September 2017

Should Teen Pregnancy Prevention Studies Randomize Students or Schools? The Power Tradeoffs Between Contamination Bias and Clustering

Evaluators of teen pregnancy prevention (TPP) programs implemented in a school environment face a difficult tradeoff in selecting the level of randomization that will give their evaluations the best chance of detecting program effects. If schools are randomized, then the study's ability to detect program impacts (that is, the study's statistical power) is reduced by larger standard errors resulting from clustering.[1] If individual youth are randomized within schools, then the study's power is potentially reduced by attenuation bias that can occur when members of the program group date members of the control group (contamination bias).

In this brief, we seek to quantify this tradeoff to help evaluators choose the best unit of randomization in different evaluation contexts to maximize study power. Using simulations, we assess the tradeoffs between contamination bias and clustering effects across evaluation contexts defined by (1) the prevalence of dating, (2) the prevalence of risky sexual behaviors, (3) the difference in the prevalence of risky sexual behaviors between those who are and are not dating other study sample members, (4) the magnitude of clustering effects (as measured by the intraclass correlation [ICC]), and (5) number of schools in the study.

In general, it is better to randomize individuals rather than schools when: (1) the prevalence of dating is low, (2) the magnitude of clustering effects is large, (3) the prevalence of risky sexual behaviors is high, and (4) the difference in the prevalence of risky sexual behaviors between those who are and are not dating is low. We provide more specific guidance for identifying the best design for different contexts. We cannot make a universal recommendation favoring one design over the other – there are realistic contexts favoring each design. We also note that while this brief is focused on statistical power, other considerations may be relevant in some contexts. For example, if researchers face no constraints on the size of the study, then statistical power would not be a consideration and other issues like face validity could take precedence. In that case, randomization of schools would always be preferred.

How Dating Could Lead to Contamination Bias

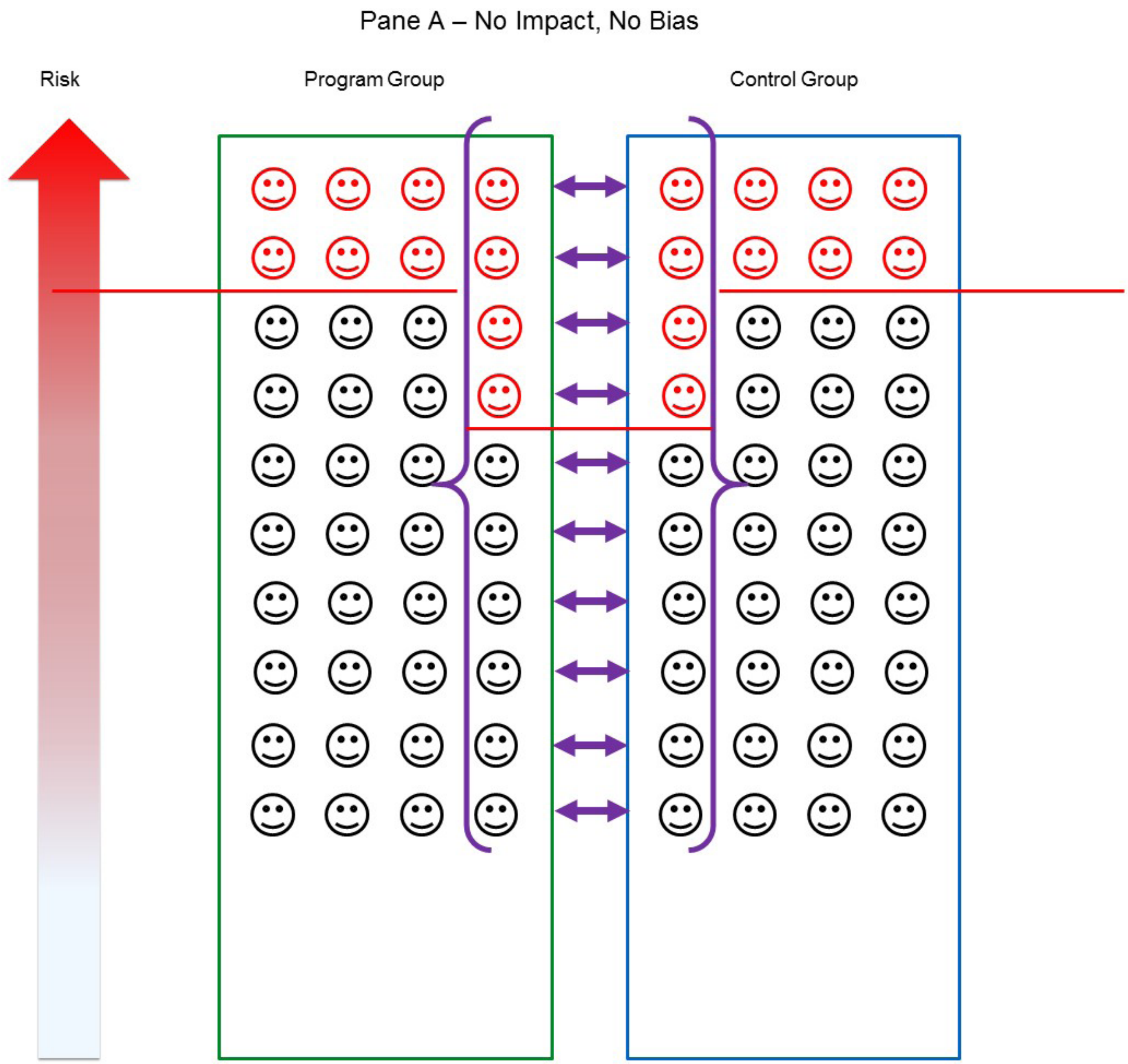
Dating relationships between students in the program and control groups create the potential for contamination bias in studies of TPP programs that goes beyond the usual concern about contamination from peer effects in studies that randomize students within schools.[2] This is because (1) the outcomes targeted by TPP programs are intrinsically germane to dating relationships and (2) individuals in a dating relationship may have considerable influence over one another. For example, an abstinence program that persuades youth in the program group to abstain from sex almost unavoidably affects the sexual behavior of their dating partners, which would lead to contamination bias when some of those partners are in the control group.[3] Similarly, if one member of a dating couple is persuaded to use a condom, the other member of the couple will also most likely be affected by that decision.[4]

The basic logic model for how dating leads to contamination bias is illustrated in Figure 1. The figure shows a program and control group, with the individuals in the groups sorted by their inclination to engage in a risky behavior like unprotected sex (people who are most inclined to engage in risky behavior are at the top). The goal of the program is to reduce the proportion of people who engage in risky behavior.

In the figure, people who do engage in the risky behavior are represented by red faces while people who do not end up engaging in risky behavior are represented by black faces. The people inside the purple brackets are the ones who are dating, and the purple arrows indicate who is dating whom. The red horizontal lines in the figure represent the risk threshold that people must cross before engaging in risky behavior. This threshold is assumed to be lower for people who are dating because we presume that people who are dating have more opportunities to engage in risky sexual behavior.[5]

In pane A of Figure 1 we present a scenario in which the program has no impact on the inclination to engage in risky behavior. We can see that there is no impact because the average inclination to engage in risky behavior is the same in both groups and, similarly, the proportion of people who actually do engage in risky behavior is also the same (that is, the number of red faces is the same in both the program group and the control group).

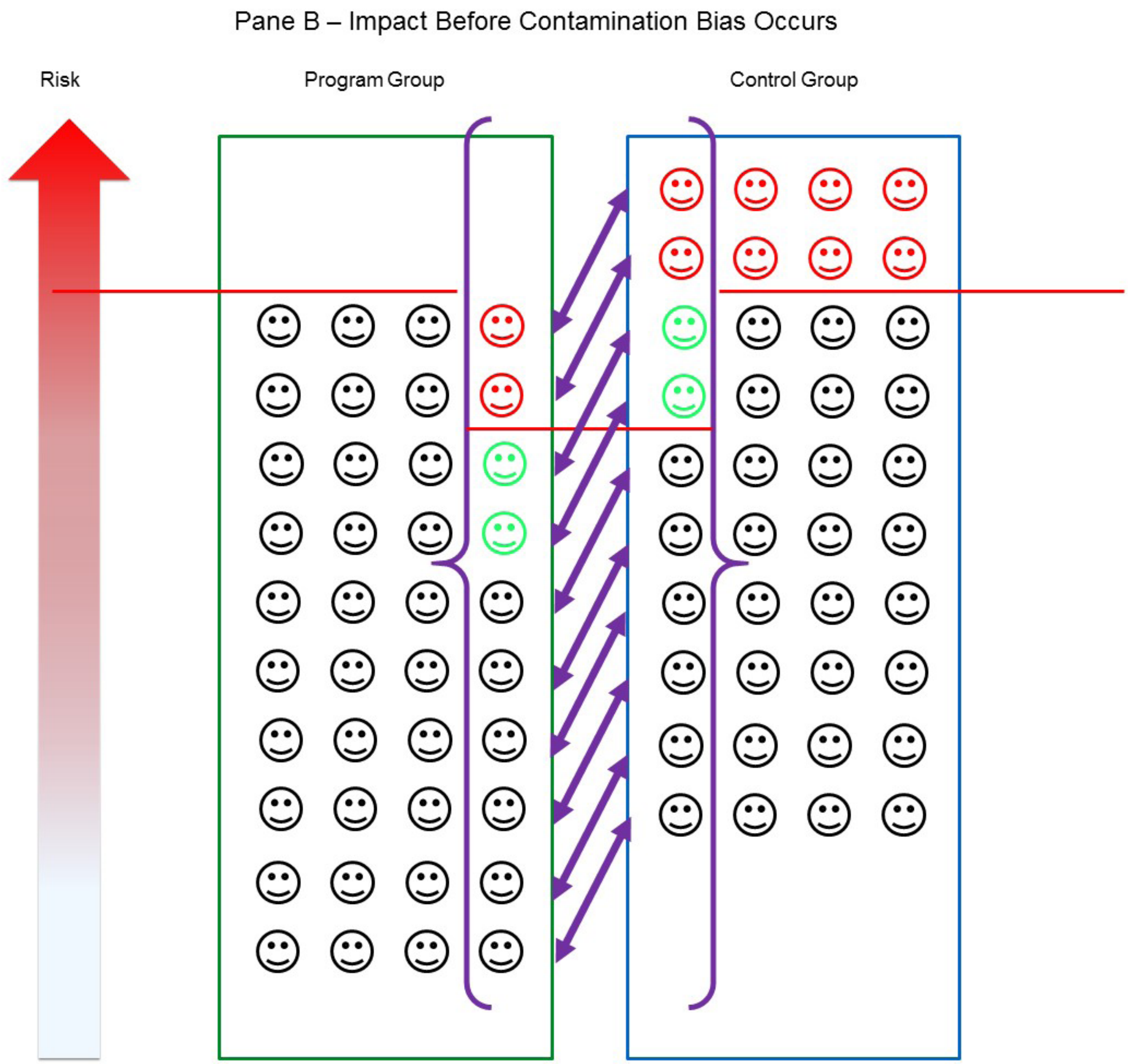
Figure 1, Pane A: A model of how dating can lead to contamination bias.



In pane B of Figure 1 we present a scenario in which the program reduces the inclination to engage in risky behavior, but the contamination effects are not yet reflected. We can see the impact because the people in the program group box are located closer to the bottom of the box (i.e. each individual's propensity to engage in risky behavior has declined) while people in the control group box are closer to the top. We also see that there are fewer red faces (fewer people who engage in risky behavior) in the program group than in the control group.

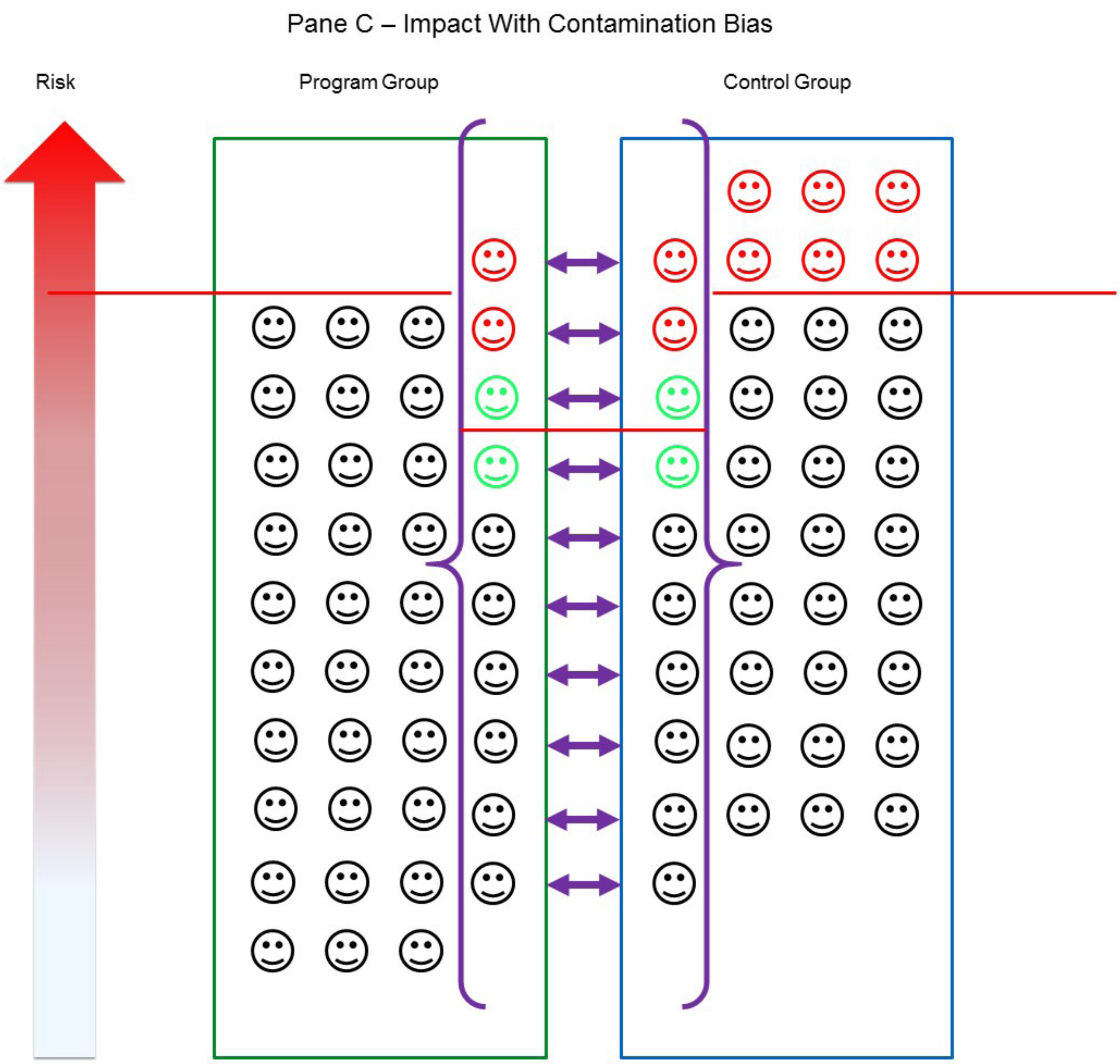
The potential for dating to lead to contamination bias is illustrated by the green faces in pane B of Figure 1. The people with the green faces in the program group are dating the people with green faces in the control group. The green people in the program group have been affected by the program – their risk has been lowered, which is visually represented by their lower position relative to Pane A. The result is that they are now below the threshold for engaging in risky behavior while their dating partners remain above that threshold. This scenario creates the potential for contamination bias.

Figure 1, Pane B: A model of how dating can lead to contamination bias.



In pane C of Figure 1 we see the potential for contamination bias realized. In this pane, we focus just on the green faces. For this example, couples make their decision to engage in risky behavior based on the *average* of their independent inclinations to engage in risky behavior—the green faces in the program group shift up relative to pane B while the green faces in the control group shift down. Without contamination there would have been just two individual in the program group engaging in risky behavior (pane B); with contamination there are three (pane C). Without contamination there would have been 10 individuals in the control group participating in risky behavior (pane B); with contamination there are 9 (pane C). Without contamination the impact on the number of people engaging in risky behavior would have been a reduction of 8; with contamination the impact is a reduction of 6.

Figure 1, Pane C: A model of how dating can lead to contamination bias.



The Factors that Lead to Contamination Bias from Dating

To assess the likely magnitude of contamination bias from dating, we use a mathematical model of how dating could affect bias combined with information from past research regarding key factors in the model (like the proportion of students who date and prevalence rates for measures of risky behavior). The mathematical model is a more formal representation of the logic model illustrated in Figure 1. The full mathematical model is presented in the appendix.[6]

In our model there are five factors that affect the magnitude of contamination bias. We break these factors into two categories: (1) those that can potentially be observed by researchers when designing a study and (2) those that most likely cannot be observed.

Factors That Affect Contamination Bias and Can Potentially Be Observed by Researchers

Researchers can use estimates of these three factors to define their study context. The estimates could come from past descriptive studies of similar populations or from descriptive analyses conducted by researchers on their samples. The third factor listed here is the most difficult to estimate, but it is possible to estimate if researchers have access to data on the prevalence of risky behavior among youth who do and do not date.

1. ***The proportion of the study sample dating another member of the study sample.*** We examine dating rates ranging from 0.05 to 0.95.
2. ***The overall prevalence rate of the outcome.*** We examine outcomes with prevalence rates ranging from 0.05 to 0.95.
3. ***Difference in the risk threshold between people dating a sample member and people not dating another sample member.*** In our model people in a dating relationship face lower barriers to engaging in risky behavior—that is, they have a lower risk threshold. This means that the prevalence rate of risky behavior is higher among people dating another member of the study sample. The prevalence rate is not zero, however, among people who are not dating another member the study sample. This is for two reasons. First, they could be dating someone who is **not** a member of the study sample (for example, someone in a different grade or a different school). Second, people who are not dating might still have opportunities to engage in risky sexual behavior, even if those opportunities are fewer and further between. We examine three magnitudes of risk threshold differences between those who are and are not dating another sample member: 0.5, 1.0, and 2.0 standard deviations of the underlying inclination to engage in risky behavior (we model the inclination to engage in risky behavior as a continuous variable following the standard normal distribution with mean 0 and variance 1). These differences in risk thresholds translate into differences in the percentage engaging in risky behavior. Examples of outcome prevalence rates corresponding to differences in risk threshold are presented in Table 1 (for these calculations we assume that 20 percent of the sample dates another member of the sample).

Unobservable Factors That Affect Contamination Bias From Dating

These two factors most likely cannot be estimated at the design stage of a study.

1. ***How dating couples are formed.*** Many factors may influence how individuals choose their dating partners. The factor that is relevant to assessing contamination bias is the inclination to engage in risky behavior. We assume that dating partners are very similar in their inclination to engage in risky behavior. This mechanism is consistent with the homophily principle (McPherson, Smith-Lovin, and Cook 2001).
2. ***How members of a couple influence each other.*** Another issue for our assessment of the magnitude of contamination bias is how a dating couple resolves their different preferences for engaging in risky behavior. We assume that the two members of the dating couple affect each other's inclination to engage in risky behavior so that they end up with an average of their independent inclinations. This is the approach illustrated in Figure 1.

Table 1: Examples of outcome prevalence rates by difference in risky behavior threshold

Risk Threshold Difference (standard deviations)	Prevalence Rate		
	Overall	Dating Sample Member	Not Dating Sample Member
0.50	0.10	0.18	0.08
0.50	0.20	0.32	0.17
1.0	0.10	0.28	0.06
1.0	0.20	0.46	0.14
2.0	0.10	0.44	0.02
2.0	0.20	0.71	0.07

While these factors cannot be directly observed at the design stage of an evaluation, we are still able to take them into account in our models and assess the sensitivity of our findings to our assumptions. In the appendix, we consider an alternative mechanism for how couples are formed in which dating couples are formed randomly with respect to their inclinations to engage in risky behavior. We also consider an alternative mechanism for how couples influence each other in which the decision to engage in risky behavior is determined by the person in the couple who is least inclined to engage in risky behavior.

Assessing the Tradeoff Between Contamination Bias and Clustering Effects

Having developed a model of how contamination bias varies across study contexts, we can now turn to assessing the tradeoff between contamination bias and clustering effects when choosing between randomization of individuals or schools. We assess the power tradeoffs between randomization of individuals and randomization of schools by comparing the minimum detectable effect size (MDES) for a design that randomizes schools to a design that randomizes students within schools, holding all other factors constant between the two designs. We assume that there is no contamination bias from dating when schools are randomized. We also assume that there are no other forms of contamination bias, for example we assume that casual interactions between members of the program and control groups do not lead to contamination. There are a total of 7 factors that affect the bias-variance tradeoff. The first 5 factors are the 5 factors that affect contamination bias (described in the previous section). In addition, two factors affect variance: (1) sample size (particularly the number of schools) and (2) the ICC.

We use simulations to calculate the MDES for both randomization of schools and randomization of students. The MDES is the smallest detectable effect on the inclination to engage in risky behavior—this is the (unobserved) direct effect of the intervention on individuals, prior to individuals being influenced by dating partners (and therefore prior to contamination). An effect of the program on individuals' inclination to engage in risky behavior is *detected* if their observed behavior is changed by a magnitude that is large enough to be statistically significant. The methods used to conduct the simulations used to compare the MDES from school randomization to individual randomization are described in the appendix.

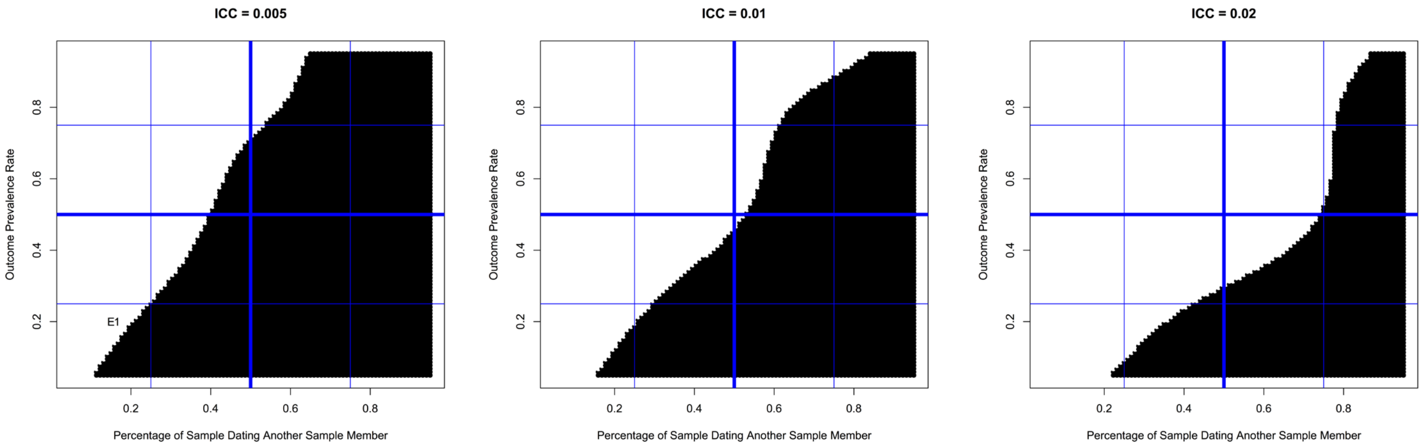
Simulation Findings

Using simulations, we examine the tradeoffs between individual randomization and school randomization across a wide range of possible study contexts. We use Figures 2-5 to show the tradeoffs across different contexts. Each figure consists of three plots. The five factors that define different contexts, the values they take, and how they are represented in the figures are:

1. ***The prevalence of dating.*** This is the proportion of the study sample who are dating another member of the study sample. We examine values ranging from 0.05 to 0.95. This factor is the horizontal axis of each plot in a figure.
2. ***The prevalence of risky sexual behaviors.*** We examine values ranging from 0.05 to 0.95. This factor is the vertical axis of each plot in a figure.
3. ***The difference in the prevalence of risky sexual behaviors between those who are and are not dating other study sample members.*** We examine this difference using the difference in risk thresholds between those who are and are not dating another member of the study sample. The differences we examine are 0.5, 1.0, and 2.0 standard deviations. This factor is varied across plots in Figures 4 and 5. In many cases researchers may have no clear basis for selecting a value for this factor, in which case we suggest assuming a difference of 1 standard deviation.[7] However, there may be some cases when there can be a basis for choosing another value—we examine such a scenario in Example 2 in Table 2, below.
4. ***Sample size.*** We examine two sample sizes, corresponding to a small-scale study and a large-scale study—8 schools (Figures 2 and 4) and 30 schools (Figures 3 and 5). We presume that researchers conducting studies with fewer schools are doing so in a context where the ICCs are smaller, so we focus on lower ICCs with 8 schools than with 30 schools. Figures reporting findings for sample sizes of 16 and 60 schools are included in the appendix.
5. ***The magnitude of clustering effects.*** This is measured by the ICC. The ICCs we examine are 0.005, 0.01, 0.02, and 0.05. This factor is varied across plots in Figure 2 (where we compare 0.005, 0.01, and 0.02) and Figure 3 (where we compare 0.01, 0.02, and 0.05). We chose these ICC values based on the distribution of ICCs reported in Glassman et al (2015).

Each plot in a figure shows whether school-level randomization or individual-level randomization will yield a lower MDES, given these five contextual factors. White indicates a lower MDES for randomization of individuals and black indicates a lower MDES for randomization of schools. The horizontal axis in each plot is the percentage of the sample dating another member of the sample. The vertical axis is the prevalence rate of the outcome. For example in plot B of Figure 2, school randomization is preferred if 50 percent of the study sample is dating and the outcome prevalence rate is 20 percent, but individual randomization is preferred if the outcome prevalence rate is 50 percent.

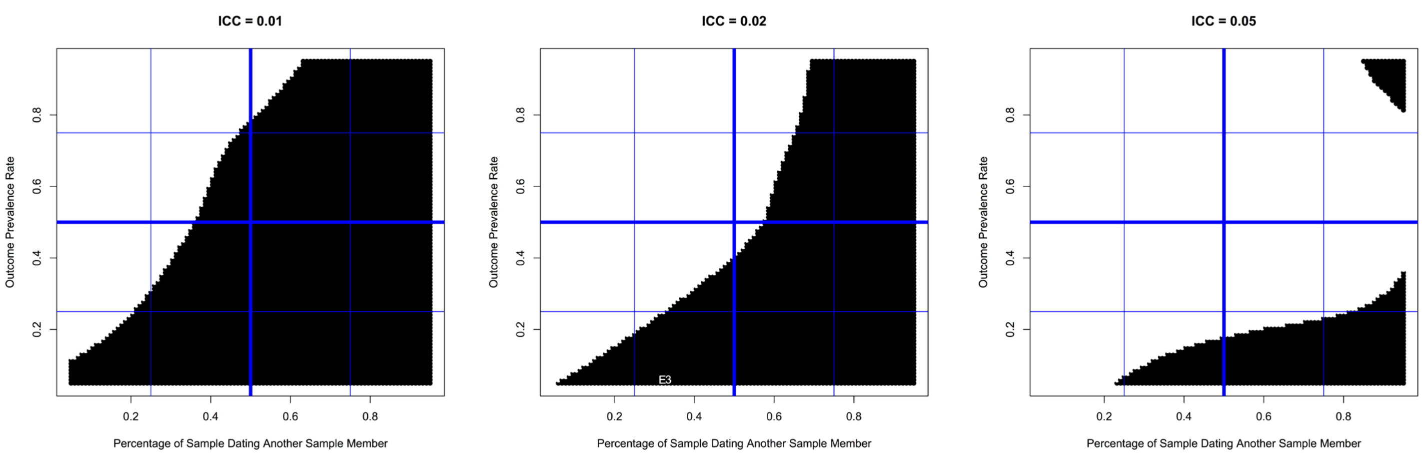
Figure 2: MDES differences between student and school randomization with 8 schools, variying the ICC
(White indicates a lower MDES for randomization of individuals; black indicates a lower MDES for randomization of schools)



Plot A **Plot B** **Plot C**

Parameter values held constant:
8 schools with 100 students per school
There is a 1 standard deviation
difference between daters and non-daters in the threshold for engaging in risky behavior
E1 indicates the point referenced in Example 1

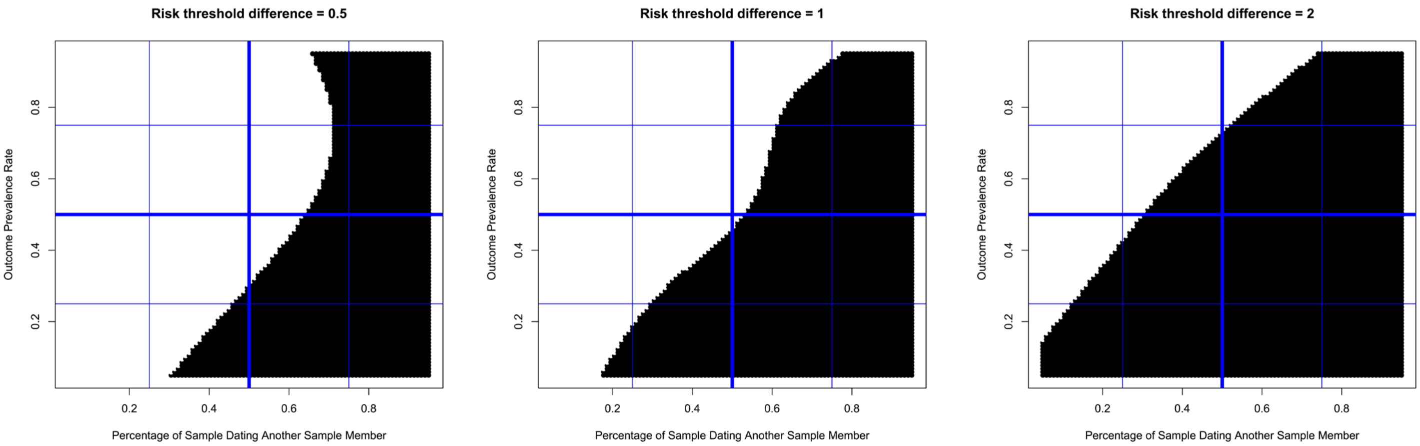
Figure 3: MDES differences between student and school randomization with 30 schools, variying the ICC
(White indicates a lower MDES for randomization of individuals; black indicates a lower MDES for randomization of schools)



Plot A **Plot B** **Plot C**

Parameter values held constant:
30 schools with 100 students per school
There is a 1 standard deviation
difference between daters and non-daters in the threshold for engaging in risky behavior
E3 indicates the point referenced in Example 3

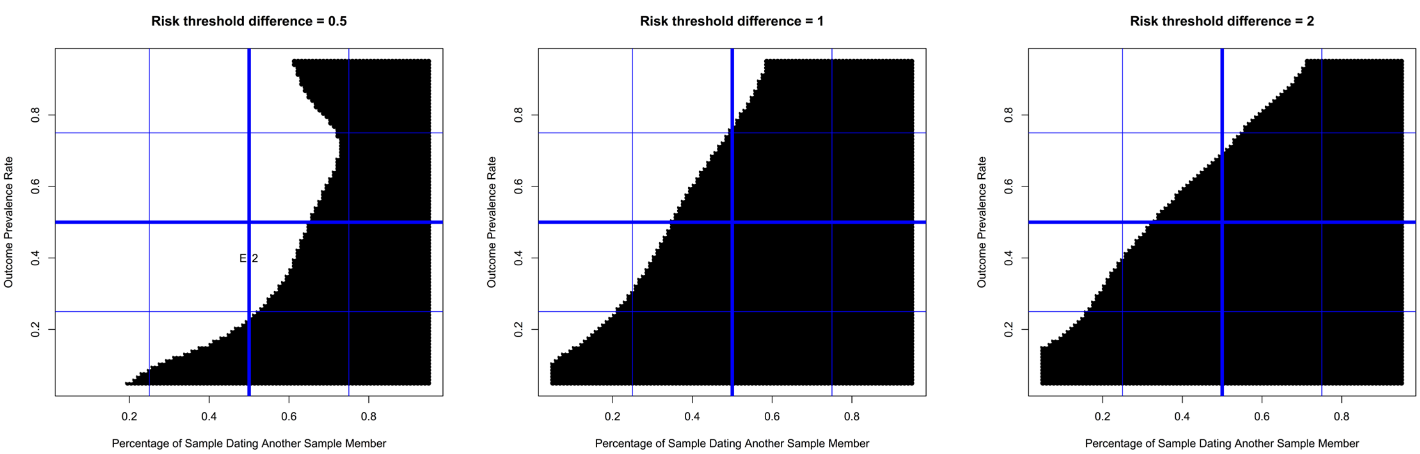
Figure 4: MDES differences between student and school randomization with 8 schools, variying the relationship between dating and risky sex (White indicates a lower MDES for randomization of individuals; black indicates a lower MDES for randomization of schools)



Plot A **Plot B** **Plot C**

Parameter values held constant:
8 schools with 100 students per school
ICC of 0.01

Figure 5: MDES differences between student and school randomization with 30 schools, variying the relationship between dating and risky sex (White indicates a lower MDES for randomization of individuals; black indicates a lower MDES for randomization of schools)



Plot A **Plot B** **Plot C**

Parameter values held constant:
30 schools with 100 students per school
ICC of 0.02
E2 indicates the point referenced in Example 2

The simulation findings show that the best design (randomization of schools or individuals) depends simultaneously on multiple factors. Two broad findings are that:

- higher values of the ICC favor randomization of individuals;
- randomization of schools is more likely to be preferred when the impact on dating couples represents a larger fraction of the overall impact (multiple factors affect this).

Examining each factor individually while holding all other factors constant we see the following patterns:

- school randomization is more likely to be preferred when the prevalence of dating is high (the more students who date, the more students who are affected by contamination bias with individual randomization);
- individual randomization is more likely to be preferred when the prevalence of a risky outcome behavior is high (in most contexts, an increase in the overall prevalence rate moves the prevalence rate for sample non-daters closer to 50 percent, which means that the percentage point impact for non-daters becomes larger,[8] which means that the impact for non-daters represents a relatively larger share of the overall impact);
- school randomization is more likely to be preferred when the risk threshold difference is high (when this difference is high, the prevalence rate for non-daters is typically very low, which means that the impact on non-daters is a smaller share of the overall impact).

Using Simulation Findings to Design an Evaluation

Evaluators can use their knowledge of their evaluation context and the results of our simulations to decide whether to randomize schools or students. In addition to data they may have for their specific context, evaluators can also draw on three large nationally representative surveys—Monitoring the Future, the National Longitudinal Study of Adolescent to Adult Health, and the Youth Risk Behavior Survey—for data related to dating and risky sexual behavior.[9] We include tables in the appendix that report statistics on risky behavior and dating from these surveys. In addition, Glassman et al (2015) reports covariate-adjusted ICCs for several risky behaviors in different contexts. Our selection of ICCs to include in our simulations was informed by Glassman et al (2015). Specifically, the median covariate-adjusted ICC reported by Glassman et al (2015) was approximately 0.01, the mean was 0.02, and the 90th percentile of ICCs was 0.05 (we use ICCs of 0.005, 0.01, 0.02, and 0.05).

We provide three examples of how to combine information about a study context with our simulation results to choose the level of randomization. The examples are summarized in Table 2. We draw on the literature referenced above for the values of the parameters specified in the table.

Table 2: Examples of applying simulation findings to evaluation contexts

	Example 1	Example 2	Example 3
Outcome	Sexual Initiation	No Condom	Risky Sex
Population	9th graders	Sexually Active 9-12 graders	9-12 graders
Proportion dating another sample member	0.12	0.50	0.32
Outcome prevalence rate	0.20	0.40	0.05
Difference in the risk threshold between those who are and are not dating another sample member (standard deviations)	1.0	0.5	1.0
ICC	0.005	0.02	0.02
Number of schools	8	30	30
Relevant Figure	Figure 2, plot A	Figure 5, plot A	Figure 3, plot B
Randomize schools or individual students?	Individual	Individual	School

Example 1: An evaluation of a program to reduce sexual initiation among 9th graders

In this example, we imagine an evaluation of a program to reduce sexual initiation that is implemented in 9th grade health classes, and that all 9th graders in study schools are included in the evaluation. The rationale for each contextual factor specified in Table 2 is:

- ***Proportion dating another sample member.*** According to Wildsmith (2013), 47% of 8th graders and 62% of 10th graders reported ever dating, so we assume 60% 9th graders date (assuming a value near the high end of the range is conservative with respect to bias). According to Arcidiacono (2010), about 50% of students date someone in their own school and about 40% date someone in their same grade. Combined, these estimates imply that about 12% (0.6*0.5*0.4) of students in the study sample will be dating another member of the study sample.

- ***Outcome prevalence rate.*** According to Kann et al (2014), about 20% of 9th graders are sexually active.
- ***Risk threshold difference.*** Giordano (2010) finds that 36% of students who are currently dating have sexual intercourse. With the outcome prevalence and sample dating rates assumed above, a 0.5 standard deviation difference in the risk threshold equates to outcome prevalence rates among sample daters and sample non-daters of 33% and 18%. A 1 standard deviation difference equates to 48% and 15%. A 2 standard deviation difference equates to 76% and 10%. While the 0.5 standard deviation difference in thresholds yields an outcome prevalence rate for sample daters close to Giordano (2010), we note that Giordano (2010) is not based on a nationally representative survey—it is based on the Toledo Adolescent Relationships Study conducted in Toledo, OH. To be conservative with respect to bias, we therefore assume a 1 standard deviation difference in the risk threshold.
- ***Number of Schools.*** The evaluation is limited to 8 schools for reasons beyond the control of the evaluators.
- ***ICC.*** Glassman et al (2015) reports ICCs for sexual initiation of about 0.005, so we use that average to inform our example.

The selected values for these contextual factors lead us to the point labeled E1 in Plot A of Figure 2. In that plot, we see that a dating rate of 0.15 and an outcome prevalence rate of 0.20 is in the range where randomization of individuals yields the smallest MDES.

Example 2: An evaluation of a program to increase condom use among sexually active high school students

In this example, we imagine an evaluation of a program to provide free condoms to sexually active high school students, with the goal of increasing condom use among the sexually active. The study sample includes all sexually active students in the school. The rationale for each contextual factor specified in Table 2 are:

- ***Proportion dating another sample member.*** Because the sample is defined as all students who were sexually active at baseline, we conservatively assume that all of these students are dating. According to Arcidiacono (2010), about 50% of students date someone in their own school, so we assume 50% of the students in the study sample are dating another member of the study sample.
- ***Outcome prevalence rate***. According to Kann et al (2014), about 40% of sexually active high school students do not use a condom.
- ***Risk threshold difference.*** The risky behavior in this study is failure to use a condom. Since we assume that everyone in the study sample was dating someone at baseline, there should be little difference in the risk threshold between someone dating a member of the study sample and someone not dating a member of the study sample. Since this is a subjective determination that cannot be directly informed by evidence, however, we do not assume a zero difference in the risk threshold. Instead, we assume a 0.5 standard deviation in the risk threshold. With the outcome prevalence and sample dating rates assumed above, a 0.5 standard deviation difference in the risk threshold equates to outcome prevalence rates among sample daters and sample non-daters of 50% and 30%.
- ***Number of Schools.*** The evaluation includes 30 schools.
- ***ICC.*** Glassman et al (2015) reports ICCs for condom use ranging from 0 to 0.15, with a mean of 0.03 and a median of 0.01. We assume an ICC of 0.02.

The selected values for these contextual factors lead us to the point labeled E2 in Plot A of Figure 5. In that plot, we see that a dating rate of 0.50 and an outcome prevalence rate of 0.40 is in the range where randomization of individuals yields the smallest MDES.

Example 3: An evaluation of a comprehensive program to reduce risky sex among high school students

In this example, we imagine an evaluation of a comprehensive program to reduce risky sex among high school students, either by reducing sexual initiation or by increasing use of condoms or other birth control methods. The study sample includes all students in the school. The rationale for each contextual factor specified in Table 2 are:

- ***Proportion dating another sample member.*** According to Wildsmith (2013), 62% of 10th graders and 66% of 12th graders reported ever dating, so we assume 64% of all high school students date (assuming a value near the high end of the range is conservative with respect to bias). According to Arcidiacono (2010), about 50% of students date someone in their own school. Combined, these estimates imply that about 32% (0.64*0.5) of students in the study sample will be dating another member of the study sample.
- ***Outcome prevalence rate.*** According to Kann et al (2014), about 5% of high school students are sexually active but not using any form of contraception.
- ***Risk threshold difference.*** We are not aware of data sources that directly inform what value we should assume for this variable. However this outcome is closely related to the out-

comes in examples 1 and 2. To be conservative with respect to contamination bias, we assume the difference in the risk threshold is 1.0 standard deviations. With the outcome prevalence and sample dating rates assumed above, a 1.0 standard deviation difference in the risk threshold equates to outcome prevalence rates among sample daters and sample non-daters of 12% and 2%.

- ***Number of Schools.*** The evaluation includes 30 schools.
- ***ICC.*** Glassman et al (2015) reports ICCs for a dichotomous measure of risky sex ranging from 0.004 to 0.08, with a mean and median of 0.03. Since we do not have simulation results for an ICC of 0.03 we assume an ICC of 0.02 since that is conservative with respect to bias.

The selected values for these contextual factors lead us to the point labeled E3 in Plot B of Figure 3. In that plot, we see that a dating rate of 0.32 and an outcome prevalence rate of 0.05 is in the range where randomization of schools yields the smallest MDES.

Conclusion

Selecting the level of randomization that will maximize the potential for an evaluation to detect program impacts is challenging. Clustering effects from randomization of schools can substantially reduce the statistical precision of impact estimates, making it difficult to detect effects. Yet randomization of individuals, while yielding more precise impact estimates, can lead to contamination bias which can also make it difficult to detect effects because contamination bias leads to smaller impact estimates.

If a researcher's goal is to design a study with the greatest potential to detect a program effect (if one exists) then there are no simple rules of thumb that can be viewed as "safe" or "conservative." It might be tempting to always randomly assign schools because then one is guaranteed that there will be no contamination bias. Such a strategy may seem appealing because bias is often an "unknown unknown"—a source of error that may or may not exist and whose magnitude is unknown. But if bias can be reasonably bounded then following such a rule of thumb can be misguided because one is also very likely to have a larger standard error with a school-randomization design, which reduces the probability that smaller impacts will be detected. In other words, if there are contexts where we can be reasonably assured that the individual-level design is more powerful, then we would be able to detect program impacts that would otherwise go undetected. An important caveat is that this discussion assumes researchers face financial or logistical constraints regarding the size of the study they can conduct. If researchers are free to conduct a study of whatever size is needed to assure a high probability of detecting the smallest substantively important impact, then there is a clear rule of thumb to follow—always randomize schools.

We contend that for most studies a better approach is to develop moderately conservative estimates of bias and then assess tradeoffs between designs that randomize schools or individuals. The simulation findings presented in this brief show that selecting the best design depends on multiple dimensions of the study context. This means that careful consideration of multiple factors is needed—rules of thumb based on one or two factors is not enough. We provide three examples of how to use information about study context, along with our simulation results, to select the level of randomization.

Endnotes

[1] See Bryk and Raudenbusch (1992), Moulton (1992), and Schochet (2008).

[2] Rhoads (2011) examined the statistical power tradeoff between randomization of schools and students in education studies where contamination due to peer effects could attenuate impact estimates. He concluded that contamination due to peer effects is likely to be too small to warrant randomization of schools rather than randomization of students within schools when the objective is to maximize the statistical power of the study.

[3] The federal evaluation of abstinence only education was criticized by some researchers for randomly assigning individual students within schools because of concerns about dating-induced attenuation bias (Weed and Lickona 2014; https://www2.cortland.edu/dotAsset/260f7ddd-526b-41e0-9ac6-7a7f0c5eaa19.pdf)

[4] When the potential outcomes of one study participant are affected by the treatment status of another participant the assumption of "non-interference" is violated (Cox 1958; Rubin 1978). The implications of the violation of this assumption depend on how it is violated. In this brief, the main implication is attenuation bias resulting from a dating couple including a member of the treatment group and a member of the control group. A secondary implication is that even when both couples are in the same treatment condition, they still influence each other and that influence alters the effects of the program. These effects are examined in the appendix.

[5] Dating can occur within the program and control groups too, but since those relationships do not lead to contamination bias, we do not highlight those relationships in this figure.

[6] The simplified model depicted in the figure is deterministic, but the actual mathematical model includes random components (see the appendix).

[7] This suggestion is based on our subjective assessment that the difference in outcome prevalence rates between students who are and are not dating another sample member should be large, but not so large that the prevalence rate is negligible among those not dating another sample member. We think the difference should be large because students in a dating relationship have more opportunities to engage in risky behavior. Yet we think that the prevalence rate should not be negligible among those not dating another sample member for two reasons. First, they could be dating someone who is not in the study sample. Second, they could participate in risky behaviors outside of a dating relationship. As another point of reference, Cohen (1988) subjectively described a difference between two groups of 0.80 standard deviations or greater as "large."

[8] This is because the same impact on the latent continuous inclination to engage in risky behavior has different percentage point impacts on

the dichotomous outcome variable depending on the prevalence rate. For example, when the prevalence rate is 5 percent, a 0.20 standard deviation impact on the latent inclination translates into a 1.7 percentage point impact. When the prevalence rate is 50 percent, a 0.20 standard deviation impact on the latent inclination translates into a 7.9 percentage point impact.

[9] Studies that report descriptive findings from these surveys include Carver (2003), Arcidiacono (2010), Wildsmith (2013), Kann et al (2014), and Child Trends (2015).

References

Arcidiacono, P, A. Beauchamp, and M. Mc Elroy. "Terms of Endearment: An Equilibrium Model of Sex and Matching." 2010.

Bryk, A. and S. Raudenbush. *Hierarchical Linear Models: Applications and Data Analysis Methods*, Newbury Park, CA: Sage, 1992.

Carver, K., Joyner, K., and Udry, J. R. "National estimates of adolescent romantic relationships." (2003).

Child Trends. "Indicators on Children and Youth: Dating." December 2015.

Cohen, J. *Statistical Power Analysis for the Behavioral Sciences*, 2nd edition. Mahwah, NJ: Lawrence Erlbaum Associates, 1988.

Connolly J, Craig W, Goldberg A, Pepler D. Mixed-Gender Groups, Dating, and Romantic Relationships in Early Adolescence. Journal of Research on Adolescence. 2004; 14:185–207.

Cox DR. Planning of Experiments. Wiley; New York: 1958.

Coyle K. et. al. "Romantic relationships: an important context for HIV/STI and pregnancy prevention programmers with young people." *Sex Education*, vol. 14, no. 5, 2013, pp. 582-596.

Furman, W. and L. Hand. "The Slippery Nature of Romantic Relationships: Issues in Definition and Differentiation." In Romance and Sex in Adolescence and Emerging Adulthood: Risks and Opportunities, ed. A. Crouter, A. Booth, pp. 171-78. 2006. Mahwah, NJ: Erlbaum.

Giordano, P., W. Manning, and M. Longmore. "Affairs of the Heart: Qualities of Adolescent Romantic Relationships and Sexual Behavior." *J. Res. Adolesc.* 20(4): 983-1013. 2010.

Glassman, J. R., Susan C. Potter, Elizabeth R. Baumler, and Karin Coyle. "Estimates of Intraclass Correlation Coefficients From Longitudinal Group-Randomized Trials of Adolescent HIV/STI/Pregnancy Prevention Programs." *Health Education & Behavior*, vol. 42, no. 4, 2015, pp. 545-553.

Goldberger, A.S. (1964) *Econometric Theory*. New York: Wiley.

Hogg, R. V., & Craig, A. T. (1978). *Introduction to mathematical statistics* (4th ed.). New York, NY: Macmillan.

Kann, L., et al. "Youth Risk Behavior Surveillance—United States, 2013." Centers for Disease Control and Prevention: Morbidity and Mortality Weekly Report, Surveillance Summaries, vol. 23, no. 4, 2014.

Maddala, G.S. (1983) *Limited Dependent and Qualitative Variables in Econometrics*. New York: Cambridge University Press.

McPherson, J. Miller, Lynn Smith-Lovin, and James M. Cook. 2001. "Birds of a Feather: Homophily in Social Networks." *Annual Review of Sociology* 27:415–44.

Moulton, Brent R. "An Illustration of a Pitfall in Estimating the Effects of Aggregate Variables on Micro Units." *Review of Economics and Statistics*, Vol. 72, No. 2, 1990, pp. 334-38.

Rhoads, C. (2011). "The Implications of 'Contamination' in Experimental Designs in Education." *Journal of Educational and Behavioral Statistics*, vol. 36, no. 1, 76-104.

Rubin, D.B. (1978) "Bayesian Inference for Causal Effects: The Role of Randomization." *The Annals of Statistics*, 6, 34–58.

Rubin, D. B. (1974). "Estimating causal effects of treatments in randomized and nonrandomized studies." *Journal of Educational Psychology*, 66, 688–701.

Schochet, Peter. "Statistical Power for Random Assignment Evaluations of Education Programs." *Journal of Educational and Behavioral Statistics*, vol. 33, no. 1, 2008, pp. 62–87.

Wildsmith, E., M. Barry, J. Manlove, and B. Vaughn. "Dating and Relationships." Adolescent health Highlight, Child Trends. October 2013.

This brief was written by John Deke from Mathematica Policy Research for the HHS Office of Adolescent Health under contract #HHSP233201300416G.