are sexually active during the interval between program completion and follow-up data collection (or any period defined as post–random assignment). Because a teenage pregnancy prevention program may influence not only whether or not a teenager uses contraception, but also whether he or she has sexual intercourse, creating a subgroup of sexually active youth using contraception is partly determined by a variable that is in itself an outcome. The compositional balance or expected equivalence on both observed and unobserved characteristics for the full sample that results from random assignment is not guaranteed in a subgroup of the sample that is defined by an outcome measure. Due to the potential for unobserved compositional differences between the treatment and control groups in such a subgroup, impact estimates for such a subgroup (from here on referred to as an *endogenous subgroup*) are subject to bias.

This brief demonstrates why estimating program effectiveness for an endogenous subgroup leads to compositional differences between the treatment and control groups that may result in biased impact estimates. Specifically, it shows the source of bias in the estimated impact of a teenage pregnancy prevention program on the likelihood that youth use contraception, among a subgroup of youth who are sexually active post–random assignment. It also provides guidance for meeting similar research objectives by constructing outcomes so that they yield unbiased estimates of program impacts.

Source of the Bias in Estimates Based on Endogenous Subgroups

To illustrate how compositional differences can be introduced when limiting analyses to those who are sexually active post–random assignment, it is helpful to first classify youth into three mutually exclusive categories. The categories—denoted by Type A, B, and C—are based on a youth's potential to be sexually active under two conditions: (1) if assigned to the treatment group and (2) if assigned to the control group (Table 1). We observe individuals in the group (treatment or control) they were assigned to, but we cannot observe which type they are. Still, this "potential outcomes" framework recognizes that their behavior might have been different had they been assigned to the other group.[1] Some youth will have sex if assigned to either group, some will have sex only if assigned to one group but not the other, and some will not have sex regardless of the group they are assigned to.

Table 1. Three Categories of Youth Based on Their Potential to Have Sexual Intercourse if Assigned to the Treatment or Control Group

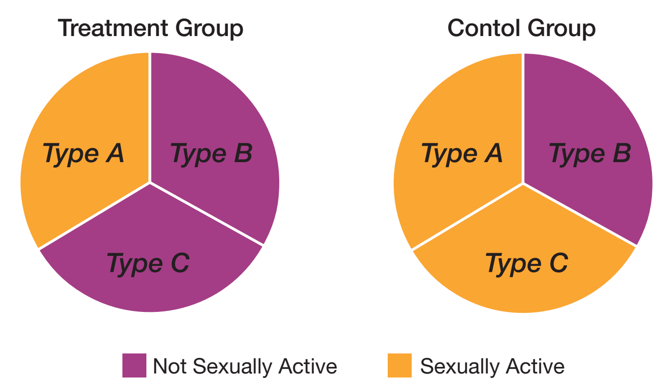
Type A	Will be sexually active if assigned to either the treatment or control group
Type B	Will not be sexually active if assigned to either the treatment or control group
Type C	Will be sexually active if assigned to the control group but not if assigned to the treatment group

Type A youth will be sexually active regardless of whether they are assigned to the treatment or control group, whereas Type B youth are the opposite—they will not be sexually active if assigned to either. For Type A and Type B youth, therefore, the intervention does not affect sexual activity. Type C youth will be sexually active if assigned to the control group but not if assigned to the treatment group—an outcome anticipated from a teenage pregnancy prevention program.[2]

Because of the nature of random assignment, the distribution of youth type in the treatment group will be the same as the distribution in the control group. For example, if each type makes up a third of the population, then each type should make up a third of the treatment group and a third of the control group. Given the compositional equivalence between the treatment and control groups, differences in outcomes (such as contraceptive use) between the two groups are valid estimates of program effectiveness, if estimated for the full sample. However, restricting the analytic sample to a subgroup of youth who are sexually active at follow-up introduces compositional differences between the two analyzed groups. As Figure 1 shows, the sexually active youth in both the treatment and control groups include Type A youth, as they will be sexually active if assigned to either group. In addition, the sexually active youth in the control group include Type C youth, who will be sexually active only if assigned to the control group. Consequently, analysis of program impacts on the proportion using contraception among youth who are sexually active at follow-up is based on the comparison of two compositionally different groups: Type A youth in the treatment group and Type A and C youth in the control group. Because of this compositional difference, the comparison of sexually active youth between the two groups will produce a biased estimate of program impacts.

If teenage pregnancy prevention programs do affect whether youth are sexually active, and the estimate of contraceptive use among the subgroup of sexually active youth is biased, the bias could be either a positive or negative. In other words, it could overstate or understate the impact estimate on contraceptive use. Whether the bias is positive or negative depends on whether Type C students in the control group are more or less likely than Type A students to use contraception. Furthermore, the biased impact estimate could lead one to erroneously conclude that there is an impact on contraceptive use at all, when in fact the impact could be on rates of sexual activity. In practice, neither the magnitude nor the direction of the bias can be calculated, but it is not credible to assume that the bias does not exist. (See the Technical Appendix for a more detailed explanation of the composition of the bias.)

Figure 1. Potential for Being Sexually Active at Follow-Up, by Type of Youth and Assignment Status



Notes: For illustration, in this figure we assume that each type of youth makes up a third of the population. In reality, we cannot observe what the distribution is. Regardless of the actual distribution between Type A, B, and C youth, however, the distribution will be the same in both the treatment and control groups because of the nature of random assignment.

Alternative Strategies for Estimating Unbiased Impacts

A biased impact estimate can be avoided through careful construction of analysis samples and outcomes, allowing researchers to answer important questions about program effects on youth engaging in sexual behaviors.

Defining the analytic sample. Two approaches can be used to construct an analytic sample that will facilitate calculation of unbiased impacts. One approach is to use pre-intervention measures (instead of post-intervention measures) of sexual activity to define the analytic sample. Given that random assignment creates two equivalent groups on observed and unobserved characteristics, the treatment and control groups are not systematically different from each other on pre-intervention (exogenous) measures, such as sexual initiation. Therefore, one could estimate the impact of the intervention on contraceptive use among the youth who were sexually active prior to the intervention. One key consideration for this approach is that, depending on the age or other characteristics of the sample, only a small proportion of the sample may be sexually active prior to the program. Therefore, taking this approach could mean restricting the analytic sample to a small proportion of the entire randomly assigned sample, thereby limiting the statistical power to detect impacts. The second approach—estimating impacts for the entire randomly assigned sample with outcome data—will also provide unbiased estimates of program impacts, and is also valuable because it can maximize the power of the study to detect impacts.

Constructing the outcome measure. No matter what approach is used to construct the analytic sample, all youth with outcome data in either the full sample or a subgroup defined by pre-intervention characteristics must be included in the analysis of impacts, regardless of whether or not they are sexually active at follow-up. The challenge is to construct the outcome measure in such a way that appropriately handles youth that are not sexually active. Some examples for doing so follow.

Measures of some outcomes, such as the prevalence of pregnancy, are relatively straightforward to define and interpret. For example, youth who were not pregnant (or did not cause a pregnancy) post–random assignment are assigned a value of zero, whether or not they have been sexually active.

An Unbiased Estimate of Contraceptive Use Among the Sexually Active Is Available in Theory but Not Practice

An unbiased estimate of the program's effect on the proportion of youth using contraception could be obtained by comparing the contraceptive use of the sexually active youth in the treatment group to the contraceptive use of the *same type of youth* in the control group. Because the sexually active in the treatment group are Type A youth, comparing the proportion of youth using contraception among Type A youth in the treatment group to the proportion among Type A youth in the control group would provide an unbiased estimate. In practice, however, the researcher cannot observe which category youth are in. For example, the researcher cannot differentiate between Type A and Type C youth in the control group. Therefore, in practice, it is not possible to select Type A youth in the control group to form a valid counterfactual to Type A youth in the treatment group.

Comparing the proportion of youth using contraception among the sexually active post–random assignment between the treatment and control groups would provide an unbiased impact estimate *only* if there are no Type C youth in the study population.[3] Under such a scenario, the sexually active in both groups would be Type A youth, and the two groups would be compositionally equivalent. Because Type C youth are, by definition, the youth whose likelihood of being sexually active is affected by the program, assuming that they are not present in the sample is the same as assuming that the program has no impact on sexual activity. In other words, the estimates for the impact on contraceptive use conditional on being sexually active could only be considered unbiased if the program could not influence whether someone is sexually active—an unrealistic expectation for a teenage pregnancy prevention program. (For a more detailed explanation, see the Technical Appendix.)

Analyzing "contraceptive use" is less straightforward because the outcome is undefined for youth who are not sexually active. This issue can be addressed by instead defining measures that can be intuitively constructed and interpreted regardless of a youth's sexual activity at the follow-up period. For example, instead of "contraceptive use," the outcome could be defined as "avoiding unprotected sex." In this case, the outcome would be dichotomous and a youth could avoid unprotected sex (and be coded 1) if they use contraception consistently or are not having sex at follow-up. Youth who are having sex and are not using contraception consistently are coded as 0.

A similar logic can be applied to an outcome that measures risk of exposure to sexually transmitted infections (STIs). A measure of risk of exposure to STIs could be defined as the proportion of youth who are sexually active but do not use condoms consistently. For this dichotomous measure of risk, these youth would be coded as 1. Youth who abstain or use condoms consistently would be coded as 0.

For measures of risk or risk avoidance, such as those described above, supplemental explorations can aid understanding. To gain a better understanding of which component of an overall measure of risk or avoidance of risk is contributing to the estimated impacts, it may be helpful to supplement the analysis by examining program impacts on the individual measures among the full sample. For example, analysis of an overall measure of risk of exposure to STIs, defined as the proportion of the full sample that is sexually active without using condoms, can be supplemented by the analysis of the two separate measures: (1) the proportion of the full sample that is sexually active (an indicator that takes on the value 1 for youth who are sexually active and 0 otherwise) and (2) the proportion of the full sample that is sexually active and uses condoms (an indicator that takes on the value 1 for sexually active youth who use condoms and 0 otherwise). Examining these results can illuminate whether a reduction in risk of exposure to STIs is driven by a reduction in sexual activity, an increase in condom use, or both.

Examining the effectiveness of teenage pregnancy prevention programs on risk behaviors, like contraceptive use, for only those youth sexually active at follow-up can produce a biased impact estimate. As explained in this brief, the bias is introduced because the compositional equivalence of the groups achieved through random assignment can no longer be assumed to be present if the analytic sample is defined by an outcome that can be influenced by the intervention. Unbiased estimates of program effectiveness can be achieved by constructing measures of risky behavior that appropriately account for youth who are not sexually active at follow-up and defining the analytic sample in such a way that maintains the integrity of random assignment.

Endnotes

[1] Rubin, D. B. "Estimating Causal Effects of Treatments in Randomized and Nonrandomized Studies." *Journal of Educational Psychology*, vol. 66, 1974, pp. 688–701.

[2] A fourth type of youth could be those who are sexually active if assigned to the treatment group but not if assigned to the control group. In this brief, we assume that there are none of these youth in the population. As demonstrated in the Technical Appendix, the general conclusion about the bias is not affected by this assumption.

[3] If the average outcome of interest for the Type C youth in the control group is the same as the average outcome for Type A youth in the control group, then the average outcome among the sexually active in the control group would be the same as if Type C youth did not exist in the study population. This would produce the same impact estimate (in terms of magnitude and direction) that would be calculated if Type C youth did not exist.

This brief was written by Silvie Colman from Mathematica Policy Research for the HHS Office of Adolescent Health under contract #HHSP23320095642WC, Task Order No. HHSP23337017T.

TECHNICAL APPENDIX

The expected effect of the program on the likelihood of consistent use of contraceptives, among youth who are sexually active can be expressed as

$$\mathrm{E}\left[Y_i \mid S_i = 1, T_i = 1\right] - \mathrm{E}\left[Y_i \mid S_i = 1, T_i = 0\right] \quad \textbf{(1)}$$

where Y_i equals 1 if youth *i* uses contraceptives consistently, and equals 0 otherwise; S_i equals 1 if youth *i* is sexually active, zero otherwise; and T_i is the treatment indicator that equals 1 for youth assigned to the treatment group, and 0 for youth assigned to the control group. Treatment status is determined by random assignment.

The source of the bias is not apparent from equation (1). To reveal the source of the bias, it is helpful to reframe equation (1) in terms of potential outcomes notation (which cannot be observed or estimated directly) and to compare the valid causal and the biased impacts.[a] For each type of outcome, every youth *i* has two potential outcomes: one that the youth would reveal if assigned to the treatment group, and one that the youth would reveal if assigned to the control group. In particular, let S_{Ti} equal 1 if youth *i* would be sexually active if assigned to the treatment group, and let S_{Ci} equal 1 if youth *i* would be sexually active if assigned to the control group. Following the same logic, let Y_{Ti} and Y_{Ci} be indicators for whether the youth would use contraceptives consistently if assigned to the treatment group and if assigned to the control group, respectively. Following this notation, equation (1) can be expressed in the potential outcomes framework as

$$\mathrm{E}\left[Y_{Ti} \mid S_{Ti} = 1\right] - \mathrm{E}\left[Y_{Ci} \mid S_{Ci} = 1\right] \quad \textbf{(2)}$$

The first term in equation (2) shows the likelihood of consistent contraceptive use under the treatment condition among youth who have sexual intercourse in the treatment group. As Table 1 indicates, these are Type A youth. The second term shows the likelihood of consistent contraceptive use under the control condition among youth who have sexual intercourse in the control group, which are Type A and C youth. Consequently, an estimate that is based on equation (1) compares the treatment outcome of Type A youth to the control outcome of Type A and C youth combined, which does not provide a valid impact estimate of the program's effectiveness in increasing the consistent use of contraceptives among youth who are sexually active in the treatment group.

Following the notation used in equation (2), the causal effect of the program among sexually active youth can be expressed as

$$\mathrm{E}\left[Y_{Ti} \mid S_{Ti} = 1\right] - \mathrm{E}\left[Y_{Ci} \mid S_{Ti} = 1\right] \quad \textbf{(3)}$$

which is the difference between the treatment and control outcome among youth who are sexually active in the treatment group (Type A). Taking the difference between equations (2) and (3)—the biased value based on the endogenous subgroup and the true causal effect of the program—reveals the source of the bias, as expressed in equation (4)

$$\underbrace{\left\{\mathrm{E}\left[Y_{Ti} \mid S_{Ti} = 1\right] - \mathrm{E}\left[Y_{Ci} \mid S_{Ci} = 1\right]\right\}}_{\text{Biased Value}} - \underbrace{\left\{\mathrm{E}\left[Y_{Ti} \mid S_{Ti} = 1\right] - \mathrm{E}\left[Y_{Ci} \mid S_{Ti} = 1\right]\right\}}_{\text{True Causal Effect of Program}} = \underbrace{\mathrm{E}\left[Y_{Ci} \mid S_{Ti} = 1\right] - \mathrm{E}\left[Y_{Ci} \mid S_{Ci} = 1\right]}_{\text{Bias}} \quad \textbf{(4)}$$

Equation (4) shows that the bias is a function of the potential outcome under the control condition $\left(Y_{Ci}\right)$ among youth who are sexually active if assigned to the treatment group $\left(S_{Ti} = 1\right)$ and youth who are sexually active if assigned to the control group $\left(S_{Ci} = 1\right)$. Specifically, it is the difference between the average outcome in the control group among Type A youth and the weighted average of the control-group outcome among Type A and C youth combined (the weights are based on the distribution of Type A and Type C youth among the sexually active in the control group). Simply put, the bias arises because the likelihood of using contraceptives consistently may be different between Type A and Type C youth in the control group.

Adding Type D youth (youth who are sexually active in the treatment group but not in the control group) to the sample will slightly alter the composition of the bias, because their presence will alter the weighted average of the outcome among youth who are sexually active if assigned to the treatment group (Type A and Type D)—the first term in the bias expressed in equation (4). The magnitude of the bias may therefore change. Regardless of the presence or absence of Type D youth, estimates for the sexually active at follow-up will be biased as long as there are Type C youth—youth whose likelihood of being sexually active is affected by the intervention.

[a] The formal presentation of the bias is based on the example provided in: Angrist, J., and S. Pishke. *Mostly Harmless Econometrics: An Empiricist's Companion*. Princeton, NJ: Princeton University Press, 2009, Section "3.2.3 Bad Control," p. 34.