Addressing Attrition Bias in Randomized Controlled Trials: Considerations for Systematic Evidence Reviews

July 2015

Addressing Attrition Bias in Randomized Controlled Trials: Considerations for Systematic Evidence Reviews

July 2015

John Deke

Emily Sama-Miller

Alan Hershey

MATHEMATICA
Policy Research

≡OPRE

CONTENTS

OVERVIEW

A well-executed randomized controlled trial (RCT) can provide highly credible evidence about program efficacy, but this credibility can be weakened if there is substantial attrition (that is, people leaving the study sample). If the characteristics of the people who leave are correlated with their group status or outcomes, this correlation could create systematic differences between the remaining program and control group members. This in turn could lead to biased estimates of program effects; the risk of bias (that is, a systematic difference between the *true* program impact and its *estimated* impact on the sample of people analyzed) increases with the attrition rate.

These issues are critical for systematic evidence reviews that assess existing studies on program effectiveness, such as the Home Visiting Evidence of Effectiveness Review (HomVEE). These reviews typically focus on RCTs and other studies that are sufficiently well-designed to conclude that a program *caused* an observed effect. Attrition could introduce bias, so systematic reviews are concerned with knowing its level in a study relative to a tolerable level in order to assess the validity of RCTs.

HomVEE uses an attrition standard adapted from the Department of Education's What Works Clearinghouse (WWC), another systematic evidence review. This standard establishes tolerable rates of attrition for the RCTs reviewed by HomVEE.[1] The standard is a boundary between high and low rates of overall attrition and differential attrition (the difference in the rates of sample loss for the program and control groups). Attrition rates above this boundary yield an unacceptably high bias. For these reviews, the maximum acceptable bias is 0.05 standard deviations.

HomVEE's population of interest includes pregnant women, and families with children age birth to kindergarten entry; the population is different than the school-age children whose test scores were the basis of the attrition standard for the WWC. Therefore, we conducted a statistical exercise to examine how the HomVEE attrition boundary would respond to changes in two fundamental assumptions: (1) the correlation between outcomes and attrition and (2) the level of attrition bias deemed "acceptable." Data came from the Early Head Start Research and Evaluation Project, in which 7 of 17 sites delivered Early Head Start primarily via home visits, and from effect sizes reported in HomVEE-reviewed studies through September 2014.

The results suggest two main conclusions. First, the HomVEE attrition boundary is relatively insensitive to changes in the assumed correlation between outcomes and attrition. Second, the attrition boundary is sensitive to how HomVEE defines an "acceptable" level of attrition bias. Specifically, when small impacts matter, small biases (possibly resulting from attrition) also matter.

As a principle of well-executed social science research, researchers attempting to detect small impacts must also worry about small biases and should strive for the lowest possible attrition rate in their studies, including rates lower than those permitted by HomVEE standards.

---

[1] For more information, visit the HomVEE website (http://homvee.acf.hhs.gov/) and the WWC website (http://ies.ed.gov/ncee/wwc/).

I.   INTRODUCTION

A well-executed randomized controlled trial (RCT) can provide highly credible evidence about program efficacy. Because study groups are formed randomly, researchers can assume the groups are equivalent on average in all respects except that only one group receives a program that is being tested. Therefore, any statistically significant differences between the outcomes of the groups at the end of an evaluation can be attributed to the program rather than to other factors.

The credibility of RCTs, however, can be weakened if there is substantial attrition. If characteristics of people who leave the sample are correlated with their group status or outcomes, the correlation may point to systematic differences between the remaining program and control group members. This could lead to biased estimates of program effects. Because researchers typically cannot fully understand why some sample members leave they may not know whether or how the leavers' characteristics are related to their group status or their outcomes. Thus, as the attrition rate rises, so does the potential for bias.

These issues are critical for the Home Visiting Evidence of Effectiveness Review (HomVEE). HomVEE focuses on studies of home visiting programs targeting pregnant women and families with children from birth to kindergarten. The review identifies, assesses, and rates the rigor of impact studies of home visiting programs that serve this population. HomVEE focuses on studies that are sufficiently well designed to estimate programs' effects, apart from other factors that may influence the target population.

HomVEE uses an attrition standard adapted from the Department of Education's What Works Clearinghouse (WWC). This standard establishes tolerable rates of attrition for the RCTs reviewed by HomVEE.[2] In this paper, we discuss whether an attrition standard based on information from education research is appropriate for use with the research that HomVEE examines. The paper also provides an example of how to assess whether the attrition standard for one systematic evidence review fits other systematic reviews, along with considerations for adopting or modifying the standard for alternative contexts.

II. BACKGROUND

RCT studies use a rigorous design and *may* receive the highest possible ratings of causal validity in both the HomVEE and WWC reviews.[3] However, excessive sample attrition will preclude such a rating. An RCT with high attrition (as defined by the attrition standard described below) can receive no more than a "moderate" rating in HomVEE. Further, high-attrition RCTs may earn that mid-level rating only if researchers can show that selected characteristics of the

---

[2] For more information, visit the HomVEE website (http://homvee.acf.hhs.gov/) and the WWC website (http://ies.ed.gov/ncee/wwc/).

[3] The study-level ratings—(1) high, (2) moderate, and (3) low—provide a measure of the review's degree of confidence that the study design could provide unbiased estimates of model impacts.

program and control groups were equivalent at baseline (before enrollment into the program).[4] Therefore, the rate of attrition is an important consideration when these reviews assess the validity of RCT studies.

The attrition standard, described in the next section relies on two measures of attrition— overall and differential—to assess whether the attrition rate is large enough to introduce an unacceptable level of bias (that is, a systematic difference) between the *true* and *estimated* impact on the analytic sample. The measures are defined as follows:

- **Overall attrition rate:** the proportion of sample members randomly assigned to the study groups for whom outcome data are *not* available

- **Differential attrition rate:** the difference in attrition rates between the program and control groups

Random assignment produces groups that are similar on average, but overall and differential attrition may lead to groups that have different baseline characteristics. This can result in biased conclusions about a program's efficacy. In other words, the evaluation may capture the impact of the characteristics that differ between the groups in addition to the program's impact. Researchers can neither observe nor statistically control for all possible factors that lead to attrition, so it is difficult or impossible to eliminate this type of bias when calculating effects.[5]

WWC's and HomVEE's attrition standards are not concerned with sample loss that is exogenous (unrelated to an individual's random assignment status) because it does not introduce bias. For example, researchers facing budget constraints may collect follow-up data from only some randomly assigned sample members. This is acceptable as long as the researchers randomly choose the sample members for follow-up. Excluding sample members randomly does not introduce bias because doing so is clearly unrelated to their treatment status.

Conversely, losing sample members because of nonrandom events that occur after random assignment is problematic because it may introduce bias. For example, if sample members are asked to consent to the evaluation after random assignment, the loss of people who do not consent introduces the possibility of bias because sample members might decide whether to consent based on their group assignment. This can lead to differences between the groups in the number and characteristics of sample members who leave the study. If this happens, there is a bigger risk that a significant difference in outcomes between the program and control groups will be attributed to the program when in fact it is due to characteristics that caused sample members to consent or refuse.

The WWC has consistently recognized that attrition can undermine the estimates of an RCT, but its specific attrition standard has evolved. The original standard consisted of cutoff values for

---

[4] See http://homvee.acf.hhs.gov/Review-Process/4/Producing-Study-Ratings/19/5 and http://ies.ed.gov/ncee/wwc/pdf/reference_resources/wwc_procedures_v2_1_standards_handbook.pdf) for details.

[5] Readers seeking more information on the concept of attrition bias and strategies for mitigating it might find helpful explanations in a recently released brief prepared by HomVEE staff: http://homvee.acf.hhs.gov/HomVEE_brief_2014-49.pdf.

overall and differential attrition rates.[6] The cutoff for the overall attrition rate was similar to the survey response rates targeted by federal agencies such as the Office of Management and Budget (OMB) and the National Center for Educational Statistics (NCES).[7] OMB, NCES, and WWC choose these cutoffs out of concern about nonresponse bias, but there is no theoretical or empirical evidence that such cutoffs limit bias. The WWC Statistical, Technical, and Analytical Team (STAT) therefore developed a model of attrition and analyzed data from past education-related RCTs to help select parameter values for that model.[8] Essentially, the model estimates the expected attrition bias, using assumptions about the relationship between outcomes and the intrinsic likelihood that sample members will leave the sample. The STAT used the model to calculate the expected bias for every combination of overall and differential attrition rates. The combinations along the boundary represent the maximum acceptable level of bias. WWC and HomVEE now refer to this maximum acceptable level as the attrition standard. The next section more fully explains how the standard was developed.

III. THE ATTRITION STANDARD

The attrition standard for WWC and HomVEE is a boundary between high and low rates of overall and differential attrition (Figure 1). Attrition rates above this boundary (the red area of the figure) yield an unacceptably high bias; rates in the green area are acceptably low. Low overall and differential attrition rates are preferable, but there is some flexibility. High rates of overall attrition may be acceptable when the differential attrition rate is very low, and the reverse is true as well.

---

[6] WWC reviews are organized by topic areas such as literacy. Originally, the principal investigator for each topic area selected the attrition rate cutoff value for the studies reviewed in that topic area. These cutoffs ranged from 20 to 30 percent for the overall attrition rate and from 5 to 10 percent for the differential attrition rate.

[7] OMB's target response rate is 80 percent for data collection (http://www.whitehouse.gov/omb/inforeg/pmc_survey_guidance_2006.pdf), and NCES's target response rate is 85 percent (http://nces.ed.gov/pubs2003/2003601.pdf). Because OMB and NCES guidelines focus on general data collection (not just RCTs), they do not include targets for the differential attrition rate.

[8] Dr. John Deke led the development of the WWC attrition model and the accompanying attrition standard, which is described in Appendix A of the *What Works Clearinghouse Procedures and Standards Handbook*, version 2.1 (http://ies.ed.gov/ncee/wwc/pdf/reference_resources/wwc_procedures_v2_1_standards_handbook.pdf); the model is presented in the WWC white paper on assessing attrition bias: http://ies.ed.gov/ncee/wwc/pdf/reference_resources/wwc_attrition_v2.1.pdf.

Figure 1. WWC and HomVEE attrition bounds
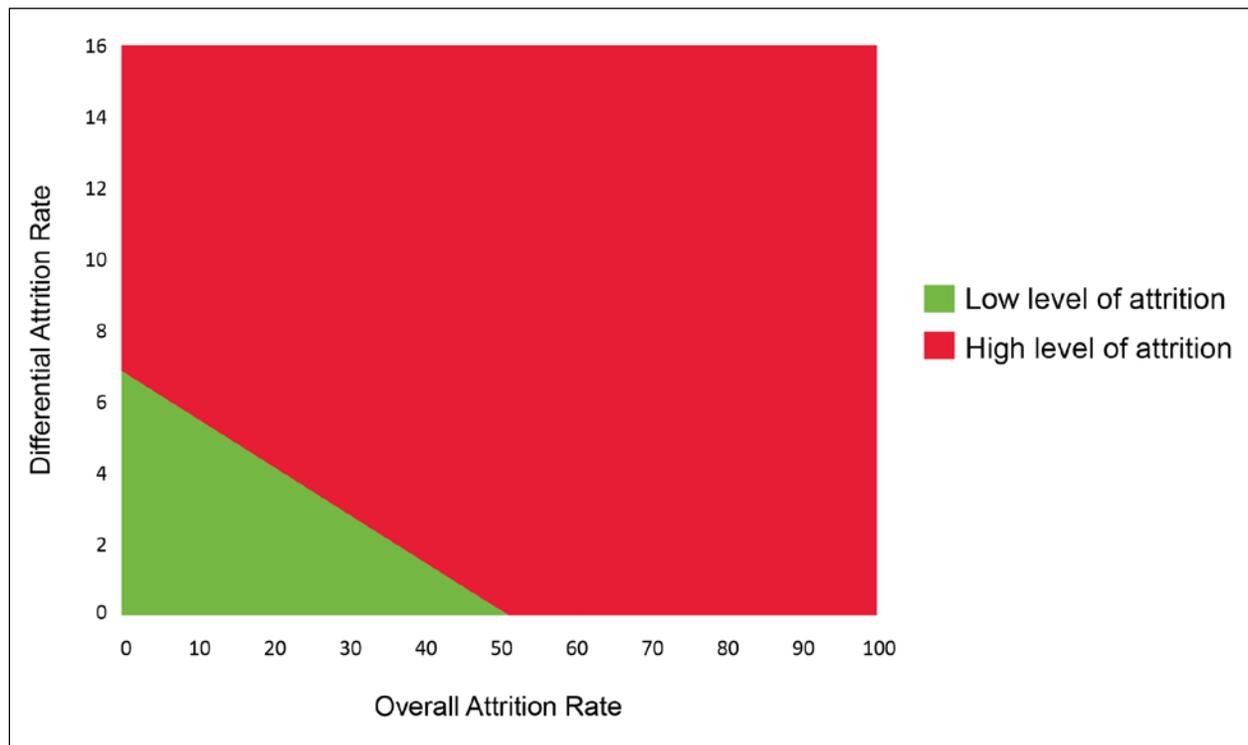
Note: This is a stylized illustration of attrition boundaries, so the dividing line in the diagram may not precisely reflect the calculated boundary. Table 1 provides specific values that lie on the boundary.

To more precisely illustrate the concept in Figure 1, Table 1 presents some specific attrition rates that lie on the boundary. The left column lists overall attrition rates, and the right column lists the differential attrition rate that corresponds to the placement of the attrition standard boundary for that overall attrition rate.

Table 1. Maximum acceptable rate of differential attrition for each overall attrition rate

Overall attrition rate	Maximum acceptable rate of differential attrition
0.10	0.063
0.15	0.059
0.20	0.054
0.25	0.048
0.30	0.041
0.35	0.033
0.40	0.026
0.45	0.018
0.50	0.010
0.55	0.003

Source: Authors' calculations using the attrition model with WWC's "conservative" parameter assumptions (which are also used by HomVEE).

Selecting the position of the attrition boundary depends on complex decisions about:

- The statistical model that defines the boundary
- The assumptions made about the relationship between outcomes and attrition
- The amount of bias considered tolerable

Making realistic assumptions about how attrition relates to outcomes is challenging because the outcomes for sample members who leave are, by definition, unknown. Researchers therefore cannot directly observe the relationships of interest. Instead, the WWC STAT indirectly estimated this relationship using data from past RCTs in education. The team used academic pre-test scores (which were available for all randomized students in the past studies) as a proxy for post-test scores (which were missing for some of the randomized sample). The STAT assumed that the relationship between the attrition rate and pre-test scores was the same as the relationship between the attrition rate and post-test scores. This was because academic pre- and post-test scores are often highly correlated.[9]

Researchers must also decide what level of attrition to allow. The STAT addressed this challenge by selecting a level that was low relative to WWC's definition of a substantively important impact (0.25 standard deviations of the outcome variable). The team defined maximum acceptable bias as one-fifth of a substantively important impact, or 0.05 standard deviations. Attrition bias exceeding this level of 0.05 is considered unacceptably high.

The STAT used data from education research to develop two attrition boundaries for WWC: optimistic and conservative. WWC uses the optimistic boundary when it seems reasonable to assume that attrition is only weakly related to the program and outcomes. This boundary is consistent with actual correlations between the attrition and pre-test scores observed in RCTs involving children in grades 1 through 9, as well as in curricular interventions the STAT examined when making assumptions about the attrition model parameters. WWC uses the conservative boundary when attrition is likely to be strongly related to the program and outcomes—for example, in studies focusing on drop-out prevention. The likelihood of dropping out of school may be highly correlated with the likelihood of attrition during follow-up data collection.[10]

HomVEE adopted the conservative boundary from WWC, and does not use the optimistic boundary.

---

[9] The correlation between academic pre-tests and post-tests is typically 0.7 to 0.8 (Bloom et al. 2005; Schochet 2008).

[10] Principal investigators for WWC topic areas may choose which boundary should be used for reviews in their areas. See http://ies.ed.gov/ncee/wwc/pdf/reference_resources/wwc_procedures_v3_0_standards_handbook.pdf, page 12. Once a principal investigator selects a boundary, it is used for all RCTs in that topic area.

IV. ARE WWC'S ASSUMPTIONS ABOUT ATTRITION SUITABLE FOR HOMVEE?

HomVEE and WWC focus on different research populations, and thus HomVEE might need to choose different attrition model parameters or cutoffs for acceptable attrition bias. Would doing so change the attrition boundary? If the boundary is affected by small changes in the parameter or cutoff values, it may be appropriate to assess whether the WWC-based values are appropriate for HomVEE. If the boundary is not affected by changes in the two factors, we can be more confident about HomVEE's existing evidence standards. Our approach to testing each of the two factors, while leaving the other unchanged, is described below.

A. Examining the parameter values for the attrition model

We cannot observe (and therefore must assume) the correlations between outcomes and the likelihood of attrition, but what if our assumptions are wrong? To answer this question, we conducted two analyses. First, we used real data from HomVEE to examine the WWC Statistical, Technical, and Analytical Team's assumption that pre-test measures are reasonable proxies for post-test measures when defining parameter values for the attrition model. Second, we conducted a thought experiment testing a wide range of parameter values to examine whether and how they affect the boundary.

Are baseline variables a good proxy for outcomes in an early childhood setting? The selection of attrition model parameters for WWC was informed, in part, by analyses of data from RCTs in education research, a field with a generally high (70 to 80 percent) correlation between pre- and post-tests. To assess whether similar analyses could inform the selection of parameter values for the HomVEE attrition model, we analyzed data from the Early Head Start Research and Evaluation Project (EHSREP; Love et al. 2002), one of the largest experimental evaluations of an early childhood intervention. Seven of 17 EHSREP sites delivered Early Head Start primarily via home visits.

We examined 70 outcomes (Table A.1) from several areas of interest to HomVEE: (1) child health, (2) maternal health, (3) cognitive development,[11] (4) social-emotional development, (5) parenting, (6) family economic self-sufficiency, (7) family violence, and (8) linkage and referrals to other services in the community. We used 58 baseline variables (Table A.2) to predict each outcome using a regression. The adjusted R^2 statistic resulting from the regression quantified the proportion of the variation in each outcome that these baseline variables predicted. Across the 70 regressions, the adjusted R^2 ranged from 0 to 0.14, with a median of 0.04. That is, the baseline variables usually explained less than 5 percent of the variation in the outcome.

We found that baseline variables, therefore, likely *do not* reliably predict outcomes in early childhood research; knowing baseline scores is less useful for predicting outcomes for young children than for older children. Only 10 of the 70 regressions indicated (through their adjusted R^2) that at least 10 percent of the variation in the outcome was explained by the baseline variables. The STAT successfully used the correlation between baseline scores and the attrition

---

[11] HomVEE uses a single outcome domain to encompass children's cognitive and social-emotional development, but EHSREP considered these separately.

rate to approximate the correlation between outcomes and attrition, but these findings make us less confident that we could do the same in HomVEE.

Given our findings, we do not recommend conducting a WWC-style analysis to select parameter values for a HomVEE attrition model. Unless we can find better predictors through additional research, it does not make sense to refine the attrition model to apply specifically to the HomVEE context. Instead, we next examined how a range of hypothetical parameter values affect the boundary.

How do relatively large changes in assumptions about the model parameters affect the attrition boundary? Increasing the correlation between attrition and outcomes increases bias. So does increasing the difference in correlation between the program and control groups. But to what extent is bias affected by these correlations? We tested how various correlations might affect the level of bias by using a range for correlations for each assumption.

We used the formulas for the WWC attrition model to calculate the average attrition bias at the boundary. These calculations were based on three alternative assumptions (which could be characterized as weak, moderate, and strong) about the correlations between the attrition and outcomes and the difference in these correlations between the program and control groups (Figure 2). Specifically, we combined:

- Three assumptions of the overall correlation between attrition and outcomes ($r = 0.22$, 0.42, and 0.62) with

- Three assumptions of the difference in that correlation between the program and control groups ($r = 0.03$, 0.06, and 0.12)

For example, in combining the assumption of $r = 0.22$ with $r = 0.12$, we assumed a relatively weak relationship between the attrition rate and outcomes *and* a relatively strong difference in that relationship between the program and control groups. This hypothetical situation could occur if baseline measures of outcomes strongly predict attrition in *either* the program or control group, but does not predict attrition very well across the groups. The current HomVEE attrition boundary corresponds to the moderate level in our thought experiment, with an overall correlation of $r = 0.42$ and a difference of $r = 0.06$.

Figure 2. Expected bias at the standard attrition boundary in HomVEE for different attrition-to-outcome correlations
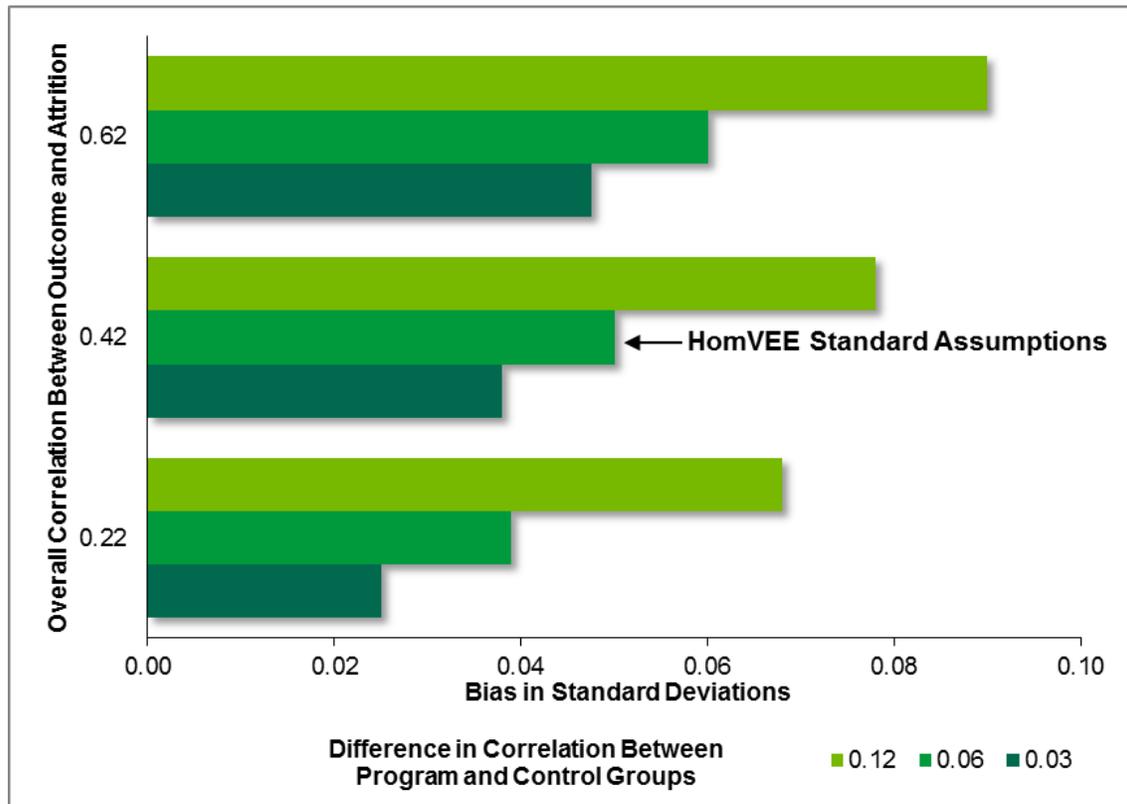
We found, as we explain below, that relatively large changes from the HomVEE standard assumptions in both the overall and differential correlations do little to change the expected bias at the boundary. This is reassuring because we are relatively uncertain about how strongly the attrition rate and outcomes are correlated in the HomVEE context. Some examples (illustrated in Figure 2) are as follows:

- Doubling the difference in the attrition-to-outcome correlation between the program and control group from $r = 0.06$ to $r = 0.12$ increases the expected bias along the boundary from 0.05 to 0.08 standard deviations.

- If we hold constant the difference in attrition-to-outcome correlation between program and control groups and increase the overall correlation of attrition to outcomes from $r = 0.42$ to $r = 0.62$, the expected bias along the boundary increases from 0.05 to 0.06 standard deviations.

B. Checking how much our definition of "acceptable" bias affects the boundary

Setting the level of acceptable attrition bias at 0.05 standard deviations (the conservative boundary used in many WWC topic areas and in HomVEE) may not always be right. It may not be suitable, for instance, if small impacts of an intervention are meaningful. The boundary is based on 0.25 standard deviation effect size. WWC defines this as substantively important based

on education literature. In the HomVEE context, which uses other outcomes that may have other typical effect sizes, a different definition of a "substantively important impact" may therefore be appropriate. This would require changing the level of acceptable bias and thus shifting the attrition boundary.

We used HomVEE data on effect sizes to examine typical effect sizes in the home visiting literature. HomVEE reports outcomes, including effect sizes, from studies with a research design of at least moderate quality, and groups these outcomes by domain (a topic area, such as child health or child development and school readiness). The reviewers list author-reported effect sizes or calculate the effect sizes themselves, if possible. Although the method for calculating effect sizes may differ by study or outcome, examining the effect sizes still provides useful information about the magnitude of reported impacts and what magnitude may be considered "substantively important" in each outcome domain.

In a HomVEE review, effect sizes vary by domain but are typically well below 0.25 (Figure 3). The median effect size for seven of the eight domains is less than 0.25 (although the mean effect size tends to be larger than 0.25, reflecting some very large effects within certain domains). Across the domains, however, the average can vary. For instance, the average of 0.89 standard deviations in the Linkages and Referrals domain is six times larger than the average effect size of 0.15 standard deviations in the Reductions in Child Maltreatment domain. This difference likely reflects the fact that some changes (such as modifying parenting behaviors) are harder to achieve than others (such as linking families to other services).

The interpretation of a "substantively important" effect size may therefore vary by domain or even by outcome, so an absolute level of acceptable bias may not be appropriate. For example, suppose a difference of 0.10 standard deviations is considered substantive for an outcome such as substantiated child abuse. A bias of 0.05 is half the size of the substantive effect and may be too large to be considered acceptable.[12] Furthermore, decision makers may use criteria other than effect size to determine whether an effect is meaningful. For instance, in a low-cost intervention, a small effect on an important outcome (such as infant mortality or child maltreatment) may be meaningful to a funder or policymaker. A similarly sized effect on a less crucial outcome (such as whether families are referred to other services) stemming from a costlier intervention may be less meaningful.

---

[12] Put differently, suppose the true effect is 0.05 of a standard deviation. With a bias of 0.05 standard deviations, the estimated effect could be almost twice as high as the true effect (0.10).

Figure 3. Average and median effect sizes in HomVEE, by domain
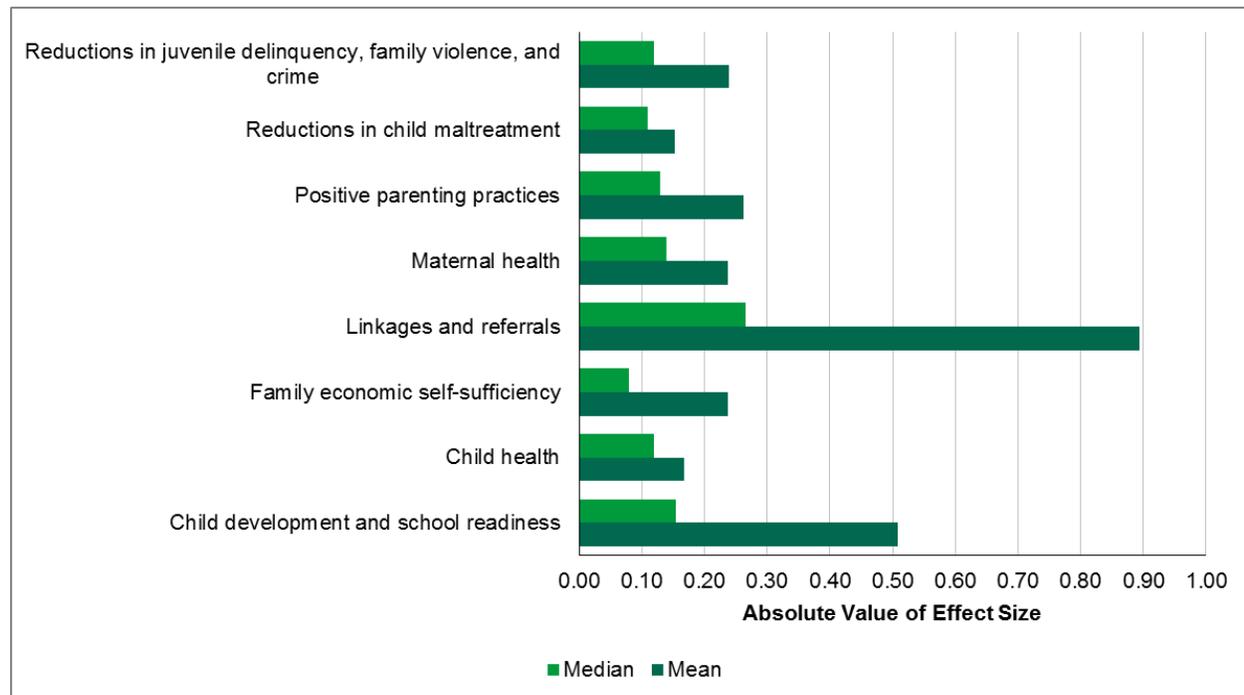
Source: HomVEE review data through September 2014.

We examined how the conservative attrition boundary would change if we altered the definition of "acceptable bias" while holding constant the underlying assumptions about correlations between attrition, the program, and outcomes in the program and control groups. We calculated where the boundary would be for levels of acceptable bias equal to 0.025, 0.05, 0.075, and 0.10 standard deviations (Figure 4). The current HomVEE boundary is that between the yellow (second from the left) and orange (third from the left) regions in the figure.

We concluded that relatively small changes from the HomVEE standard assumptions about the acceptable level of attrition bias substantially change the location of the attrition boundary. For example, reducing acceptable bias from 0.05 to 0.025 standard deviations shifts the boundary inward (Figure 4), to the line between the green (first on the left) and yellow (second from the left) regions. At this level, only low attrition rates would be acceptable. This is a particularly important consideration for researchers conducting studies in which impacts smaller than 0.25 standard deviations are substantively important.

Figure 4. Attrition bounds when acceptable bias varies and correlations are held constant
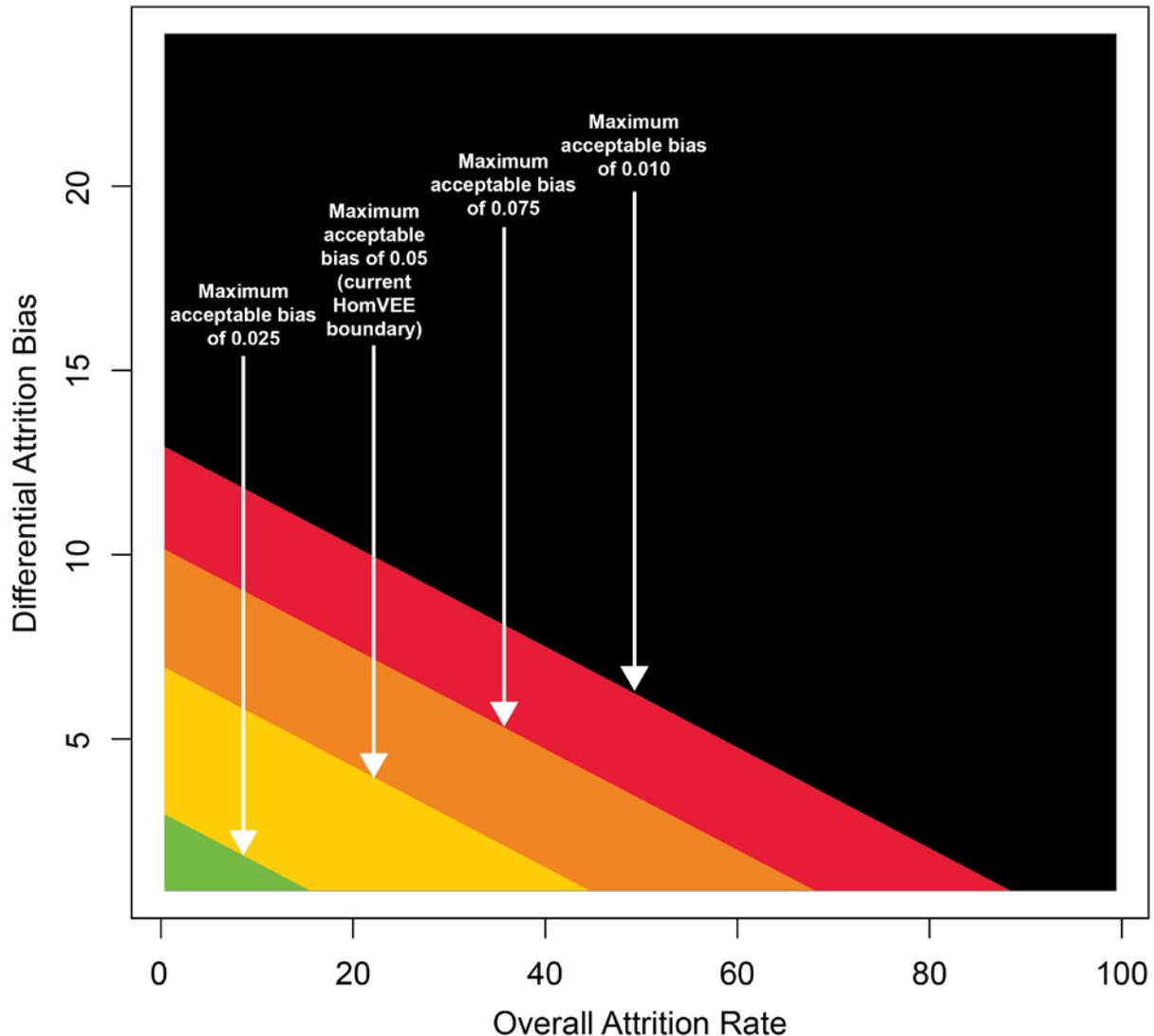
Note: This is a stylized illustration of attrition boundaries, so the dividing lines in the diagram may not precisely reflect the calculated boundaries.

V. DISCUSSION

In this paper we examined how the HomVEE attrition boundary would respond to changes in two fundamental assumptions: (1) the correlation between outcomes and attrition and (2) the level of attrition bias deemed "acceptable."

We found that the HomVEE attrition boundary is relatively insensitive to changes in the assumed correlation between outcomes and attrition. When we made relatively large changes to the HomVEE standard assumptions about the correlation between attrition and outcomes, or about the difference in this correlation between groups, the expected bias at the attrition boundary barely changed. This finding is reassuring because, unlike in the education

context, we found a weak relationship between pre- and post-intervention measures of outcomes in early childhood. In other words, the fact that we can't predict the size of the correlations between key variables isn't necessarily a danger to our HomVEE assumptions.

However, we also found that the attrition boundary is sensitive to how HomVEE defines an "acceptable" level of attrition bias. As Figure 4 illustrates, attrition bounds can change drastically as the maximum level of bias changes. This suggests the need for a much tighter boundary on permissible rates of sample attrition in contexts where small impacts are "substantively important." That is, when small impacts matter, small biases also matter.

Given these findings, HomVEE (and other evidence reviews) might consider using different definitions of acceptable bias—and therefore different attrition boundaries—in different contexts, but refining this definition may prove challenging. A single definition ignores the considerable variation in effect sizes across outcomes and domains that we have observed in studies HomVEE reviewed, and disregards the context-dependent nature of designating an impact as "substantively important." Yet, setting multiple attrition boundaries may cause confusion and conflict with HomVEE's need to remain transparent and consistent. Furthermore, additional research would be needed to determine how to implement the variation in attrition boundaries so as to minimize the possibility of applying an overly restrictive boundary. HomVEE must determine how to balance some stakeholders' need for clear, consistent definitions against others' need to know which programs have sufficiently strong evidence of effectiveness in key domains.

Regardless of how HomVEE and other evidence reviews ultimately address the sensitivity of the ideal attrition boundary to study context, the implication of these findings for researchers is clear. As a principle of well-executed social science research, researchers attempting to detect small impacts must also worry about small biases and should strive for the lowest possible attrition rate in their studies, including rates lower than the maximum permitted by HomVEE standards.

REFERENCES

Bloom, H., L. Hayes, and A. Black. "Using Covariates to Improve Precision." New York: MDRC, 2005.

Love, John M., Ellen E. Kisker, Christine M. Ross, Peter Z. Schochet, Diane C. Paulsell, Kimberly Boller, Jill M. Constantine, and Cheri A. Vogel. "Making a Difference in the Lives of Infants and Toddlers and Their Families: The Impacts of Early Head Start." Washington, DC: U.S. Department of Health and Human Services, Administration for Children and Families, June 2002.

Schochet, P. "Statistical Power for Random Assignment Evaluations of Education Programs." *Journal of Educational and Behavioral Statistics*, vol. 33, 2008, pp. 62–87.

APPENDIX

Table A.1. Descriptions of outcome variables, by domain

Outcome description	Number of variables[a]	Variable type
Cognitive development		
Bayley MDI score	3	Continuous
ECLS-K Fifth-Grade Math Assessment	1	Continuous
PPVT standard score	1	Continuous
PPVT-III standard score	2	Continuous
WJ Applied Problems standard score	1	Continuous
Social-emotional development		
BBRS: Emotional Regulation (measures child's cooperation with the interviewer)	3	Continuous
BBRS: Orientation/Engagement (measures child's ability to change tasks and test materials)	3	Continuous
SDQ Anger/Distractibility Subscale	1	Continuous
SDQ Peer Relations Subscale	1	Continuous
SDQ Sad/Lonely/Anxious Subscale	1	Continuous
Parenting		
NCATS child total (measures the skills the child brings to the parent-child interaction)	1	Continuous
NCATS parent total (measures the skills the parent brings to the parent-child interaction)	1	Continuous
NCATS total (the sum of the child and parent measures)	1	Continuous
Parent Supportive Presence Puzzle score	1	Continuous
Parent supportiveness/cognitive stimulation (observation of parent-child interaction during Three-Bag Assessment)	1	Continuous
Safety measures (e.g., home has working smoke alarms)	18[b]	Binary and continuous
Bedtime routine (child has a regular bedtime routine)	3	Binary
Read daily (parent reads to child at least once a day)	3	Binary
Child health		
Child health status (parent-reported five-point scale)	2	Categorical
Well-baby visits (child has had recommended number of well-baby check-ups)	2	Binary
Child received any immunizations	1	Binary
Number of times child has stayed overnight in a hospital	3	Continuous
Hospitalizations due to accident or injury	2	Binary
Maternal health		
Mother's health status (self-reported five-point scale)	3	Categorical
Number of births	1	Continuous
CES-D Short Form Scale	2	Continuous
Linkages and referrals		
Parent received any employment services	1	Binary
Parent received any housing services	1	Binary
Parent received education services	1	Binary
Family received mental health services	1	Binary
Respondent received transportation services	1	Binary
Family violence		
Child witnessed violence	2	Binary and continuous
Family economic self-sufficiency		
Income per month	1	Continuous

Source: Early Head Start Research and Evaluation Project.

[a]Some outcomes have several variables because they were collected during multiple follow-up waves.

[b]This outcome has one composite variable and 17 binary safety measures.

BBRS = Bayley Behavior Rating Scale; CES-D = Center for Epidemiologic Studies-Depression; ECLS-K = Early Childhood Longitudinal Study, Kindergarten; MDI = Mental Developmental Index; NCATS = Nursing Child Assessment Teaching Scale; PPVT = Peabody Picture Vocabulary Test; SDQ = Strengths and Difficulties Questionnaire; WJ = Woodcock-Johnson.

Table A.2. Descriptions of baseline variables, by domain

Baseline description	Number of variables	Variable type
Family and household characteristics		
Mother's characteristics		
Age (at random assignment, at birth of child)	2	Continuous
Race	1	Categorical
English as primary language	1	Binary
English language ability	1	Categorical
Pregnant with focus child	1	Binary
Child's characteristics		
Age	1	Continuous
Gender	1	Binary
Firstborn	1	Binary
Number of children in the household (under age 5 and from 6 to 17)	2	Continuous
Number of adults in the household	1	Categorical
Adult male in household	1	Binary
Number of moves in the past year	1	Continuous
Parenting		
Applicant was previously in Head Start or child development program	1	Binary
Child health		
Any prenatal care	1	Binary
Trimester began prenatal care	1	Categorical
Child's weight at birth	1	Continuous
Child weighed less than 2,500 grams at birth	1	Binary
Child was born more than three weeks early	1	Binary
Stayed in hospital after birth	2	Binary
Concerns about health and development	1	Binary
Evaluated on health and development	1	Binary
Child had health risks (established, medical/biological, environmental, any)	4	Binary
Child covered by health insurance	1	Binary
Maternal health		
Possibly depressed (based on CES-D scale)	3[a]	Binary
Family economic self-sufficiency		
Received services: AFDC/TANF, Medicaid, WIC, Food Stamps,[b] SSI, public housing assistance, welfare	7	Binary
Poverty level and income as percentage of poverty level	2	Continuous
Inadequate supplies or services (e.g., food, housing)	8	Binary
Owns home	1	Binary
Primary caregiver's living arrangements	1	Categorical
Primary caregiver's educational attainment	1	Categorical
Primary caregiver's occupation	1	Categorical
Maternal demographic risks (e.g., not employed, in school or training)	5[c]	Binary and categorical

Source: Early Head Start Research and Evaluation Project.

[a]CES-D Long Form, Short Form, and dummy variable for possibly being depressed, based on either form.

[b]This study took place before the October 2008, when the program name changed from Food Stamps to the Supplementary Nutrition Assistance Program.

[c]One composite variable and four binary risks.

AFDC = Aid to Families with Dependent Children; CES-D = Center for Epidemiologic Studies-Depression; SSI = Supplemental Security Income; TANF = Temporary Assistance for Needy Families; WIC = Special Supplemental Nutrition Program for Women, Infants, and Children.