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

November 2014 • Brief 5

Sample Attrition in Teen Pregnancy Prevention Impact Evaluations

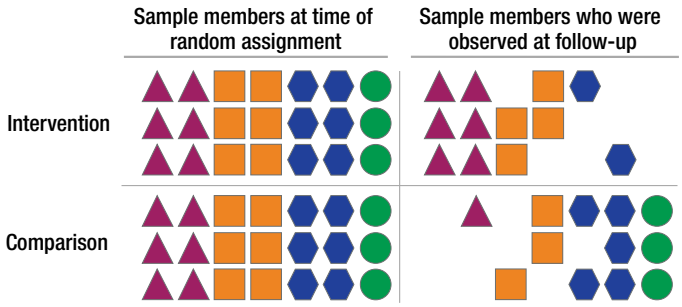
A randomized controlled trial (RCT) allows for an unbiased test of program impact, provided that the impact is estimated using the full sample that was initially assigned to condition. Random assignment ensures that the assigned intervention and comparison groups are similar on all pre-intervention characteristics (any differences will be due to random sampling error). Therefore, any differences in outcomes observed across groups after the intervention can be attributed to the effect, or "impact," of the intervention. Sample attrition is a key threat to achieving such unbiased impact estimates. In this brief, we discuss how attrition affects individual- and cluster-level RCTs, how it is assessed, and strategies to limit it. We pay particular attention to meeting the requirements of the current U.S. Department of Health and Human Services (HHS) Evidence Standards for Teen Pregnancy Prevention (TPP) Evaluations.

What is attrition in impact evaluations, and why is it a problem?

Attrition occurs when randomly assigned sample members are lost from the analysis due to nonconsent, item nonresponse, or entire survey nonresponse.[1] The loss of study participants can bias the study's impact estimates by creating differences in the distribution of characteristics of the intervention and comparison groups. The intervention may affect whether or not an individual will participate throughout the study period and complete a follow-up assessment. Therefore, people who drop out of a study may be very different from those who do not drop out. For example, some intervention group members may drop out of a study soon after experiencing the program because they do not find the services useful. As a result, in RCTs where the initially assigned groups are equivalent on key baseline variables, attrition can produce final samples that are not comparable. Therefore, when outcomes are compared in the final samples (which will be subsets of the samples originally assigned to condition), the resulting impact estimates will be biased due to underlying differences between the intervention and comparison groups being used to estimate the impacts. See Figure 1 for a visual example of this.

In Figure 1, at the time of random assignment, the intervention and comparison groups are equivalent on background characteristics. (In this example, assume the colors of the sample members represent their proclivity to engage in risky/unprotected sexual activity.) However, at the follow-up period, there was some sample attrition, and only a subset of the initially assigned sample members is observed. In this example, these remaining sample members have very different background characteristics (intervention group is predominantly magenta/orange, and comparison group is predominantly blue/green). If impacts are estimated using this sample, any post-intervention differences would conflate intervention effects with the fact that these subsamples have very different baseline characteristics.

Figure 1. Illustration of non-equivalence of baseline characteristics due to sample attrition

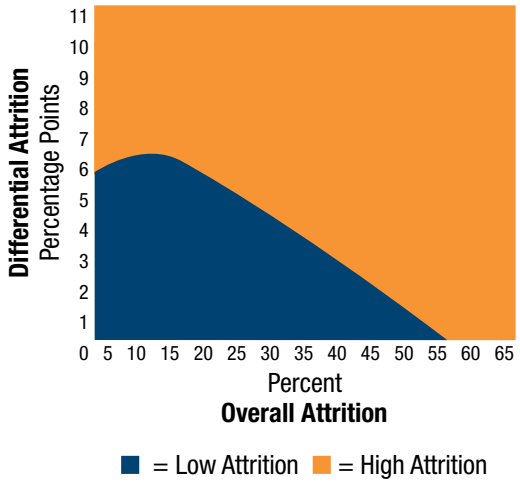


How is attrition assessed against HHS evidence standards?

The HHS evidence review assesses the level of sample attrition against standards established by the U.S. Department of Education's What Works Clearinghouse (WWC).[2] As Figure 2 shows, the attrition standards recognize a trade-off between "overall" and "differential" attrition, where overall attrition reflects the total amount of nonresponse in the sample as a whole (including both intervention and comparison groups—this kind of attrition is shown on the horizontal axis of Figure 2), and differential attrition reflects the differences in the attrition rates between the intervention and comparison groups (shown on the vertical axis of Figure 2).

Office of Adolescent Health

Figure 2. Standard for assessing sample attrition in study quality ratings



- **Blue region (low attrition).** This area of the figure shows an allowable combination of overall and differential attrition that will limit bias due to nonresponse.
- **Orange region (high attrition).** This area of the figure shows a combination of overall and differential attrition that does not adequately limit bias. When a study has attrition levels in this region, the observed impact is likely to contain substantial bias due to nonresponse.

Studies with relatively little overall attrition can meet standards with moderate differential attrition, but studies with relatively severe overall attrition require a lower level of differential attrition to meet standards. Therefore, the cutoff for an acceptable level of sample attrition is tied not to the extent of overall attrition only or differential attrition only, but rather to a combination of the two. For example, for studies with a relatively low overall attrition rate of 10 percent, the attrition standard allows a rate of differential attrition up to approximately 6 percentage points. However, for studies with a higher overall attrition rate of 30 percent, the attrition standard requires a lower rate of differential attrition, at approximately 4 percentage points. See Appendix A for a table of attrition values that provides more detail than Figure 2.

The method for calculating sample attrition differs depending on whether the study randomly assigns people to condition (individual-level RCT) or clusters to condition, such as assigning schools to intervention or comparison conditions (cluster RCT).

Individual-level RCT

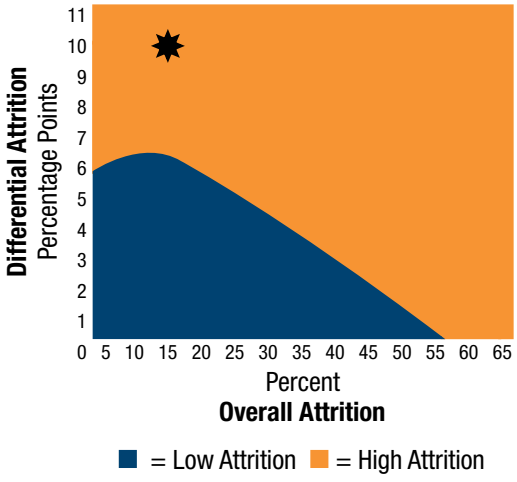
For an individual-level RCT study design, the attrition calculation is a simple comparison of sample sizes observed at follow-up relative to the sample sizes at the time of random assignment. The following box provides an example of the calculations used to produce both an overall and a differential attrition rate.

Example individual-level RCT attrition calculation

Consider a study with 100 youth assigned to the intervention condition and 100 youth assigned to the comparison condition. Assume that follow-up data were obtained from 80 youth in the intervention condition (20 youth attrite, which represents a 20 percent attrition rate in the intervention group) and 90 youth in the comparison condition (10 youth attrite, which represents a 10 percent attrition rate in the comparison group). Thus, the overall attrition rate is 30/200 = 15%, and the differential attrition is = 20% – 10% = 10 percentage points.

When the combination of overall and differential attrition in this example is plotted in Figure 3, we see that this combination falls within the orange region. That is, a combination of 15 percent overall attrition on the X axis and 10 percentage point differential attrition on the Y axis results in a point in the orange/high attrition area of Figure 3. This implies that attrition bias exceeds the desired thresholds; therefore, the authors would be required to demonstrate baseline equivalence of the sample on observed characteristics. See the **HHS evidence review protocol, version 3.0** for more details on establishing baseline equivalence.

Figure 3. Example individual-level RCT illustrates "high" level of attrition



Cluster-level RCT

For cluster-level RCTs, in which people are assigned to intervention and comparison conditions in groups (for example, schools or classrooms), attrition is calculated in two steps:

1. **Cluster attrition assessment.** The number of clusters initially assigned to condition is compared against the

number of clusters that contribute youth sample (subcluster) members to the impact analysis sample to produce overall and differential cluster-attrition rates. The combination of the overall and differential attrition rates is examined relative to the attrition figure. If there is high cluster attrition, the study must demonstrate baseline equivalence. If the study has low cluster attrition, then youth attrition is assessed.

2. **Youth attrition assessment.** The assessment of youth attrition is similar to the assessment of attrition in an individual-level RCT, with one exception. For cluster RCTs, attrition is calculated by comparing the same ratio of youth with follow-up data to youth randomly assigned, but the calculation includes youth in only the clusters contributing to the impact analysis (the clusters that did not attrite). This modification was done to prevent double-counting of sample attrition (at the cluster and youth levels). Table 1 provides an example of this.

Table 1 shows a cluster RCT in which 40 groups were randomly assigned to condition (20 to the intervention condition, and 20 to the comparison condition), where each group contained 100 youth at the time of random assignment. One cluster in the intervention condition dropped out after random assignment; therefore, the overall cluster attrition rate is 2.5 percent and differential attrition is 5 percentage points. In the youth attrition calculation, youth attrition is calculated relative to the number of youth in clusters that did not attrite, rather than to the initial number of youth in all clusters at random assignment, to guard against double-counting those youth in the attrition calculations. Therefore, in Table 1, in the calculation of the youth attrition rate for the intervention group, the denominator is 1,900 youth, rather than 2,000. This produces an overall youth attrition rate of 20 percent and a differential attrition is 0 percentage points.

Figure 4. Both cluster and subcluster attrition levels from Table 1 result in "low" levels of sample attrition.

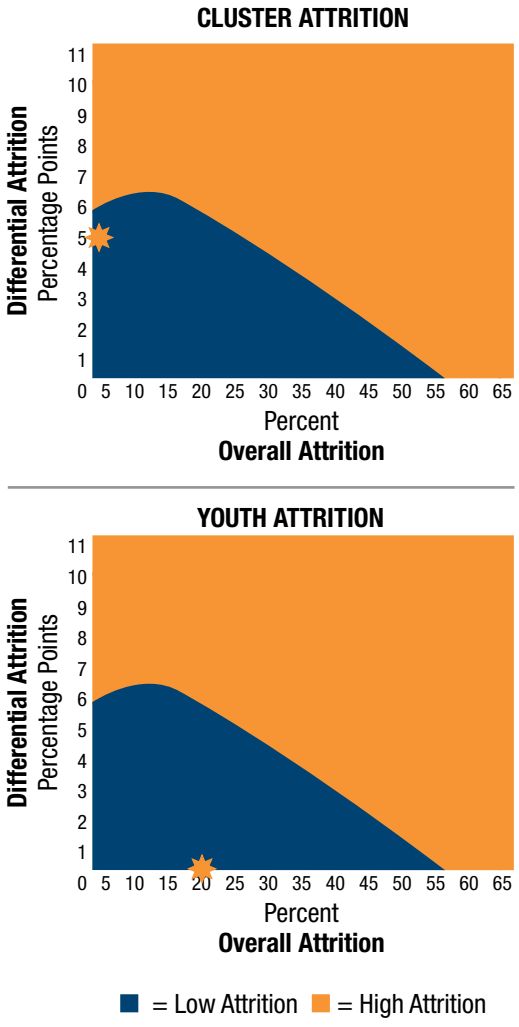


Table 1. Example of assessing youth attrition when there is cluster-level attrition

Cluster attrition calculation			
	Intervention	**Comparison**	**Overall**
Number of clusters in initial random assignment	20	20	40
Number of clusters observed at follow-up	19	20	39
Cluster attrition rate	5% = (20 – 19) /20	0%= (20 – 20) / 20	2.5%= (40 – 39) / 40
Youth Attrition Calculation			
	Intervention	**Comparison**	**Overall**
Number of youth randomly assigned in all clusters	2,000	2,000	4,000
Number of youth randomly assigned in clusters that did not attrite	1,900	2,000	3,900
Number of youth observed at follow-up	1,520	1,600	3,120
Youth attrition rate	20% = (1,900 – 1,520) / 1,900	20% = (2,000 – 1,600) / 2,000	20% = (3,900 – 3,120) / 3,900

As Figure 4 shows, both cluster- and subcluster-level attrition fall within the acceptably low range when plotted on the attrition standards graph.

Note: According to current HHS evidence standards, cluster RCTs with low attrition at the cluster level but high attrition at the subcluster level are assigned the moderate study rating. Cluster RCTs also receive a moderate rating if sample members were added during the intervention period (for example, if a study of a multiyear pregnancy prevention program for high school students included in the impact analysis new students who transferred into the school the year after the program began).

Quasi-experimental designs

Attrition standards are not applied to quasi-experimental studies. This is because these studies are reviewed based on the baseline equivalence of their final analytic samples, from which there is no attrition.

Strategies for limiting attrition in TPP evaluations

Attrition is driven by the loss of sample members who were initially randomized but were not included in the ultimate impact analysis. Common sources of attrition in TPP evaluations include nonconsent after random assignment, dropping out of a study, and item or full survey nonresponse at the focal follow-up period used to estimate intervention impacts.

As described earlier, the attrition calculations are based on two key sets of numbers: (1) the number of youth (and clusters, if applicable) assigned to each condition; and (2) the number of youth (and clusters, if applicable) observed at follow-up. Therefore, researchers must keep track of these numbers carefully at the design and analysis phases, and understand what to do if their study is likely to fail the attrition standard. The following strategies can be used to help limit the threat of sample attrition:

- Collect follow-up data from all people assigned to condition, even if they do not complete the program or if they have a low dose of the program.
- Plan to conduct follow-up assessment using several modes to allow for multiple opportunities to gather data from respondents. Consider mailing the assessments to youth who move or providing assessments online for those absent for in-person data collection.
- Plan several days of in-person data collection at each location, to the extent possible.
- Collect extensive contact information at baseline and update this information throughout the study to enable the study team to locate follow-up nonresponders.
- When possible, conduct consent before random assignment, because nonconsent after random assignment is considered a form of attrition.
- When possible, use incentives to obtain higher response rates.

Finally, although this does not address attrition, it is good practice to collect baseline assessments of the outcome of interest, because they can be used to (1) improve precision of the impact estimate, and (2) establish baseline equivalence for the study to receive a moderate evidence rating (if the study does have high attrition).

Reviews of studies with high levels of sample attrition

If a study has problematic levels of sample attrition, that study will not be eligible to achieve the highest rating under HHS evidence standards. However, if the study establishes that the final analytic sample is equivalent at baseline on key variables that influence the outcome of interest, the study will still be eligible for a moderate rating. See the TPP Eval TA brief on **matching techniques** for recommended approaches to creating comparison groups that are equivalent on observable characteristics.

Endnotes

[1] When there are multiple outcomes to be examined and some item non-response across the outcomes, the TPP Eval TA team recommends identifying a single, common analytic sample that does not have missing data across the outcomes of interest, and using that common sample for the purposes of analysis and attrition calculations. Using a common analytic sample will produce an easy-to follow and understandable presentation of the analyses across multiple outcome measures. If, however, there is substantial item-non response across two or more outcomes, then it is recommended to consider each outcome as requiring its own, unique analytic sample, which will require multiple attrition scenarios for the various outcomes examined.

[2] The WWC has two attrition thresholds. Selection of the threshold for a particular topic is contingent on the likelihood of attrition being related to the outcome. Because many TPP programs are voluntary, the HHS evidence review selected the WWC's conservative attrition threshold, which accounts for the fact that attrition might be related to the outcomes when estimating the potential bias due to attrition. For more information on the WWC attrition standards, see the "**Assessing Attrition Bias**" white paper on the WWC website.

References

Mathematica Policy Research. "Identifying Programs That Impact Teen Pregnancy, Sexually Transmitted Infections, and Associated Sexual Risk Behaviors Review Protocol Version 3.0." Retrieved from http://tppevidencereview.aspe.hhs.gov/pdfs/Review_protocol_v3.pdf.

U.S. Department of Education, Institute of Education Sciences, What Works Clearinghouse. "Procedures and Standards Handbook Version 3.0." Retrieved from http://ies.ed.gov/ncee/wwc/pdf/reference_resources/wwc_procedures_v3_0_standards_handbook.pdf.

This brief was written by Russell Cole and Seth Chizeck from Mathematica Policy Research for the HHS Office of Adolescent Health under contract #HHSP233201300416G.

APPENDIX A:
Highest differential attrition for a sample to maintain low attrition, by overall attrition.

Overall Attrition	Differential Boundary	Overall Attrition	Differential Boundary	Overall Attrition	Differential Boundary
0	5.7	22	5.2	44	2.0
1	5.8	23	5.1	45	1.8
2	5.9	24	4.9	46	1.6
3	5.9	25	4.8	47	1.5
4	6.0	26	4.7	48	1.3
5	6.1	27	4.5	49	1.2
6	6.2	28	4.4	50	1.0
7	6.3	29	4.3	51	0.9
8	6.3	30	4.1	52	0.7
9	6.3	31	4.0	53	0.6
10	6.3	32	3.8	54	0.4
11	6.2	33	3.6	55	0.3
12	6.2	34	3.5	56	0.2
13	6.1	35	3.3	57	0.0
14	6.0	36	3.2	58	-
15	5.9	37	3.1	59	-
16	5.9	38	2.9	60	-
17	5.8	39	2.8	61	-
18	5.7	40	2.6	62	-
19	5.5	41	2.5	63	-
20	5.4	42	2.3	64	-
21	5.3	43	2.1	65	-

Source: What Works Clearinghouse. “Procedures and Standards Handbook Version 3.0.”