Analytic Technical Assistance and Development
U.S. Department of Education
July 2017

Comparing Impact Findings from Design-Based and Model-Based Methods: An Empirical Investigation

Tim Kautz

Peter Z. Schochet

Charles Tilley

Mathematica Policy Research

NATIONAL CENTER FOR EDUCATION EVALUATION AND REGIONAL ASSISTANCE
Institute of Education Sciences
U.S Department of Education

NCEE 2017–4026

The National Center for Education Evaluation and Regional Assistance (NCEE) conducts unbiased, large-scale evaluations of education programs and practices supported by federal funds; provides research-based technical assistance to educators and policymakers; and supports the synthesis and the widespread dissemination of the results of research and evaluation throughout the United States.

July 2017

This report was prepared for the Institute of Education Sciences (IES) by Decision Information Resources, Inc. under Contract ED-IES-12-C-0057, Analytic Technical Assistance and Development. The content of the publication does not necessarily reflect the views or policies of IES or the U.S. Department of Education nor does mention of trade names, commercial products, or organizations imply endorsement by the U.S. Government.

Kautz, T., Schochet, P., & Tilley, C. (2017). *Comparing Impact Findings from Design-Based and Model-Based Methods: An Empirical Investigation* (NCEE 2017–4026). Washington, DC: U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance, Analytic Technical Assistance and Development.

Acknowledgments

The authors are very grateful for helpful comments provided by Jamie Diamandopoulos, Mark Dynarski, Ruth Neild, and Alex Resch, as well as excellent programming and support from Matthew Jacobus, Emma Kopa, and Bethany Simard. All errors are our own.

Contents

Figures

Tables

Appendix Tables

Executive summary

A new design-based theory has recently been developed to estimate impacts for randomized controlled trials (RCTs) and basic quasi-experimental designs (QEDs) for a wide range of designs used in social policy research (Imbens & Rubin, 2015; Schochet, 2016). These methods use the potential outcomes framework and known features of study designs to connect statistical methods to the building blocks of causal inference. They differ from model-based methods that have commonly been used in education research, including hierarchical linear model (HLM) methods and robust cluster standard error (RCSE) methods for clustered designs. In comparison to model-based methods, the design-based methods tend to make fewer assumptions about the nature of the data and also more explicitly account for known information about the experimental and sampling designs. While these theoretical differences suggest the corresponding estimates might differ, it is unclear how much of a practical difference it makes to use design-based methods versus more conventional model-based methods.

This study addresses this question by re-analyzing nine past RCTs in the education area using both design- and model-based methods. The study uses real data, rather than simulated data, to better explore the differences that would arise in practice. In order to investigate the full scope of differences between the methods, the study uses data generated from different types of randomization designs commonly used in social policy research: (1) non-clustered designs in which individuals are randomized, (2) clustered designs in which groups are randomized, (3) non-blocked designs in which randomization is conducted for a single population, and (4) blocked (stratified) designs in which randomization is conducted separately within partitions of the sample. The study conducts the design-based analyses using *RCT-YES*, a free software package funded by the Institute of Education Sciences (IES) that applies design-based methods to a wide range of RCT designs (www.rct-yes.com).

The report focuses on two analyses that compare model- and design-based methods, both of which suggest there is little substantive difference in the results between the two methods. For both analyses, the study uses a reference model-based method that is similar to the one used in the original evaluation. In the first analysis, the study compares the reference model-based method to a design-based method with underlying assumptions that most closely align with those of the reference model-based method. In the second analysis, the report presents a sensitivity check that compares the reference model-based method to an alternative design-based method. In particular, the alternative method is based on the default settings in the *RCT-YES* software, which correspond to an alternative set of plausible assumptions. The findings from both analyses suggest that model- and design-based methods yield very similar results in terms of the magnitude of impact estimates, statistical significance of the impact estimates, and implications for policy.

To contextualize the differences in impact estimates between design- and model-based methods, the report also presents a third analysis, which compares estimates from two commonly used model-based methods: (1) HLM methods and (2) linear models with ordinary least squares (OLS) assumptions and RCSE to account for clustering. Importantly, this analysis suggests that the differences between the design- and model-based methods (with similar assumptions) are no greater than the differences that would arise between commonly used, model-based methods.

Based on the studies we consider, the impact findings using the various design- and model-based methods are similar. The magnitude of the differences, however, sometimes depends on the specific underlying assumptions, for example, how clusters (such as schools) and blocks (such as school districts) are weighted for the analysis. An advantage of design-based methods is that these assumptions are often explicit and need to be specified by the analyst, whereas for model-based methods these assumptions are often implicit. Our study suggests that researchers should select estimators with assumptions that best suit the goals of their study regardless of whether they use a design- or model-based approach. Moreover, researchers should consider the tradeoffs between different assumptions, and how these assumptions affect the interpretation of findings.

I. Introduction

A new design-based theory has recently been developed to estimate impacts for randomized controlled trials (RCTs) and basic quasi-experimental designs (QEDs) for a wide range of designs used in social policy research (Imbens & Rubin, 2015; Schochet, 2016). These methods use the potential outcomes framework to connect statistical methods to the building blocks of causal inference.

The design-based methods differ from model-based methods that have commonly been used in education research, including hierarchical linear model (HLM) methods and robust cluster standard error (RCSE) methods for clustered designs. In comparison to model-based methods, the design-based methods tend to make fewer assumptions about the nature of the data and also more explicitly account for known information about the experimental and sampling designs. While these differences might affect both the impact estimates and their precision, it is unclear how much of a practical difference it makes to use design-based methods versus more conventional model-based methods.

This study addresses this question by re-estimating the impacts on key outcomes from nine past RCTs in the education area using both design- and model-based HLM and RCSE methods. In order to explore the full scope of differences between the methods, we selected RCTs for the study that cover a full range of randomization designs commonly used in social policy research: (1) non-clustered designs in which individuals are randomized, (2) clustered designs in which groups are randomized, (3) non-blocked designs in which randomization is conducted for a single population, and (4) blocked (stratified) designs in which randomization is conducted separately within partitions of the sample. The study conducts the design-based analyses using *RCT-YES*, a free, IES-funded software package that applies design-based methods to a range of experimental and non-experimental designs (www.rct-yes.com). The report is geared toward researchers with some training in statistical methods who are interested in new approaches to impact estimation and are considering using design-based methods for impact evaluations that they are conducting or supporting.

The report presents three sets of analyses that correspond to different comparisons between methods. In particular, we compare a reference model-based method to three alternative methods: a design based approach most aligned with the reference model, a design based approach with its implicit default assumptions found in the *RCT-YES* software package, and an alternate model-based approach (i.e., HLM or linear modeling with cluster-robust standard errors). Figure 1, below, describes these three analyses, their purpose, and main findings.

First, we compare the reference model-based method to a design-based approach with underlying assumptions that most closely align with those of the reference model-based method. In this first analysis, the magnitudes of the estimated impacts are indistinguishable between the design-based and the model-based methods. Approximately 3 percent of the impact estimates on outcomes differ in terms of statistical significance between the two methods, and, even in these few cases, the differences in *p*-values are small and would not substantively change the policy conclusions.

Second, as a sensitivity analysis, we compare the reference model-based method to an alternative design-based method, specifically, a method based on the default settings in *RCT-YES*. The default settings in *RCT-YES* correspond to a potentially alternative set of plausible assumptions (although in some cases, they align with the matched assumptions). These alternative sets of assumptions were developed for *RCT-YES* based on consultations with a panel of methodological experts. In this second analysis, the results from the model- and design-based methods are similar, but the differences are slightly bigger than in the first analysis. These differences are limited to clustered designs and are consistent with differences in weighting schemes and standard error calculations between the default and alternative settings analyses.

Third, to contextualize the differences in impact estimates between design- and model-based methods, we use data from nine RCTs to compare the two commonly used, model-based methods: (1) HLM methods and (2) linear models with ordinary least squares (OLS) assumptions and RCSE to account for clustering. This third analysis suggests that the differences between design-based methods and the original, model-based methods (with similar assumptions) are no greater than the differences between these two commonly used, model-based methods.

The competing impact estimation methods used by our study are all asymptotically unbiased, but they rely on different assumptions, and they could have different finite sample properties that could vary by context. It is reassuring that, on the basis of the RCT datasets considered in this report, the different estimation methods typically yield similar findings. Nonetheless, prospective users of design- or model-based methods should select an approach with assumptions that best suit the goals of their study, which could reflect theoretical considerations. Two such considerations are: (1) how clusters and blocks are weighted to align with the impact parameter of interest and (2) whether the study results are to be formally generalized outside the study sample.

The report proceeds as follows. Section II outlines the theory underlying design-based and model-based HLM and RCSE methods, and why impacts might differ across these approaches. Section III describes our overall estimation approach. Section IV summarizes the main findings. Section V provides some conclusions and discussion. Appendix I provides details on the methods used in the report. Appendix II provides details for each evaluation and supporting tables of impact estimates for each outcome.

Figure 1. Schematic of Analyses and Key Findings

Model-based vs. Design-based

1. *Matching Assumptions*

- **Comparison**
  Reference model-based method vs. design-based method with similar assumptions
- **Purpose**
  Standardizes the assumptions to compare essential features of both methods
- **Main findings**
  - Magnitude of impacts are indistinguishable
  - Only a few (small) differences in statistical significance between the two methods
  - The differences would not substantively change the policy conclusions

2. *Alternative Assumptions*

- **Comparison**
  Reference model-based method vs. design-based method with potentially alternative assumptions (specifically, the default settings in *RCT-YES*)
- **Purpose**
  Provides a sensitivity analysis to the first analysis under a plausible set of alternative assumptions
- **Main findings**
  - Differences in impacts and statistical significance are slightly bigger than in the first analysis
  - Additional differences are limited to clustered designs and are consistent with differences in weighting and standard error calculations

Model-based RCSE vs. Model-based HLM

3. *Linear models with OLS assumptions vs. HLM*

- **Comparison**
  Reference model-based method (linear models with OLS assumptions and RCSE to account for clustering) vs. alternative model-based methods (HLM methods)
- **Purpose**
  Provides a baseline for expected estimation differences between methods
- **Main findings**
  - The differences between design-based methods and model-based methods are no greater than the differences that would arise between these two commonly used, model-based methods when the assumptions are aligned

II. Background

This section provides some background on design-based, HLM, and RCSE methods and outlines the main theoretical differences, which are important for understanding potential differences between impact findings using the three approaches. Because this study focuses on how results compare across the methods most commonly used in practice, we focus on how the models are typically estimated in education evaluations. For example, we discuss how each method implicitly weights observations and clusters if the specification does not explicitly include weights. Each of these methods can be modified from their standard use.[1] The section also briefly discusses the difference between finite- and super-population models, which is another dimension that can affect the impact findings using design-based methods. Schochet (2016) provides a more comprehensive discussion.

II.A. What are design-based methods?

Design-based estimators are derived from the known features of experiments and are based on the Neyman-Holland-Rubin causal inference model (Holland, 1986; Rubin, 1974). Consider the simplest RCT design where individuals are randomly assigned to either a treatment group that is offered an intervention or a control group that is not. The study follows participants for a period of time, and collects outcome data, such as achievement test scores, on the sample. Let y_i denote the outcome variable for individual i.

Ideally, an evaluator would measure each individual's "potential" outcome in the treatment condition $\left(Y_{Ti}\right)$ and in the control condition (Y_{Ci}). With this information, it would be possible to calculate each individual's treatment effect, $\left(Y_{Ti} - Y_{Ci}\right)$, as well as the average treatment effect, $\beta_{ATE} = \overline{Y}_T - \overline{Y}_C$, the impact parameter that is estimated for most evaluations in the education field and our focus in this report.[2]

Unfortunately, it is not possible to observe both Y_{Ti} and Y_{Ci}, because only one or the other can be observed depending on the random assignment results. Therefore, it is also not possible to directly calculate individual treatment effects. This can be seen more formally by expressing the relationship between the observed outcome, y_i, and the potential outcomes, Y_{Ti} and Y_{Ci}, as:

(1) $$y_i = T_i Y_{Ti} + \left(1 - T_i\right) Y_{Ci},$$

[1]See Raudenbush & Bryk (2002) for a comprehensive treatment of HLM and Cameron & Miller (2015) for a review of RCSE methods.

[2]There are a range of other policy-relevant parameters that can be estimated with both model- and design-based methods.

where T_i is an indicator variable that equals 1 for those assigned to the treatment group and 0 for those assigned to the control group. Equation (1) simply states that $y_i = Y_{Ti}$ is observed for those in the treatment group and $y_i = Y_{Ci}$ is observed for those in the control group.

Design-based theory uses the relation in Equation (1) to develop estimators for the unobserved average treatment effect, β_{ATE}. The first step is to add $T_i\overline{Y}_T$ and $(1-T_i)\overline{Y}_C$ to both sides of Equation (1)—which does not change the equation—and to rearrange terms in the equation to produce the following regression equation:

(2) $y_i = \beta_0 + \beta_{ATE}T_i + u_i$,

where $\beta_0 = \overline{Y}_C$ is the intercept, $\beta_{ATE} = \overline{Y}_T - \overline{Y}_C$ is the average treatment effect parameter of interest, and u_i is the model "error" term.[3] Importantly, u_i is random only because T_i is random; the potential outcomes of an individual are assumed to be fixed for the study (not sampled from a population distribution).

The design-based model in Equation (2) has statistical properties that differ from the standard linear model typically used to estimate impacts for RCTs. For example, the error term, u_i, does not have mean 0 or constant variance and is correlated with the regressor, T_i. Yet it can be shown that estimating this model using standard OLS produces a differences-in-means impact estimator based on the observed data, $\hat{\beta}_{ATE} = \overline{y}_T - \overline{y}_C$, that has the following statistical properties (see, for example, Schochet, 2016):

- ***Unbiased***, meaning that the estimator will, on average, equal the true impact parameter across all possible random assignment results
- ***Normally distributed in large samples***, so that standard *t*-tests or *z*-tests can be used to test the null hypothesis of zero average treatment effects
- ***Simple variance estimator*** with separate variance terms for the treatment and control groups

The key feature of design-based theory, then, is that it uses the random assignment process to *build* the impact estimation model in Equation (2). In contrast, model-based approaches specify an ad hoc model structure (for example, the standard OLS model) that is assumed to be true to ensure unbiased estimators. But it is not possible to fully verify these model assumptions.

[3] We use the term "error" here because the random u_i enter the equation in a similar way to model-based errors, but they differ in the sense that they are not an unexplained part of the model.

II.B. Design-based and model-based methods for clustered RCTs

In this section, we discuss how design-based methods for modeling clustered RCTs differ from two commonly used model-based methods (OLS with RCSE [which we hereafter refer to as the RCSE method] and HLM). We focus on clustered designs because they highlight some of the general differences between design- and model-based methods. First, we introduce each of the three approaches. Second, we summarize the key differences between them.

Design-based methods for clustered RCTs

The above theory for individual-level assignment can be extended to clustered designs where groups (such as schools or classrooms) are randomly assigned to the treatment and control groups instead of individuals. For example, a common design used in education RCTs randomly assigns schools and collects outcome data for students.

For clustered designs, to fix concepts, we focus here on design-based methods that average the individual data to the group level to make clear how the estimators are consistent with the level of randomization. For example, in an RCT where schools are randomized, we focus on estimators where student-level data (such as test scores) would be averaged to the school level. In our analysis, however, we use design-based methods that estimate impacts using individual-level data rather than aggregate data, because this approach allows us to include individual-level covariates in the regression models to improve precision (as discussed in Schochet, 2013 and later in this section).

In our setting, define potential outcomes (student averages) for school j in the treatment and control conditions as $\overline{Y}_{Tj}$ and $\overline{Y}_{Cj}$. The school-level treatment effect is $(\overline{Y}_{Tj} - \overline{Y}_{Cj})$, which cannot be observed because only $\overline{Y}_{Tj}$ or $\overline{Y}_{Cj}$ is observed, but not both. The impact parameter of interest, $\beta_{ATE,Clust} = \overline{\overline{Y}}_{TW} - \overline{\overline{Y}}_{CW}$, is a weighted average of these school-level treatment effects, which can also be expressed as a weighted average of student-level treatment effects.

The design-based approach can weight clusters in a variety of ways, depending on the impact parameters (estimands) of interest. For example, in a clustered design with school-level randomization, schools could be weighted (1) equally, to obtain impacts for the average school in the sample; (2) based on student sample sizes, to obtain impacts for the average student in the sample; or (3) using "precision" weighting, where schools whose mean student outcomes are measured more precisely are given larger weight in the analysis. While design-based methods can accommodate any of these possible weighting schemes, the random assignment mechanism aligns most closely with the equal-weighting of clusters (this scheme is also the default setting in *RCT-YES*).

Parallel to Equation (1) for the non-clustered design above, the observed mean outcome, $\bar{y}_j$, relates to the potential outcomes, $\bar{Y}_{Tj}$ and $\bar{Y}_{Cj}$, as follows:

$$(3) \quad \bar{y}_j = T_j \bar{Y}_{Tj} + \left(1 - T_j\right) \bar{Y}_{Cj},$$

where T_j equals 1 for schools assigned to the treatment group and 0 for schools assigned to the control group. As before, adding $T_j \bar{Y}_{Tj}$ and $(1-T_j)\bar{Y}_{Cj}$ to both sides of this equation and rearranging terms forms a regression model similar to Equation (2) where $\bar{y}_j$ is regressed on T_j, with the model error term, u_j, defined by the randomization process:

$$(4) \quad \bar{y}_j = \beta_{0,Clus}^{des} + \beta_{ATE,Clus}^{des} T_j + u_j.$$

Estimating this model using weighted least squares yields a weighted differences-in-means impact estimator that, in large samples, is unbiased (consistent) and normally distributed with a simple variance estimator (shown later). Standard *z*-tests or *t*-tests can be used for hypothesis testing where the degrees of freedom are based on the number of clusters in the sample.

HLM methods for clustered RCTs

In contrast to design-based methods, HLM methods start with assumptions about the data generating process that are not directly motivated by the design of the experiment or directly linked to the potential outcomes framework. For example, the standard HLM model for estimating impacts for clustered RCTs assumes the following relationship between the outcome and treatment status:

$$(5) \quad y_{ij} = \beta_{0,Clus}^{hlm} + \beta_{ATE,Clus}^{hlm} T_j + \nu_j + \varepsilon_{ij},$$

where y_{ij} is the outcome of individual i in school j, T_j is an indicator for whether a school received the treatment, ν_j is a school-specific random effect (intercept), and ε_{ij} is an individual-specific random effect. The total error for individual i in cluster j is $u_{ij} = \nu_j + \varepsilon_{ij}$.[4]

In the design-based approach, the distribution of the error term is *derived* based on the structure of the experiment, whereas in the HLM model the error term is *assumed* to have a particular structure and to be drawn from a particular distribution. The standard HLM assumptions about the error terms are: $\nu_j \sim N\left(0, \sigma_\nu^2\right)$; $\varepsilon_{ij} \sim N\left(0, \sigma_\varepsilon^2\right)$; and ν_j and ε_{ij} enter the model linearly and are independent. Furthermore, HLM assumes the same covariance between individuals within a cluster

[4] Raudenbush & Bryk (2002) discuss extensions to this formulation.

($cov\left(u_{ij}, u_{i'j}\right) = \sigma_v^2$ for $i \neq i'$), and that errors across clusters are not correlated ($cov\left(u_{ij}, u_{i'j'}\right) = 0$ for $i \neq i'$ and $j \neq j'$). Importantly, this model imposes a constraint that the variance is the same for the treatment and control clusters, which generally does not hold if treatment effects are heterogeneous; this is not the case for the design-based method, which allows for heterogeneity of impacts yielding different variances for the two research groups.[5] Equation (5) is typically estimated using a form of maximum likelihood.

Importantly, the weights for HLM also differ from design-based methods. Unless weights are specified otherwise, the HLM method selects a version of precision weights so that weights at the individual level are $w_{ij} = \dfrac{1}{n_j \sigma_v^2 + \sigma_\varepsilon^2}$ and the weights at the cluster level are $w_j = \dfrac{n_j}{n_j \sigma_v^2 + \sigma_\varepsilon^2}$ (Raudenbush & Bryk, 2002). In this way, larger clusters receive a greater weight, but the weight is a non-linear function of the size of each cluster. To apply these weights, the variance components must be estimated and inserted in place of the unknown values. This weighting scheme arises because HLM is a super-population model where the impact findings are assumed to generalize to the school-level impact in some broader population, whereas the design-based framework more closely aligns with a finite-population framework where the impacts are assumed to pertain to the study sample only (see Section II.C below for further discussion of this topic).

In sum, the key features of how HLM accounts for clustering are that the HLM approach estimates impacts using individual-level data and an ad hoc model specification, where the error structure is assumed to be true at each HLM level. We discuss additional differences later in this section.

OLS and RCSE methods for clustered RCTs

It is common in certain fields to obtain standard errors for RCTs from OLS models that are robust to model misspecification. These estimators include robust standard errors for non-clustered designs (Huber, 1967; White, 1980) and extensions to clustered designs using generalized estimating equations (Liang & Zeger, 1986). These estimators share features with the design-based estimators and take the form:

$$(6) \quad y_{ij} = \beta_{0,Clus}^{RCSE} + \beta_{ATE,Clus}^{RCSE} T_j + \varepsilon_{ij},$$

where the components are defined analogously to the HLM model. The total error for individual i in cluster j is $u_{ij} = \varepsilon_{ij}$. In this framework, the key assumption is that the errors are uncorrelated across clusters, but may be correlated within clusters. In general, for a covariate matrix X (which

[5] This assumption can be relaxed but has not traditionally been done in practice.

includes T_j), the RCSE approach uses the "sandwich" estimator to obtain variance estimates for the estimated OLS coefficient estimates:

$$(7)\quad c\left(X'X\right)^{-1}\left[\sum_{j=1}^{m}X_j^{'}\hat{u}_j\hat{u}_j^{'}X_j\right]\left(X'X\right)^{-1},$$

where m is the total number of clusters, $\hat{u}_j$ is the vector of OLS residuals for cluster j, and X_j is matrix of covariates for cluster j. $c=\left[\frac{m}{(m-1)}\frac{(n-1)}{(n-K)}\right]$ is a small-sample correction factor that is sometimes applied, where n is the total student sample size and K is the number of model covariates.

There are some differences between the RCSE and HLM models in their assumptions about the error terms. RCSE assumes that the covariances between individuals in the same cluster can vary so that $cov\left(u_{ij},u_{i'j}\right)=\sigma_{i,i'}^2$ for $i \neq i'$. This assumption differs from the HLM approach described above, in which the covariances are assumed to be the same between all individuals in the same cluster. Both RCSE and HLM, however, assume that the error terms for individuals across clusters are not correlated. As in the design-based methods, the error terms in the RCSE model are not assumed to follow a particular distribution (unlike the HLM model, which typically assumes normally distributed errors).

Both the RCSE and HLM approaches analyze data at the individual level, but the cluster-level weighting scheme differs for the two approaches. Unless weights are specified otherwise, the RCSE approach weights each individual equally so that the impact estimates pertain to the average individual in the study population, not to the average cluster (e.g., schools). In this way, schools with larger populations receive more weight because they contain more students. The individual-level weights are $w_{ij}=1$ and the cluster-level weights are $w_j=n_j$. As mentioned earlier, HLM uses precision weighting and the design-based approach assigns equal weight to each cluster, so the RCSE approach places the greatest weight on students.

Summary of differences between methods for clustered RCTs

The key conceptual difference between design- and model-based methods is that the design-based estimators are explicitly derived from the known structure of the experiment, whereas model-based methods are derived from a set of ad hoc assumptions. This conceptual difference leads to differences in three key aspects of the model structure that could affect the impact estimates and their precision: (1) the implicit weights, (2) assumptions about the error term, and (3) the level of data used. Table 1 summarizes these differences.

While all three methods can accommodate flexible weights, studies often use the default weights implicit in each method (see the first column of Table 1). Two conditions must hold for the differences in weighting to lead to differences in estimates: (1) the clusters contain different numbers of individuals, and (2) the treatment effects are heterogeneous. If all clusters have the same number of individuals, then the weighting schemes are the same across the three methods. The differences will be greatest when the clusters have different sample sizes and the treatment effects differ across clusters systematically with respect to cluster size. In this case, the weights correspond to conceptually different parameters. When clusters are weighted equally, the parameter represents the intervention effect for a student in the average school, whereas if students are weighted equally, the parameter represents the intervention effect on the average student. An advantage of the design-based methods is that it is transparent how weights enter the variance estimators, which is more difficult to discern for the HLM and RCSE methods.

Even when the impact parameters of interest (weighting schemes) are the same for the design- and model-based methods, the standard error calculations might differ due to differing assumptions about the error term summarized in Table 1. For example, the design-based and RCSE methods yield the same differences-in-means impact estimate if clusters are weighted by their size. However, the standard error estimators will differ as shown by Equations (8) and (9) below:

$$(8)\quad V\hat{a}r\left(\hat{\beta}^{des}_{ATE,clus}\right)=\left(\frac{m_T}{m_T-1}\right)\frac{1}{n_T^2}SS_T+\left(\frac{m_C}{m_C-1}\right)\frac{1}{n_C^2}SS_C-\frac{1}{m}\left(\frac{m_T}{n_T}\sqrt{\frac{SS_T}{m_T-1}}-\frac{m_C}{n_C}\sqrt{\frac{SS_C}{m_T-1}}\right)^2 \text{ and}$$

$$(9)\quad V\hat{a}r\left(\hat{\beta}^{RCSE}_{ATE,clus}\right)=\left(\frac{n-1}{n-2}\right)\left(\frac{m}{m-1}\right)\frac{1}{n_T^2}SS_T+\left(\frac{n-1}{n-2}\right)\left(\frac{m}{m-1}\right)\frac{1}{n_C^2}SS_C,$$

where $SS_T=\sum_{j:T_j=1}^{m_T}\sum_{i=1}^{n_j}\sum_{i'=1}^{n_j}\left(y_{ij}-\bar{y}_T\right)\left(y_{i'j}-\bar{y}_T\right)$, $SS_C=\sum_{j:T_j=0}^{m_c}\sum_{i=1}^{n_j}\sum_{i'=1}^{n_j}\left(y_{ij}-\bar{y}_C\right)\left(y_{i'j}-\bar{y}_C\right)$, $\bar{y}_T=\frac{1}{n_T}\sum_{j:T_j=1}^{m_T}\sum_{i=1}^{n_j}y_{ij}$, $\bar{y}_C=\frac{1}{n_C}\sum_{j:T_j=0}^{m_C}\sum_{i=1}^{n_j}y_{ij}$, n_T and n_C are the numbers of individuals in the treatment and control groups, $n=n_T+n_C$, m_C and m_T are the numbers of clusters in the treatment and control groups, and $m=m_T+m_C$.

There are two main differences between Equations (8) and (9). First, the design-based version includes a finite-population correction, the term subtracted at the end of Equation (8). This correction is a lower bound on the true, yet unidentified correction term that represents the extent of heterogeneity of treatment effects across the sample. This correction reduces the standard error of the design-based method compared to the RCSE method, and matters most when there is considerable heterogeneity of treatment effects across clusters. This term is only present in the finite-population model (see Section II.C for a discussion of finite-population models). Second, even without the finite-population correction, the two equations differ in how the number of clusters enters the equations, which tends to lead to smaller standard errors for the RCSE method compared to the design-based method, especially in studies with few clusters and many individuals.

So far this discussion has focused on models without covariates, but both the design- and model-based approaches allow for baseline covariates in the regression models to improve precision of the impact estimates and to adjust for treatment-control differences due to random sampling or missing data. However, there are differences. First, from a theoretical perspective, model-based methods typically assume the covariates enter the model additively, whereas design-based methods do not require this assumption because the covariates do not enter the "true" RCT data-generating process in Equations (2) and (4). However, covariates can be added to the model in the typical way using a design-based variant of the OLS multiple regression estimator.[6] It is remarkable that this approach yields impact estimators that are consistent and asymptotically normal, with variance estimators similar to those in Equation (8) except that it uses mean squared residuals from the fitted regression models in place of the SS_T and SS_C terms. With covariates, parallel adaptations to the RCSE variance estimator in Equation (9) do not apply without additional assumptions.

Second, from a practical perspective, for clustered designs, the design-based approach often accounts for covariates at the cluster level only and not at the individual level, whereas the model-based approaches typically account for covariates at both levels. This difference arises because design-based methods often aggregate the data at the cluster level for estimation. However, this aggregation does not typically have a large effect on the precision of the impact estimates, because the leading variance term for clustered designs is the variation in impacts between clusters (which cluster-level covariates can explain) rather than within clusters (which individual-level covariates can explain). However, design-based methods can also accommodate individual-level covariates. The theory underlying this approach is based on results in Schochet (2013), and involves estimating models for clustered designs using individual-level data (and weights) rather than cluster-level data. The approach uses weighted least squares to estimate the models to obtain the impact estimates and aggregates the model residuals to the cluster level to calculate the variance estimators discussed earlier. In this report, we use the design-based methods that adjust for covariates using individual-level data.

Finally, all three methods can accommodate blocked designs where random assignment is conducted *separately* within different subpopulations of the sample (for example, by site or grade level). The design-based approach uses this simple random assignment process in each block to develop impact estimators rather than specifying an ad hoc estimation model to account for the blocks. For the finite-population model, where the study blocks are treated as fixed for the study, the design-based estimators from above apply to each block separately, and can then be averaged to obtain overall

[6] See Freedman (2008), Schochet (2010), and Lin (2013) for discussions about the validity of using covariates in design-based estimators.

impact findings. For the super-population model, where the study blocks are treated as a random sample, the form of the variance estimator for the simple differences-in-means estimators differs but is still simple to apply in practice for both the clustered and non-clustered design:

$$(10)\quad V\hat{a}r\left(\hat{\beta}^{des}_{ATE,blk}\right) = \frac{1}{\left(B-1\right)B\overline{w}^{2}}\sum_{b=1}^{B}(w_{b}\hat{\beta}_{ATE,b} - \overline{w}\hat{\beta}^{des}_{ATE,blk})^{2},$$

where B is the number of blocks, $\hat{\beta}_{ATE,b}$ is the impact estimate in block b, w_b is the weight for block b, and $\overline{w}$ is the average block weight (see Schochet [2016] for details).

Blocked designs have the same types of issues with weighting blocks as clustered designs have with weighting clusters. By default, *RCT-YES* weights blocks by the number of randomized units in the block. The block-level weighting scheme for the HLM and RCSE procedures will depend on how the blocks are accounted for in the analysis (for example, by including fixed block indicators as covariates or estimating random slope models for HLM).

Table 1. Summary of Differences between Design-Based and Model-Based Approaches for Analysis of Clustered RCTs

Method	Default weights for cluster-level observations	Default weights for individual observations	Error structure	Assumptions about error structure	Level of data
Design-based	1	$1/n_j$	$u_j = p\left(\overline{Y}_{Tj} - \overline{\overline{Y}}_T\right) + (1-p)\left(\overline{Y}_{Cj} - \overline{\overline{Y}}_C\right) + \left[\left(\overline{Y}_{Tj} - \overline{\overline{Y}}_T\right) - (\overline{Y}_{Cj} - \overline{\overline{Y}}_C)\right]T_j - p\left(\overline{Y}_{Tj} - \overline{Y}_{Cj}\right),$ where p is the fraction in the treatment group	n/a	Typically the cluster, but could be the individual
Model-based (HLM)	$\frac{n_j}{n_j\sigma_\nu^2 + \sigma_\varepsilon^2}$	$\frac{1}{n_j\sigma_\nu^2 + \sigma_\varepsilon^2}$	$u_{ij} = \nu_i + \varepsilon_{ij}$	$\nu_i \sim N\left(0, \sigma_\nu^2\right), \varepsilon_{ij} \sim N\left(0, \sigma_\varepsilon^2\right),$ ν_i and ε_{ij} are independent	Individual
Model-based (RCSE)	n_j	1	$u_{ij} = \varepsilon_{ij}$	$cov\left(u_{ij}, u_{i'j}\right) = \sigma_{i,i'}^2$ for $i \neq i'$ $cov\left(u_{ij}, u_{i'j'}\right) = 0$ for $i \neq i'$ and $j \neq j'$	Individual

Note: All three methods allow for different weights but are often used with the weights listed in this table. n_j is the number of individuals in cluster j.

II.C. Finite-population and super-population models

Up to this point, we have focused on design-based estimators for finite-population models, which assume the impact findings pertain to the study sample only. Researchers can also select a super-population model, which assumes the impact findings generalize to a broader population. Inference for finite-population models assumes that the sample is fixed for the study and accounts for the variability in estimates solely based on the assignment of treatment status among the study sample. Intuitively, if the researchers repeated random assignment on the study sample, the treatment and control groups would comprise different sample members, leading to different impact estimates. Inference for super-population models, however, must also account for variability in estimates due to the assumed (or actual) sampling of study members from a broader population; drawing an entirely new sample from this population would generate different impact estimates.

The choice of whether to use a finite-population or super-population model should depend on sample design and theory. For instance, the finite-population approach allows the researcher to assess whether the intervention can work in particular contexts (rather than whether it does work) so it might be appropriate for efficacy trials with purposively selected samples. The super-population model, however, might be germane if the sample is purposefully randomly sampled from a broader population (which is rare in social policy RCTs) or for large multisite trials implemented in a variety of settings. This approach might also be justified for studies in which the study population is well defined and the conceptual model for the evaluation suggests that the study findings would be applicable to that population. The approach might also be justified under the assumption that policymakers will generalize the findings anyway, especially if the intervention will not be tested soon in other contexts (which is typically the case for social policy RCTs).

The super-population methods tend to yield less precise impact estimates than the finite-population methods because they need to account for the additional source of variability from random sampling. For example, consider the standard error estimators for the differences-in-means impact estimator for the non-clustered, non-blocked design described in Section II.A. Equations (11) and (12) present the variance estimators for the finite-population and super-population parameters:

$$(11)\ \hat{Var}\left(\hat{\beta}_{ATE,FP}\right)=\frac{S_T^2}{np}+\frac{S_C^2}{n(1-p)}-\frac{\left(S_T-S_C\right)^2}{n} \text{ and}$$

$$(12)\ \hat{Var}\left(\hat{\beta}_{ATE,SP}\right)=\frac{S_T^2}{np}+\frac{S_C^2}{n(1-p)},$$

where p is the proportion of individuals in the treatment group, $S_T^2 = \frac{1}{np-1}\sum_{i:T_i=1}^{np}\left(y_i - \overline{y}_T\right)^2$, and $S_C^2 = \frac{1}{n(1-p)-1}\sum_{i:T_i=0}^{n(1-p)}\left(y_i - \overline{y}_C\right)^2$. Note that $V\hat{a}r\left(\hat{\beta}_{ATE,FP}\right) \le V\hat{a}r\left(\hat{\beta}_{ATE,SP}\right)$. The reason for this difference is that the super-population model assumes sampling of individuals, which adds another source of variation beyond how individuals are assigned to a particular treatment group.[7]

As discussed in Imai, King, and Nall (2009), Imbens (2004), and Schochet (2016), the super-population model is more complex for clustered designs than for non-clustered designs, because assumptions must be made about multilevel sampling of schools and students from broader populations. Specifically, under the clustered super-population model, it can be assumed that (1) schools are fixed for the study, but students are randomly sampled within the study schools from a broader student population (the cluster average treatment effect [CATE]); (2) schools are randomly sampled from a broader school population, but the student sample is fixed for the study (the unit average treatment effect [UATE]); or (3) both schools and students are randomly sampled from broader populations (the population average treatment effect [PATE]). The situation becomes even more complex for blocked designs, which adds another layer of potential sampling.

For this study, it is important to understand that model-based methods typically assume a version of a super-population model. In contrast, a finite population model is the default setting used by *RCT-YES*. The only exception is that, by default, *RCT-YES* adopts the super-population framework for matched pair designs, which are common in the studies that are presented in this report.

For matched pair designs, the way in which standard errors are calculated and covariates are treated in the analysis accounts for some of the differences in this report's impact findings between the model- and design-based methods. A matched pair design is a special case of a blocked design in which each block contains a single treatment and control unit (either an individual for non-clustered designs or a group for clustered designs). For these designs, *RCT-YES* estimates the PATE parameter and the variance is estimated using Equation (10) above. Estimates of PATE are based on super-population assumptions in which the study sample represents a large population at all levels of the design (for example, the sampling of both students and schools from larger, similar populations). These assumptions lead to larger estimated standard errors than the finite-population analysis used by some of the study RCTs, which treat blocks as fixed and control for indicators of block membership as covariates in the models. For matched pair designs, the way in which covariates are treated in the analysis also accounts for some of the differences in this report's impact findings

[7] The theoretical finite population correction depends on the variance term pertaining to the individual treatment effects, S_τ^2, which is not identifiable because it is not possible to observe an individual in both the treatment and control conditions. Note, however, that $S_\tau^2 > \left(S_T - S_C\right)^2$. Thus, design-based estimators use Equation (11) as an upper bound (Schochet, 2016).

between the model- and design-based methods. In particular, including covariates in the estimation for a non-clustered matched pair design does not affect the impact estimate but does affect the standard error.[8]

[8] The PATE parameter is estimated using a two-stage method described in Schochet (2016).

III. Methods for comparison

The goal of this study is to compare design-based methods with commonly used, model-based methods by applying both methods to "real life" data from past RCTs in the education area. Another common way to compare different methods is to use a Monte Carlo approach in which data are simulated. This study uses "real life" data to complement the Monte Carlo analysis of Schochet (2016) by providing evidence on how much difference these methods make in practical applications.

In order to select methods that are commonly used in practice, the study estimates impacts by using a model-based method similar to the one used in the original evaluation study. In this way, the reference model-based methods used in this report represent what might have been used in the absence of design-based methods. Appendix I and Appendix II specify which model-based and design-based estimators were used for each study. Schochet (2016) provides a comprehensive formulation of the design-based estimators underlying each of the methods used in this report.

While the study uses a reference model-based method that captures the essence of those used in the original study, it sometimes modifies them slightly so that the essential features can be compared more directly to the design-based methods.[9] For example, both model- and design-based specifications include the same sets of covariates and account for missing covariates in the same way. The study standardizes these estimation features because they are not fundamental aspects of either model- or design-based approaches but nevertheless could affect the results.

This report focuses on estimating versions of the average treatment effect (ATE), some of which represent effects for a finite-population and some for a super-population. This parameter is the mean difference in outcomes between treatment and control groups (see Section II.A for a discussion). Both design- and model-based methods can estimate a range of other policy-relevant parameters. For example, some studies estimate the variation in impacts across sites or the treatment-on-the-treated (TOT) parameter (the effect of the intervention on the group that actually participated in the program). We focus on the ATE because it is the most commonly reported parameter in the original evaluation studies.

For evaluations that use clustered, blocked designs, the report presents results from two sets of models: (1) models that explicitly account for the blocks and (2) models that do not account for the blocks. The report presents both sets of analyses to provide additional evidence on clustered, non-blocked designs, because relatively few of these designs were identified in a search of available RCT datasets. Moreover, theory suggests the difference between model- and design-based methods is likely

[9] All model-based analyses were conducted using Stata 14. It is possible that other software packages could yield slightly different results if they apply different underlying formulae (Albright & Marinova, 2010).

to be the greatest for clustered designs (Schochet, 2016). Re-analyzing blocked, clustered designs as both blocked and non-blocked designs provides more evidence where it might be most informative.

The goal of this study is to compare the results from two different methods, not to replicate the findings from the original study or to provide new evidence from the evaluations that could directly inform policy. With this goal in mind, estimation methods and assumptions were chosen to facilitate comparisons between model- and design-based methods.

The report presents three separate sets of analyses that represent different comparisons between design-based and model-based methods:

- **Matched assumptions analyses.** These analyses compare the model-based methods to design-based methods that most closely match the assumptions of the original model-based methods.[10] For example, both methods use the same weighting scheme. This analysis provides information on what researchers employing design-based methods would obtain when adhering to the assumptions of commonly used, model-based methods.
- **Alternative assumptions analyses.** These analyses serve as a sensitivity check by comparing the model-based methods to an alternative set of design-based procedures, particularly the default settings in *RCT-YES.* These default settings represent a plausible alternative to the matched assumptions specification, and were informed by an expert panel of methodologists, including Guido Imbens and Donald Rubin, who are key developers of the underlying design-based methods. Additionally, the exploration of the default settings in *RCT-YES* could be directly relevant to practitioners, because *RCT-YES* is a publicly available software package that applies design-based methods for a wide range of evaluation designs. In several cases, the specifications in the matched assumptions analysis and alternative assumptions analysis are identical because the default *RCT-YES* settings align with the original assumptions.

  The alternative assumptions analyses explore the effect of using designed-based procedures that make different assumptions than the reference model-based method. For example, if the original study did not include sampling weights,[11] this analysis does not include explicitly defined weights in the model-based or design-based setup. The default weights implicit in these two approaches, however, can be different for clustered or blocked designs. These alternative specifications do not necessarily yield more (or less) correct answers than in the matched assumptions analyses, but instead yield parameters with different interpretations and underlying assumptions. Comparing the differences between the two sets of assumptions sheds light on the empirical importance of different design features.

[10] One feature we do not attempt to standardize is which tests and degrees of freedom are used for hypothesis testing.
[11] Sampling weights adjust the data for differences in the probability that observations were selected into the sample, e.g., they might account for oversampling of a particular subgroup.

- **Comparison of model-based approaches (HLM methods versus linear models with OLS assumptions and RCSE to account for clustering).** To contextualize the differences between the model-based methods and design-based methods in the first two analyses, the study compares estimates from HLM methods to linear models with OLS assumptions and RCSE to account for clustering (the reference model-based method for all of the evaluations considered in this report). These analyses provide a baseline level of what to expect in terms of how sensitive the estimates might be to slightly different methods.

III.A. Selecting key design features for the matched and alternative assumptions analyses

This section outlines key estimation features for the matched and alternative assumptions analyses. First, it discusses settings that are the same in these two analyses. Second, it highlights key differences between these two analyses. Appendix II provides additional details for each RCT and empirical specification.

Design features that are the same between the "matched assumptions" approach and the "alternative assumptions" approaches

A number of design features are the same for both the matched assumptions and the alternative assumptions analyses. Some of these features differ from the original evaluations, which might explain any differences between the findings in this report and previous evaluations. The following summaries describe those features:

Outcomes. For each study, the analysis includes the main (confirmatory) outcome variables presented in the original analysis. It excludes outcomes that were used for sensitivity analyses.

Covariates. The same covariates are used for both the model- and design-based analyses. When possible, the specifications include the covariates adopted in the original study, although in some cases, the publicly available versions of the datasets do not include the full set of covariates. In these cases, the specifications include covariates that are similar to those used in the original study. For both the design- and model-based approaches, we use individual-level covariates for clustered designs, even though design-based methods often use cluster-level averages (see Section II).

Missing covariates. Since imputation of missing covariates applies to both design- and model-based methods, both analyses use the same imputation methods.[12]

[12] Specifically, *RCT-YES*'s default settings are employed. When a covariate is missing 30 percent of the time or less for each outcome for both the treatment and control groups, mean imputation is conducted separately for the treatment and control groups. If the covariate is missing for 30 percent or more of the observations for either the treatment or control group, the covariate is excluded from the analysis.

Missing outcomes. Observations with missing data for outcomes are excluded, a common approach in both model- and design-based analyses.

Binary outcomes. In some cases, the original model-based analysis uses a nonlinear specification for binary outcomes, such as a logit model. This study uses a linear specification for the model-based methods to facilitate comparability across RCTs, because the goal is not to compare differences that arise between linear and non-linear specifications.

Block sample size checks. We conduct checks to ensure that there are enough treatment and control members within blocks to estimate impacts and standard errors. If there are too few treatment and control members for a block, it is excluded from the analysis. Specifically, for models with a block fixed effect and a block-by-treatment effect term, blocks are included if they contain at least two treatments and at least two controls with outcome data that vary so that standard errors can be calculated in each block. For matched pair designs or models with a block fixed effect (and no block-by-treatment effect), blocks are included if they contain at least one treatment and one control with outcomes that vary. These checks ensure that the data used in the analysis align with the original experimental design and that standard errors can be calculated.

Design features that differ between the "matched assumptions" approach and the "alternative assumptions" approach

There are three key differences between the "matched assumptions" and "alternative assumptions" analyses:

Block-by-treatment interactions. For the matched assumptions analyses, the study uses a method that is consistent with the model-based method: if the original study includes block-by-treatment interactions, then the specifications include those for the design-based approach, but if the original study uses block fixed effects only, the specifications include only block fixed effects in the design-based approach. For the alternative assumptions analysis, the study uses block-by-treatment interactions for the design-based method, regardless of what was applied in the original study.

Weights. For both the matched assumptions and alternative assumptions analyses, if the original analysis includes nonresponse weights or sampling weights, the specifications include these weights in the model-based and design-based methods. If the original analysis does not use weights, the matched assumptions analysis uses the implicit weights used in the original study. The alternative assumptions analysis uses a weighting scheme commonly employed in design-based procedures, as implemented by *RCT-YES*.[13]

[13] The default weights for *RCT-YES* differ by design. By default, *RCT-YES* weights individuals equally for non-clustered designs and weights clusters equally for clustered designs. Similarly, *RCT-YES* weights blocks by their total numbers of observations for non-clustered, blocked designs and by their total numbers of clusters for clustered, blocked designs.

Finite- vs. super-population assumptions. With design-based methods, the researcher must decide whether the results apply to a finite population or a super population. In the evaluations in this report, the original model-based methods are most consistent with finite-population assumptions.[14] For most designs, finite-population assumptions are also appropriate for design-based analyses. Therefore, for these designs, the finite-population model is assumed for both the matched assumptions analyses and the alternative assumptions analyses. The one exception is matched pair designs for which the design-based procedures estimate PATE, a super-population parameter. Therefore, for matched pair designs, in the matched assumptions analyses, the study specifies finite-population assumptions in the design-based approach, and for the alternative assumptions analyses, the study specifies super-population assumptions in the design-based approach.

III.B. Comparing an alternative model-based method to the reference model-based method (HLM vs. linear models with OLS assumptions and RCSE to account for clustering)

Model-based methods tend to fall into one of two categories: (1) those that use a linear model with OLS assumptions and account for clustering by using RCSE and (2) those that use an HLM approach and account for clustering by using cluster-level random effects. The study investigates the sensitivity of the results to the underlying assumptions by comparing these two model-based methods. It can also be viewed as a comparison between the reference model-based method and an alternative model-based method, since the reference model-based method for all evaluations is a linear model with OLS assumptions and RCSE when applicable.

The approach is similar to comparing design-based and model-based methods. To standardize the analyses, this study accounts for blocking by using blocked fixed effects in both models. To simplify estimation, neither model uses sampling weights. This study adopts the same approaches outlined in the preceding discussion for handling other design features, including the choice of outcomes and covariates, the treatment of missing outcomes and covariates, estimation methods for binary outcomes, and minimum sample size requirements for blocked designs. This analysis uses standard HLM methods for one-level and two-level designs (see appendix Table I.1).

[14] Model-based methods often assume a super-population at all levels of sampling, e.g., HLM methods often assume that clusters or blocks are drawn from a super population. The model-based methods from the original studies that we considered do not assume super-population models for the sampling of clusters and blocks, but are consistent with sampling of individuals from a super-population. For simplicity, we assume a finite population for all levels and designs in the matched settings comparison. We have, however, run the analysis assuming a super-population model in which individuals are resampled and we find very similar results.

IV. Summary of main results

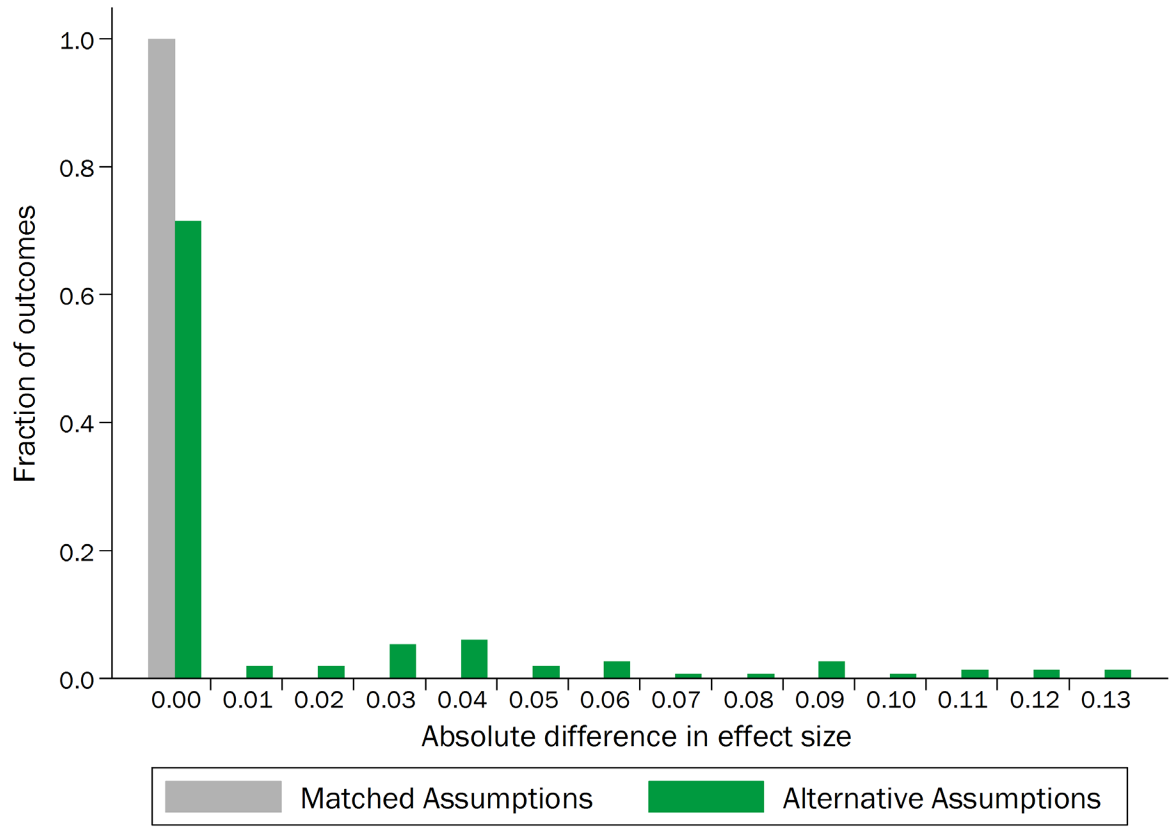
This section presents a comparison of the impact estimates and measures of statistical significance by using the model- and design-based methods for all outcomes across the study RCTs. To summarize these findings, the study creates aggregate statistics in three steps. First, for each estimation method and RCT, the study estimates impacts, standard errors, and *p*-values for each outcome. Second, the study converts the impact estimates and standard errors into effect size units by dividing the impacts and standard errors by the standard deviation of the outcome in the control group. Third, the report presents statistics that compare the effect sizes, standard errors, and *p*-values across the estimation methods.

Figure 2 provides a summary of our main findings. It displays the distribution of the absolute difference in effect sizes between the model-based methods and the design-based methods for both the matched assumptions analysis and the alternative assumptions. It reveals two patterns which are consistent with our more in-depth analysis. First, the differences for the matched assumptions analysis are indistinguishable across all outcomes. Second, it reveals that differences arise for the alternative assumptions analysis, suggesting that the estimates are sensitive to the assumptions underlying the design-based model. In the remaining sections, we further explore these differences.

For each study, we estimate impacts separately for one subgroup, because subgroup analyses are a common feature of many evaluations and have smaller sample sizes that could lead to some differences in findings. The subgroup estimates do not enter the aggregate statistics presented in this section.

Appendix II presents non-aggregated results, providing impact estimates separately for each outcome. Because the results for the subgroups show similar patterns to those of the full sample, we present these results in Appendix II as well.

Figure 2. Differences in Effect Sizes Between Matched and Alternative Assumptions Analyses



Note: Figure pertains to 151 outcomes across all studies.

IV.A. Main findings for the matched assumptions analyses

Table 2 summarizes the differences between the model- and design-based estimates for the matched assumptions analyses. The results correspond to the main impact findings from each study (excluding the estimates from the subgroup analyses). Detailed descriptions of the contents of the tables are provided in the accompanying notes below each table.

Overall, the findings from the design-based methods and the model-based methods are similar in terms of the magnitude of the impacts and the level of statistical significance. Across all RCTs, there are no differences in the magnitude of the estimates. This finding arises because both approaches use the same linear regression specifications and weights. Across all outcomes and RCTs, 5 of 151 estimates (3.31 percent) differ in significance.

- For non-clustered designs, the estimates from design- and model-based methods are indistinguishable. There are no outcomes for which the statistical significance differs.

- For clustered designs, the estimates from design- and model-based methods are indistinguishable in terms of the difference in the magnitude of the impacts, but there are a few differences in statistical significance. Across the outcomes for clustered RCTs, 5.21 percent of the estimates differ in significance. For three of the six clustered RCTs, there are no differences in statistical significance. For the remaining three RCTs, the estimates for five outcomes differ in statistical significance based on a 0.05 threshold. For all five outcomes, the *p*-values from one of the two methods are within 0.027 of being classified as the same significance level as the other, meaning at least one *p*-value is near the 0.05 threshold, with one above the threshold and the other below. The study explores these cases in detail to highlight that even when there are differences in statistical significance, they are relatively small and are partly due to the convention of using 0.05 as the cutoff for determining statistical significance:
  - For the Roads to Success study (see appendix Section II.F), a single outcome differs in statistical significance between the two methods. The model-based method yields an estimate that is statistically significant at the 5 percent level, whereas the design-based method yields a *p*-value of 0.066, which is 0.016 above the cutoff for statistical significance.
  - For the Health Teacher study (see appendix Section II.H), a single outcome differs in statistical significance. The model-based method yields a statistically significant estimate, while the design-based method does not. However, the *p*-value for the model-based estimate is 0.023, which 0.027 below the cutoff for statistical significance, and would not be classified as significant using a stricter threshold of 1 percent.
  - For the Power Through Choices (blocked version) study (see appendix Section II.I), three outcomes differ in statistical significance between the two methods. In all cases, the model-based method yields an estimate that is statistically significant at the 5 percent level. However, in two of these cases, the associated *p*-values are 0.045 and 0.048, which are 0.005 and 0.002 below the cutoff for statistical significance, respectively.
- The results for the subgroup estimates mirror those for the full sample specifications. The subgroup estimates, however, tend to exhibit slightly larger differences between the model- and design-based methods, possibly because the subgroup estimates are less precisely estimated, which leads to greater variability in general. See the supporting tables in Appendix II for a complete set of subgroup analysis findings for each evaluation.
- As a separate sensitivity analysis, the study conducts the comparisons without covariates. In this analysis, the results for the model- and design-based methods align even more closely.
- As another sensitivity analysis, we use cluster-level (not individual-level) data when estimating design-based specifications for clustered designs. This approach is often adopted with design-based methods (see Section II). In this analysis, the general patterns are similar, but there are larger differences in the average impacts.

Table 2. Summary of Differences between Design-Based and Model-Based Approaches for Impact Estimates (Matched Assumptions)

Study	Design-based specification		Model-based specification			Summary of results			
	RCT-YES design	Finite vs. super-population	Method	Block-by-treatment interactions	Weights	Number of estimates	Mean difference in effect size	Mean difference in SE (effect size)	Percentage of estimates that differ in significance
Job Corps	Non-clustered, non-blocked	Finite pop. (ATE)	OLS	N/A	✓	20	0.000	0.000	0.0
Teen Options to Prevent Pregnancy	Non-clustered, non-blocked	Finite pop. (ATE)	OLS	N/A		27	0.000	0.000	0.0
Teach For America	Non-clustered, blocked	Finite pop. (ATE)	OLS, Block FE	✓	✓	2	0.000	0.000	0.0
Charter School Impacts	Non-clustered, blocked	Finite pop. (ATE)	OLS, Block FE	✓	✓	4	0.000	0.005	0.0
Teacher Prep	Non-clustered, blocked	Finite pop. (ATE)	OLS, Block FE	✓	✓	2	0.000	0.000	0.0
Roads to Success	Clustered, non-blocked	Finite pop. (ATE)	OLS, RCSE	N/A	✓	17	0.000	0.023	5.9
TFA and Teaching Fellows Program	Clustered, non-blocked	Finite pop. (ATE)	OLS, RCSE	N/A	✓	2	0.000	0.000	0.0
Health Teacher	Clustered, non-blocked	Finite pop. (ATE)	OLS, RCSE	N/A	✓	27	0.000	0.025	3.7
POWER through Choices	Clustered, non-blocked	Finite pop. (ATE)	OLS, RCSE	N/A		24	0.000	0.002	0.0
TFA and Teaching Fellows Program	Clustered, blocked (Block FE)	Finite pop. (ATE)	OLS, Block FE, RCSE		✓	2	0.000	0.005	0.0
POWER through Choices	Clustered, blocked (Block FE)	Finite pop. (ATE)	OLS, Block FE, RCSE			24	0.000	0.026	12.5

Note: The table presents results that compare estimates from model-based methods and design-based methods as implemented by *RCT-YES*. The settings in *RCT-YES* are chosen to align most closely to the original model-based methods. ***RCT-YES* design** indicates which randomization design is specified in *RCT-YES*. **Finite vs. super-population** indicates whether finite- or super-population assumptions are selected in *RCT-YES*. **Method** indicates the method used for the model-based estimation. **OLS** indicates that the model is estimated by using ordinary least squares. **Block FE** indicates that the model includes block fixed effects. **RCSE** indicates that robust cluster standard errors are used. **Block-by-treatment interactions** indicates whether the model-based specification includes block-by-treatment interaction terms so that a different treatment effect is estimated for each block. **Weights** indicates whether the model-based specification includes explicit weights. If explicit weights are not specified, the implicit weights in the model-based method are specified in *RCT-YES* so that the weights are standardized between the two methods. **Number of estimates** indicates the number of different outcomes for which impacts are estimated. **Mean difference in effect size** is the average of the absolute value of the difference in effect size between the model- and design-based approaches, in which the effect sizes are calculated by dividing the impact estimate by the standard deviation of the outcome in the control group. **Mean difference in SE (effect size)** is the average of the absolute value of the difference in standard errors in effect size units between the model- and design-based approaches, in which the standard errors are calculated by dividing the standard error of the impact estimate by the standard deviation of the outcome in the control group. **Percentage of estimates that differ in significance** indicates the percentage of estimates that differ in significance at the 5 percent level.

IV.B. Main findings for the alternative assumptions analyses

The results of the alternative assumptions analyses are similar to the matched assumptions analyses with a few exceptions that accord with statistical theory underlying these estimators. Table 3 summarizes the differences for the alternative assumptions analyses.

- Overall, findings using the design-based and model-based methods are similar in both the magnitude of the impacts and the level of statistical significance, but the differences are slightly larger than for the matched setting analyses. Across all RCTs, the largest average difference in estimated effect sizes across outcomes is 0.051 (while there were no differences for the matched assumptions analyses). Across all outcomes and RCTs, 4.64 percent of estimates differ in significance (two more total outcomes than in the matched assumptions analyses).
- One reason that the alternative assumptions analyses align closely with the matched assumptions analyses presented in Table 2 is that the assumptions differ for only 4 out of 11 studies; that is, in most cases widely used design-based assumptions align closely with those in the original studies. The bold font in the name of the studies in Table 3 indicates that the alternative assumptions differ from the matched assumptions.
- The only noticeable differences in the magnitude of impacts are for the POWER Through Choices (clustered, non-blocked) and POWER Through Choices (clustered, blocked) studies (see appendix Sections II.I and II.K). While these differences could theoretically arise for several possible reasons as outlined in Section III.A, additional analyses reveal that the differences are primarily due to the difference in weights between the two methods. When standardizing the weights but using the other alternative design-based assumptions, the mean differences become negligible. This finding shows that results can be sensitive to different weighting schemes.
- Compared to the matched assumptions analyses, the alternative assumptions analysis of POWER Through Choices (clustered, blocked) yield one additional estimate that differs in significance between the model- and design-based approaches. In particular, the impact on "Intentions to use a condom" is statistically significant for the model-based method but not for the design-based method. This finding arises for two reasons: (1) the impact estimate for the design-based method is closer to zero largely due to the difference in weights and (2) it is estimated less precisely because it is a super-population parameter.[15] The second difference is consistent with the theory on statistical sampling for super-population models because the

[15] The model-based impact estimate and standard error of "Intentions to use a condom" are 14.07 and 2.55, respectively. For the alternative assumptions analysis, the design-based impact estimate and standard error are 8.84 and 4.47, respectively. For the design-based analysis with standardized weights, the impact estimate and standard error are 14.07 and 3.80, respectively.

super-population model accounts for sampling which increases the variance in the impact estimates (see Section II.C).

- Despite the small differences in POWER Through Choices, the analyses from the alternative assumptions and matched assumptions analyses provide similar estimates and suggest the same policy conclusions. Note that neither the matched assumptions analysis nor the alternative assumptions analysis is more correct than the other—they simply have different interpretations and assumptions.

IV.C. Main findings for the comparison of an alternative model-based method to the reference model-based method (HLM vs. linear models with OLS assumptions and RCSE to account for clustering)

To benchmark the findings from Table 2 and Table 3, the study conducted analyses to examine how results compare between HLM estimators and linear models with OLS assumptions and RCSE used in the original studies. Table 4 summarizes the differences between the original model-based method and an alternative model-based method.

The differences in estimates between these two model-based methods are similar in prevalence to the differences between the matched assumptions analysis and the alternative assumptions analysis. Across all RCTs, the largest average difference in estimated effect sizes is 0.021. This difference is larger than that in the matched assumptions analysis but smaller than that in the alternative assumptions analysis. Across all outcomes and RCTs, 3.31 percent of estimates differ in significance (the same as in the alternative assumptions analysis and more than in the matched assumptions analysis). As with the earlier analyses, most (4 out of 5) differences in statistical significance arise in clustered RCTs.

These findings suggest that the differences between design-based and model-based methods with similar assumptions are no greater than those between different model-based approaches.

Table 3. Summary of Differences between Design-Based and Model-Based Approaches for Impact Estimates (Alternative Assumptions Analysis)

Study	Design-based specification		Model-based specification			Summary of results			
	RCT-YES design	Finite vs. super-population	Method	Block-by-treatment interactions	Weights	Number of estimates	Mean difference in effect size	Mean difference in SE (effect size)	Percentage of estimates that differ in significance
Job Corps	Non-clustered, non-blocked	Finite pop. (ATE)	OLS	N/A	✓	20	0.000	0.000	0.0
Teen Options to Prevent Pregnancy	Non-clustered, non-blocked	Finite pop. (ATE)	OLS	N/A		27	0.000	0.000	0.0
Teach For America	Non-clustered, blocked	Finite pop. (ATE)	OLS, FE	✓	✓	2	0.000	0.000	0.0
Charter School Impacts	Non-clustered, blocked	Finite pop. (ATE)	OLS, FE	✓	✓	4	0.000	0.005	0.0
Teacher Prep	Non-clustered, blocked	Finite pop. (ATE)	OLS, FE	✓	✓	2	0.000	0.000	0.0
Roads to Success	Clustered, non-blocked	Finite pop. (ATE)	OLS, FE, RCSE	N/A	✓	17	0.000	0.023	5.9
TFA and Teaching Fellows Program	Clustered, non-blocked	Finite pop. (ATE)	OLS, RCSE	N/A	✓	2	0.000	0.000	0.0
Health Teacher	Clustered, non-blocked	Finite pop. (ATE)	OLS, RCSE	N/A	✓	27	0.000	0.025	3.7
POWER Through Choices	Clustered, non-blocked	Finite pop. (ATE)	OLS, RCSE	N/A		24	0.050	0.016	4.2
TFA and Teaching Fellows Program	Clustered, blocked	Super pop. (PATE)	OLS, Block FE, RCSE		✓	2	0.000	0.010	0.0
POWER Through Choices	Clustered, blocked	Super pop. (PATE)	OLS, Block FE, RCSE			24	0.051	0.031	16.7

Note: The table presents results that compare estimates from model-based methods and design-based methods as implemented by *RCT-YES*. The default settings in *RCT-YES* are used to estimate the design-based models. Bolded study names indicate that the alternative assumptions differ from the matched assumptions. ***RCT-YES* design** indicates which randomization design is specified in *RCT-YES*. **Finite vs. super-population** indicates whether finite- or super-population assumptions are selected in *RCT-YES*. **Method** indicates the method used for the model-based estimation. **OLS** indicates that the model is estimated by using ordinary least squares. **Block FE** indicates that the model includes block fixed effects. **RCSE** indicates that robust cluster standard errors are used. **Block-by-treatment interactions** indicates whether the model-based specification includes block-by-treatment interaction terms so that a different treatment effect is estimated for each block. **Weights** indicates whether the model-based specification includes explicit weights. If explicit weights are not specified, the default weights are used in *RCT-YES*. **Number of estimates** indicates the number of different outcomes for which impacts are estimated. **Mean difference in effect size** is the average of the absolute value of the difference in effect size between the model- and design-based approaches, in which the effect sizes are calculated by dividing the impact estimate by the standard deviation of the outcome in the control group. **Mean difference in SE (effect size)** is the average of the absolute value of the difference in standard errors in effect size units between the model- and design-based approaches, in which the standard errors are calculated by dividing the standard error of the impact estimate by the standard deviation of the outcome in the control group. **Percentage of estimates that differ in significance** indicates the percentage of estimates that differ in significance at the 5 percent level.

Table 4. Summary of Differences between the Alternative and Reference Model-Based Specifications (HLM vs. Linear Models with OLS Assumptions and RCSE to Account for Clustering)

Study	Model-based specifications		Summary of results			
	Reference	Alternative	Number of estimates	Mean difference in effect size	Mean difference in SE (effect size)	Percentage of estimates that differ in significance
Job Corps	OLS	MLE	20	0.000	0.000	0.0
Teen Options to Prevent Pregnancy	OLS	MLE	27	0.000	0.000	0.0
Teach For America	OLS, Block FE	MLE, Block FE	2	0.000	0.000	0.0
Charter School Impacts	OLS, Block FE	MLE, Block FE	4	0.000	0.000	0.0
Teacher Prep	OLS, Block FE	MLE, Block FE	2	0.000	0.000	50.0
Roads to Success	OLS, RCSE	MLE, Cluster RE	17	0.021	0.044	5.9
TFA and Teaching Fellows Program	OLS, RCSE	MLE, Cluster RE	2	0.015	0.005	0.0
Health Teacher	OLS, RCSE	MLE, Cluster RE	27	0.001	0.012	11.1
POWER Through Choices	OLS, RCSE	MLE, Cluster RE	24	0.018	0.004	0.0
TFA and Teaching Fellows Program	OLS, Block FE, RCSE	MLE, Block FE, Cluster RE	2	0.000	0.005	0.0
POWER Through Choices	OLS, Block FE, RCSE	MLE, Block FE, Cluster RE	24	0.004	0.008	0.0

Note: The table presents results that compare estimates from the original model-based methods and alternative model-based specification. **Reference specification** indicates the method used for the original model-based estimation. **Alternative specification** indicates the method used for the alternative model-based estimation. **OLS** indicates that the model is estimated by using ordinary least squares. **Block FE** indicates that the model includes block fixed effects. **RCSE** indicates that robust cluster standard errors are used. **MLE** indicates that the model is estimated by using maximum likelihood. **Cluster RE** indicates that the model includes random effects at the cluster-level. **Number of estimates** indicates the number of different outcomes for which impacts are estimated. **Mean difference in effect size** is the average of the absolute value of the difference in effect size between the original and alternative model-based approaches, in which the effect sizes are calculated by dividing the impact estimate by the standard deviation of the outcome in the control group. **Mean difference in SE (effect size)** is the average of the absolute value of the difference in standard errors in effect size units between the original and alternative model-based approaches, in which the standard errors are calculated by dividing the standard error of the impact estimate by the standard deviation of the outcome in the control group. **Percentage of estimates that differ in significance** indicates the percentage of estimates that differ in significance at the 5 percent level. Both the original and alternative specifications account for blocked designs that use block fixed effects without interaction terms.

V. Conclusions

The results of this study suggest that the design-based methods and commonly used, model-based methods yield similar estimates for "real-world" RCTs that span a variety of different designs and contexts. When assumptions which most closely match the model-based methods are used, design-based methods yield estimates that are similar to the model-based methods both in the size of the impacts and in statistical significance. The estimates for non-clustered designs are particularly similar. The few differences for clustered designs are small and unlikely to change policy conclusions. Importantly, the differences between design- and model-based estimates are no larger than those that arise between commonly used model-based methods (linear models with OLS assumptions and RCSE to account for clustering vs. HLM estimators).

When an alternative set of design-based assumptions are used as sensitivity analysis, design-based methods still yield estimates that are similar to model-based methods. The few additional differences can be explained by differences in weights and assumptions about sampling. These design-based estimates are neither more nor less correct than the model-based estimates but simply have different interpretations.

The results from this study highlight that competing impact estimation methods that are asymptotically unbiased but rely on different assumptions can yield different estimates in finite samples. The design-based and model-based estimators examined in this report fall into this class of estimators. It is reassuring that, based on the datasets considered in this report, different estimation methods typically yield similar findings, especially when the underlying assumptions are aligned. This finding suggests that users of any method (including design-based methods) should select the assumptions that best suit the goals of their study and may want to examine the robustness of their study findings by using alternative specifications—in particular, the way in which blocks and clusters are weighted for the analysis and the choice of the finite- versus super-population models.

References

Albright, J. J., & Marinova, D. M. (2010). *Estimating multilevel models using SPSS, Stata, SAS, and R.* Unpublished manuscript.

Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the royal statistical society. Series B (Methodological)*, 289-300.

Cameron, A. C., & Miller, D. L. (2015). A practitioner's guide to cluster-robust inference. *Journal of Human Resources, 50*(2), 317-372.

Chaplin, D., Bleeker, M., & Booker, K. (2010). *Roads to Success: Estimated impacts of an education and career planning program during middle school.* Final report. (No. 6388-700). Princeton, NJ: Mathematica Policy Research.

Clark, M. A., Chiang, H. S., Silva, T., McConnell, S., Sonnenfeld, K., Erbe, A., & Puma, M. (2013). *The effectiveness of secondary math teachers from Teach For America and the Teaching Fellows programs.* (No. NCEE 2013-4015). Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education.

Constantine, J., Player, D., Silva, T., Hallgren, K., Grider, M., & Deke, J. (2009). *An evaluation of teachers trained through different routes to certification.* (No. NCEE 2009-4043). Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education.

Decker, P. T., Mayer, D. P., & Glazerman, S. (2004). *The effects of Teach For America on students: Findings from a national evaluation.* Princeton, NJ: Mathematica Policy Research.

Freedman, D.A. (2008). On regression adjustments to experimental data. *Advances in Applied Mathematics, 40,* 180-193.

Gleason, P., Clark, M., Tuttle, C. C., & Dwoyer, E. (2010). *The evaluation of charter school impacts: Final report.* (No. NCEE 2010-4029). Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education.

Goesling, B., Colman, S., Scott, M., & Cook, E. (2014). Impacts of an enhanced family health and sexuality module of the HealthTeacher middle school curriculum. Princeton, NJ: Mathematica Policy Research.

Goesling, B., Covington, R. D., Manlove, J., Berry, M., Oman, R. F., & Vesley, S. (2015). *Interim impacts of the POWER Through Choices program.* Princeton, NJ: Mathematica Policy Research.Holland, P. W. (1986). Statistics and causal inference. *Journal of the American Statistical Association, 81*(396), 945-960.

Hothorn, T., Bretz, F., & Westfall, P. (2008). Simultaneous inference in general parametric models. *Biometrical Journal, 50*(3), 346-363.

Huber, P. J. (1967). The behavior of maximum likelihood estimates under nonstandard conditions. *Procedures of the Fifth Berkeley Symposium on Math and Statistical Probability*, *1*, 221-233.

Imai, K., King, G., & Nall, C. (2009). The essential role of pair matching in cluster-randomized experiments, with application to the Mexican Universal Health Insurance Evaluation. *Statistical Science*, 24, 29-53.

Imbens, G. (2004). Nonparametric estimation of average treatment effects under exogeneity: A review. *Review of Economics and Statistics,* 86, 4-29.

Imbens, G., & Rubin, D. (2015). Causal inference for statistics, social, and biomedical sciences: An introduction. Cambridge UK: Cambridge University Press.

Liang, K., & Zeger, S. (1986). Longitudinal data analysis using generalized linear models. *Biometrika, 73*, 13-22.

Lin, W. (2013). Agnostic notes on regression adjustments to experimental data: Reexamining Freedman's critique. *The Annals of Applied Statistics, 7(1),* 295-318.

Raudenbush, S. W., & Bryk, A. (2002). Hierarchical linear models: Applications and data analysis methods. Newbury Park, CA: Sage.

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

Schochet, P.Z. (2010). Is regression adjustment supported by the Neyman model for causal inference? *Journal of Statistical Planning and Inference, 140*, 246–259.

Schochet, P. Z. (2013). Estimators for clustered education RCTs using the Neyman model for causal inference. *Journal of Educational and Behavioral Statistics, 38*(3), 219-238.

Schochet, P. Z. (2016). *Statistical theory for the RCT-YES software: Design-based causal inference for RCTs.* (No. NCEE 2015–4011). Washington, DC: U.S: Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance, Analytic Technical Assistance and Development.

Schochet, P. Z., Burghardt, J., & Glazerman, S. (2001). *National Job Corps study: The impacts of Job Corps on participants' employment and related outcomes*. Princeton, NJ: Mathematica Policy Research.
Smith, K., Rotz, D., Goesling, B., Cook, E., Murphy, K., & Stevens, J. (2015). *Interim impacts of the Teen Options to Prevent Pregnancy Program.* Princeton, NJ: Mathematica Policy Research.

White, H. (1980). A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity. *Econometrica*, *48*, 817-838.

Appendix I. Hierarchical linear model methods

This appendix provides details for the hierarchical linear model (HLM) methods used in the alternative assumptions analysis. For each of the four designs (non-clustered, non-blocked; non-clustered, blocked; clustered, non-blocked; clustered, blocked), the estimation equation and abbreviation used in Table 4 of the main text are detailed in Table I.1.

Table I.1. Summary of HLM Methods used for Alternative Assumptions Analysis

Design	Abbreviation used in results table	Estimating equation	Definitions and assumptions
Non-clustered, non-blocked	MLE	$y_i = \beta_0 + \beta_1 X_i + \beta_{ATE} T_i + \varepsilon_i$	i – individual y_i – outcome T_i – treatment status of individual X_i – vector of covariates ε_i – individual error, $\varepsilon_i \sim N\left(0, \sigma_\varepsilon^2\right)$
Non-clustered, blocked	MLE, Block FE	$y_{ik} = \beta_{0k} + \beta_1 X_{ik} + \beta_{ATE} T_{ik} + \varepsilon_{ik}$	i – individual k – block y_{ik} – outcome T_{ik} – treatment status X_{ik} – vector of covariates β_{0k} – block fixed effect ε_{ik} – individual error, $\varepsilon_{ik} \sim N\left(0, \sigma_\varepsilon^2\right)$
Clustered, non-blocked	MLE, Cluster RE	$y_{ij} = \beta_0 + \beta_1 X_{ij} + \beta_{ATE} T_j + \nu_j + \varepsilon_{ij}$	i – individual j – cluster y_{ij} – outcome T_j – treatment status of cluster X_{ij} – vector of covariates ε_{ij} – individual error, $\varepsilon_{ij} \sim N\left(0, \sigma_\varepsilon^2\right)$ ν_j – cluster random effect, $\nu_j \sim N\left(0, \sigma_\nu^2\right)$ ν_j and ε_{ij} are independent

Design	Abbreviation used in results table	Estimating equation	Definitions and assumptions
Clustered, blocked	MLE, Block FE, Cluster RE	$y_{ijk} = \beta_{0k} + \beta_1 X_{ijk} + \beta_{ATE} T_{jk} + \nu_{jk} + \varepsilon_{ijk}$	i – individual j – cluster k – block y_{ijk} – outcome T_{jk} – treatment status of cluster X_{ijk} – vector of covariates β_{0k} – block fixed effect ε_{ijk} – individual error, $\varepsilon_{ijk} \sim N\left(0, \sigma_{\varepsilon}^2\right)$ ν_{jk} – cluster random effect, $\nu_{jk} \sim N\left(0, \sigma_{\nu}^2\right)$ ν_{jk} and ε_{ijk} are independent

Note: All equations are estimated using maximum likelihood.

Block FE = block fixed effects; Cluster RE = cluster-level random effects; MLE = maximum likelihood estimate.

Appendix II. Detailed description of studies and results

This appendix provides details for each evaluation. We describe the model-based and design-based methods used for each study as well as any specific settings applied in *RCT-YES*. See Schochet (2016) for the estimating equations used for each of the settings in *RCT-YES*.

For each study, two types of results are presented:

- **Impact estimates.** For each study, we present three tables of impact estimates: (1) the model-based analysis, (2) the design-based analysis with settings that align most closely to the model-based analysis ("matched settings"), and (3) the design-based analysis with *RCT-YES*'s default settings ("default settings").

  For each outcome, the tables present estimates of the control group mean (weighted when appropriate), the impact, the standard error, and the *p*-value. The stars indicate that the impact is statistically significant based on tests of a single hypothesis. For studies in which we adjust for multiple hypotheses, the "^" indicates that the results are statistically significant at the 5 percent level after applying the Benjamini-Hochberg correction within outcome domains. The tables that present the design-based estimates also present information on how the design-based estimates compare to the model-based estimates, including the differences in the effect sizes, standard errors of the effect sizes, and *p*-values.

- **Estimates of key features that might explain the differences.** For each study, we present a fourth table that details several features of the data that could potentially explain differences between the treatment and control groups:

  - **Standard deviation of outcomes and outcome residuals.** In contrast to many model-based methods (e.g., standard HLM models), design-based methods allow for differences in the variance in outcomes between the treatment and control groups.

  - **Intracluster correlations.** The intracluster correlation (ICC) is an important component of power calculations for cluster randomized controlled trials (RCTs). It is defined as the average correlation between outcome residuals within a cluster. Because of its importance in power calculations, we report estimates of the ICC for each method. With the assumptions under standard HLM models (and random effects models), the ICC is the fraction of the total variance across individuals due to variation between groups (e.g., schools). For model-based methods, such as ordinary least squares (OLS) with fixed effects and robust cluster standard error (RCSE), other estimates of the ICC are more appropriate. We also present estimates of the ICC from *RCT-YES*.[16]

[16] For a description of how *RCT-YES* calculates the ICC, see Schochet (2016).

II.A. Job Corps

Original study. Schochet et al. (2001).

Sponsor agency. Employment and Training Administration, U.S. Department of Labor.

Description of intervention. Job Corps is an intensive and comprehensive program that provides employment assistance to disadvantaged youths aged 16 to 24. The program offers many services, including academic education, vocational training, residential living, health care and health education, counseling, and job placement assistance.

Randomization design. Non-clustered, non-blocked. In the original study, 5,977 eligible applicants were randomly assigned to a control group. During the same time period, 9,409 eligible applicants were randomly assigned to the program (treatment) group. The members of the treatment and control groups were randomly selected from a fully national sample of eligible Job Corps applicants.

Model-based method. Ordinary least squares (OLS).

We use OLS to estimate the following equation:

$$(13)\quad Y_i = \alpha + \delta T_i + \varepsilon_i,$$

where Y_i is the outcome for person i, T_i is a treatment indicator, and ε_i is an individual-specific error term. We assume that ε_i is uncorrelated between individuals. To be consistent with the main specification from the original study, the model does not include covariates.

Design-based specification: RCT-YES model. Non-clustered, non-blocked.

Design-based specification: RCT-YES parameter (matched settings). Finite population (ATE).

Design-based specification: RCT-YES parameter (default settings). Finite population (ATE).

Covariates. In our analysis, we do not include covariates. This is consistent with the differences-in-means approach in the original report.

Weights. Yes. We use weights for both the model-based and design-based approaches. The weights account for both sampling and nonresponse. In particular, the weights account for sampling because some groups had different probabilities of being selected into the research sample.

Multiple hypothesis correction. None. We do not apply a multiple hypothesis correction, because one was not applied to the main results in the original study. The original study, however, did conduct sensitivity checks to address the issue of multiple hypothesis testing.

Main differences from original analyses. Our model-based methods differ in several ways from the original analysis:

- The standard errors in the original study accounted for clustering caused by the selection of areas for in-person interviewing at baseline. All members of the study sample were contacted by phone within 45 days after assignment. Members in selected areas who could not be reached by phone in the first 45 days were released into the field, and in-person interviews were attempted instead of phone interviews. The original study accounted for possible correlation between outcomes among youths who were sampled after the 45-day period in the same geographic areas. We do not account for this feature in either the design- or model-based estimates.

- The original study used a differences-in-means approach. For consistency with other evaluations, we estimate a linear regression with a constant term and treatment effect.

Outcomes. We estimate the model for the following outcomes:

Ever enrolled in education or training program: Enrollment in an education or training program during the 48 months after random assignment.

Hours per week ever in education or training program: Hours per week ever in an education or training program during the 48 months after random assignment.

Ever took academic classes: Ever took academic classes during the 48 months after random assignment.

Hours per week ever in academic classes: Hours per week ever in academic classes during the 48 months after random assignment.

Ever took vocational training: Ever received vocational training during the 48 months after random assignment.

Hours per week ever in vocational training: Hours per week ever received vocational training during the 48 months after random assignment.

Ever enrolled in a vocational program: Ever attended vocational, technical, or trade school during the 48 months after random assignment.

Ever enrolled in a two-year college program: Ever attended a two-year college during the 48 months after random assignment.

Ever enrolled in a four-year college program: Ever attended a four-year college during the 48 months after random assignment.

Received a vocational degree: Received a vocational degree or certificate during the 48 months after random assignment.

Received a college degree: Received a two-year or four-year college degree during the 48 months after random assignment.

Earnings per week, Q16: Earnings per week, 16 quarters after random assignment.

Ever employed, Q16: Ever employed, 16 quarters after random assignment.

Weeks employed, Q16: Percentage of weeks employed, 16 quarters after random assignment.

Hours worked, Q16: Hours employed per week, 16 quarters after random assignment.

Amount of benefits received: Dollar amount of AFDC/TANF, food stamp, SSI/SSA, or GA benefits received during the 48 months after random assignment.

Ever arrested or charged with a delinquency or criminal complaint: Ever arrested or charged with a delinquency or criminal complaint during the 48 months after random assignment.

Convicted, pled guilty, or adjudged delinquent: Ever convicted, pled guilty, or adjudged delinquent during the 48 months after random assignment.

Served time in jail for convictions: Ever served time in jail for convictions during the 48 months after random assignment.

Weeks in jail for convictions: Number of weeks ever in jail for convictions during the 48 months after random assignment.

Table II.A.1. Job Corps: Evaluation Impact Estimates (Model-Based)

Outcome	Sample	Control Mean	Impact	SE	p-value	Effect Size
Amount of benefits received	Full Sample	4155.73	-459.76	134.93	0.001*	-0.06
	Subgroup:				0.855	
	Male	2075.56	-461.85	165.78	0.005*	-0.10
	Female	7068.16	-508.59	195.28	0.009*	-0.06
Convicted, pled guilty, or adjudged delinquent	Full Sample	25.17	-3.08	0.80	0.000*	-0.07
	Subgroup:				0.126	
	Male	34.89	-3.97	1.01	0.000*	-0.08
	Female	10.98	-1.56	1.21	0.198	-0.05
Earnings per week, Q16	Full Sample	199.35	18.11	4.07	0.000*	0.09
	Subgroup:				0.464	
	Male	225.55	20.84	5.23	0.000*	0.10
	Female	161.13	14.84	6.30	0.018*	0.08
Ever arrested or charged with a delinquency or criminal complaint	Full Sample	32.56	-3.74	0.87	0.000*	-0.08
	Subgroup:				0.035*	
	Male	43.52	-5.06	1.09	0.000*	-0.10
	Female	16.47	-1.48	1.31	0.257	-0.04
Ever employed, Q16	Full Sample	0.69	0.02	0.01	0.006*	0.04
	Subgroup:				0.496	
	Male	0.71	0.02	0.01	0.090	0.04
	Female	0.65	0.03	0.01	0.022*	0.06
Ever enrolled in a four-year college program	Full Sample	3.37	-0.09	0.35	0.784	0.00
	Subgroup:				0.861	
	Male	2.50	-0.05	0.45	0.910	0.00
	Female	4.62	-0.17	0.54	0.746	-0.01
Ever enrolled in a two-year college program	Full Sample	12.31	-0.82	0.62	0.190	-0.02
	Subgroup:				0.051	
	Male	8.89	0.18	0.81	0.828	0.01
	Female	17.22	-2.28	0.96	0.018*	-0.06
Ever enrolled in a vocational program	Full Sample	28.63	-2.52	0.86	0.003*	-0.06
	Subgroup:				0.703	
	Male	24.97	-2.82	1.11	0.011*	-0.07
	Female	33.86	-2.16	1.33	0.104	-0.05
Ever enrolled in education or training program	Full Sample	0.72	0.21	0.01	0.000*	0.47
	Subgroup:				0.000*	
	Male	0.69	0.23	0.01	0.000*	0.50
	Female	0.76	0.17	0.01	0.000*	0.40
Ever took academic classes	Full Sample	0.57	0.24	0.01	0.000*	0.50
	Subgroup:				0.739	
	Male	0.56	0.24	0.02	0.000*	0.50
	Female	0.58	0.23	0.02	0.000*	0.48

Outcome	Sample	Control Mean	Impact	SE	p-value	Effect Size
Ever took vocational training	Full Sample	0.28	0.46	0.01	0.000*	1.05
	Subgroup:				0.070	
	Male	0.26	0.47	0.01	0.000*	1.11
	Female	0.32	0.43	0.02	0.000*	0.95
Hours per week ever in academic classes	Full Sample	2.50	0.58	0.11	0.000*	0.15
	Subgroup:				0.038*	
	Male	2.54	0.40	0.14	0.003*	0.10
	Female	2.42	0.85	0.17	0.000*	0.22
Hours per week ever in education or training program	Full Sample	4.12	3.45	0.11	0.000*	0.65
	Subgroup:				0.673	
	Male	3.87	3.49	0.15	0.000*	0.67
	Female	4.48	3.40	0.17	0.000*	0.62
Hours per week ever in vocational training	Full Sample	0.94	2.20	0.09	0.000*	0.92
	Subgroup:				0.636	
	Male	0.82	2.23	0.11	0.000*	1.01
	Female	1.12	2.15	0.14	0.000*	0.82
Hours worked, Q16	Full Sample	26.40	1.46	0.43	0.001*	0.06
	Subgroup:				0.931	
	Male	28.96	1.46	0.56	0.009*	0.06
	Female	22.65	1.54	0.67	0.022*	0.07
Received a college degree	Full Sample	1.54	-0.21	0.22	0.355	-0.02
	Subgroup:				0.211	
	Male	1.08	0.02	0.29	0.947	0.00
	Female	2.21	-0.55	0.35	0.117	-0.04
Received a vocational degree	Full Sample	15.19	22.27	0.80	0.000*	0.62
	Subgroup:				0.238	
	Male	13.21	21.45	1.04	0.000*	0.63
	Female	18.10	23.37	1.26	0.000*	0.61
Served time in jail for convictions	Full Sample	17.95	-2.11	0.71	0.003*	-0.05
	Subgroup:				0.084	
	Male	25.96	-3.02	0.89	0.001*	-0.07
	Female	6.24	-0.61	1.07	0.570	-0.03
Weeks employed, Q16	Full Sample	59.02	2.79	0.85	0.001*	0.06
	Subgroup:				0.955	
	Male	61.71	2.78	1.11	0.012*	0.06
	Female	55.10	2.87	1.33	0.031*	0.06

Outcome	Sample	Control Mean	Impact	SE	p-value	Effect Size
Weeks in jail for convictions	Full Sample	6.57	-0.58	0.44	0.192	-0.02
	Subgroup:				0.168	
	Male	10.59	-1.04	0.56	0.066	-0.03
	Female	0.71	0.18	0.68	0.791	0.03

Note: The p-value on the subgroup rows tests for differences in impacts across subgroup categories. Effect sizes are calculated by dividing the impact estimate by the standard deviation of the outcome in the control group.
* Difference is statistically significant at the 0.05 level, two-tailed test.

Table II.A.2. Job Corps: Evaluation Impact Estimates (Design-Based, Matching Assumptions)

Outcome	Sample	Control Mean	Impact	SE	p-value	Effect Size	Absolute Value of Difference Compared to Model-Based Specification		
							Dif. Effect Size	Dif. SE (Effect Size)	Dif. p-value
Amount of benefits received	Full Sample	4155.73	-459.76	140.33	0.001*	-0.06	0.00	0.00	0.000
	Subgroup:				0.874				0.019
	Male	2075.56	-461.85	118.49	0.000*	-0.10	0.00	0.01	0.005
	Female	7068.16	-508.59	270.15	0.060	-0.06	0.00	0.01	0.051
Convicted, pled guilty, or adjudged delinquent	Full Sample	25.17	-3.08	0.85	0.000*	-0.07	0.00	0.00	0.000
	Subgroup:				0.123				0.003
	Male	34.89	-3.97	1.21	0.001*	-0.08	0.00	0.00	0.001
	Female	10.98	-1.56	0.98	0.113	-0.05	0.00	0.01	0.085
Earnings per week, Q16	Full Sample	199.35	18.11	4.25	0.000*	0.09	0.00	0.00	0.000
	Subgroup:				0.463				0.001
	Male	225.55	20.84	5.95	0.000*	0.10	0.00	0.00	0.000
	Female	161.13	14.84	5.60	0.008*	0.08	0.00	0.00	0.010
Ever arrested or charged with a delinquency or criminal complaint	Full Sample	32.56	-3.74	0.92	0.000*	-0.08	0.00	0.00	0.000
	Subgroup:				0.037*				0.002
	Male	43.52	-5.06	1.26	0.000*	-0.10	0.00	0.00	0.000
	Female	16.47	-1.48	1.17	0.204	-0.04	0.00	0.00	0.053
Ever employed, Q16	Full Sample	0.69	0.02	0.01	0.010*	0.04	0.00	0.00	0.004
	Subgroup:				0.526				0.030
	Male	0.71	0.02	0.01	0.100	0.04	0.00	0.00	0.010
	Female	0.65	0.03	0.02	0.037*	0.06	0.00	0.02	0.015
Ever enrolled in a four-year college program	Full Sample	3.37	-0.09	0.37	0.796	0.00	0.00	0.00	0.012
	Subgroup:				0.876				0.015
	Male	2.50	-0.05	0.41	0.901	0.00	0.00	0.00	0.009
	Female	4.62	-0.17	0.67	0.796	-0.01	0.00	0.01	0.050

Outcome	Sample	Control Mean	Impact	SE	p-value	Effect Size	Absolute Value of Difference Compared to Model-Based Specification		
							Dif. Effect Size	Dif. SE (Effect Size)	Dif. p-value
Ever enrolled in a two-year college program	Full Sample	12.31	-0.82	0.66	0.218	-0.02	0.00	0.00	0.028
	Subgroup:				0.081				0.030
	Male	8.89	0.18	0.75	0.815	0.01	0.00	0.00	0.013
	Female	17.22	-2.28	1.19	0.055	-0.06	0.00	0.01	0.037
Ever enrolled in a vocational program	Full Sample	28.63	-2.52	0.91	0.006*	-0.06	0.00	0.00	0.003
	Subgroup:				0.726				0.023
	Male	24.97	-2.82	1.13	0.012*	-0.07	0.00	0.00	0.001
	Female	33.86	-2.16	1.50	0.151	-0.05	0.00	0.00	0.047
Ever enrolled in education or training program	Full Sample	0.72	0.21	0.01	0.000*	0.47	0.00	0.00	0.000
	Subgroup:				0.000*				0.000
	Male	0.69	0.23	0.01	0.000*	0.50	0.00	0.00	0.000
	Female	0.76	0.17	0.01	0.000*	0.40	0.00	0.00	0.000
Ever took academic classes	Full Sample	0.57	0.24	0.01	0.000*	0.50	0.00	0.00	0.000
	Subgroup:				0.756				0.017
	Male	0.56	0.24	0.02	0.000*	0.50	0.00	0.00	0.000
	Female	0.58	0.23	0.02	0.000*	0.48	0.00	0.00	0.000
Ever took vocational training	Full Sample	0.28	0.46	0.01	0.000*	1.05	0.00	0.00	0.000
	Subgroup:				0.089				0.019
	Male	0.26	0.47	0.02	0.000*	1.11	0.00	0.02	0.000
	Female	0.32	0.43	0.02	0.000*	0.95	0.00	0.00	0.000
Hours per week ever in academic classes	Full Sample	2.50	0.58	0.11	0.000*	0.15	0.00	0.00	0.000
	Subgroup:				0.052				0.014
	Male	2.54	0.40	0.14	0.004*	0.10	0.00	0.00	0.001
	Female	2.42	0.85	0.18	0.000*	0.22	0.00	0.00	0.000

Outcome	Sample	Control Mean	Impact	SE	p-value	Effect Size	Absolute Value of Difference Compared to Model-Based Specification: Dif. Effect Size	Dif. SE (Effect Size)	Dif. p-value
Hours per week ever in education or training program	Full Sample	4.12	3.45	0.11	0.000*	0.65	0.00	0.00	0.000
	Subgroup:				0.684				0.011
	Male	3.87	3.49	0.15	0.000*	0.67	0.00	0.00	0.000
	Female	4.48	3.40	0.18	0.000*	0.62	0.00	0.00	0.000
Hours per week ever in vocational training	Full Sample	0.94	2.20	0.09	0.000*	0.92	0.00	0.00	0.000
	Subgroup:				0.633				0.003
	Male	0.82	2.23	0.11	0.000*	1.01	0.00	0.00	0.000
	Female	1.12	2.15	0.14	0.000*	0.82	0.00	0.00	0.000
Hours worked, Q16	Full Sample	26.40	1.46	0.46	0.001*	0.06	0.00	0.00	0.000
	Subgroup:				0.934				0.003
	Male	28.96	1.46	0.61	0.017*	0.06	0.00	0.00	0.008
	Female	22.65	1.54	0.66	0.021*	0.07	0.00	0.00	0.001
Received a college degree	Full Sample	1.54	-0.21	0.24	0.393	-0.02	0.00	0.00	0.038
	Subgroup:				0.275				0.064
	Male	1.08	0.02	0.27	0.942	0.00	0.00	0.00	0.005
	Female	2.21	-0.55	0.45	0.219	-0.04	0.00	0.01	0.102
Received a vocational degree	Full Sample	15.19	22.27	0.81	0.000*	0.62	0.00	0.00	0.000
	Subgroup:				0.251				0.013
	Male	13.21	21.45	1.02	0.000*	0.63	0.00	0.00	0.000
	Female	18.10	23.37	1.33	0.000*	0.61	0.00	0.00	0.000
Served time in jail for convictions	Full Sample	17.95	-2.11	0.75	0.005*	-0.05	0.00	0.00	0.002
	Subgroup:				0.074				0.010
	Male	25.96	-3.02	1.11	0.007*	-0.07	0.00	0.01	0.006
	Female	6.24	-0.61	0.77	0.427	-0.03	0.00	0.01	0.143

Outcome	Sample	Control Mean	Impact	SE	p-value	Effect Size	Absolute Value of Difference Compared to Model-Based Specification		
							Dif. Effect Size	Dif. SE (Effect Size)	Dif. p-value
Weeks employed, Q16	Full Sample	59.02	2.79	0.90	0.002*	0.06	0.00	0.00	0.001
	Subgroup:				0.958				0.003
	Male	61.71	2.78	1.16	0.016*	0.06	0.00	0.00	0.004
	Female	55.10	2.87	1.43	0.045*	0.06	0.00	0.00	0.014
Weeks in jail for convictions	Full Sample	6.57	-0.58	0.47	0.222	-0.02	0.00	0.00	0.030
	Subgroup:				0.127				0.041
	Male	10.59	-1.04	0.77	0.179	-0.03	0.00	0.01	0.113
	Female	0.71	0.18	0.20	0.371	0.03	0.00	0.08	0.420

Note: The p-value on the subgroup rows tests for differences in impacts across subgroup categories. Effect sizes are calculated by dividing the impact estimate by the standard deviation of the outcome in the control group.
* Difference is statistically significant at the 0.05 level, two-tailed test.

Table II.A.3. Job Corps: Evaluation Impact Estimates (Design-Based, *RCT-YES* Default Assumptions)

Outcome	Sample	Control Mean	Impact	SE	p-value	Effect Size	Absolute Value of Difference Compared to Model-Based Specification: Dif. Effect Size	Dif. SE (Effect Size)	Dif. p-value
Amount of benefits received	Full Sample	4155.73	-459.76	140.33	0.001*	-0.06	0.00	0.00	0.000
	Subgroup:				0.874				0.019
	Male	2075.56	-461.85	118.49	0.000*	-0.10	0.00	0.01	0.005
	Female	7068.16	-508.59	270.15	0.060	-0.06	0.00	0.01	0.051
Convicted, pled guilty, or adjudged delinquent	Full Sample	25.17	-3.08	0.85	0.000*	-0.07	0.00	0.00	0.000
	Subgroup:				0.123				0.003
	Male	34.89	-3.97	1.21	0.001*	-0.08	0.00	0.00	0.001
	Female	10.98	-1.56	0.98	0.113	-0.05	0.00	0.01	0.085
Earnings per week, Q16	Full Sample	199.35	18.11	4.25	0.000*	0.09	0.00	0.00	0.000
	Subgroup:				0.463				0.001
	Male	225.55	20.84	5.95	0.000*	0.10	0.00	0.00	0.000
	Female	161.13	14.84	5.60	0.008*	0.08	0.00	0.00	0.010
Ever arrested or charged with a delinquency or criminal complaint	Full Sample	32.56	-3.74	0.92	0.000*	-0.08	0.00	0.00	0.000
	Subgroup:				0.037*				0.002
	Male	43.52	-5.06	1.26	0.000*	-0.10	0.00	0.00	0.000
	Female	16.47	-1.48	1.17	0.204	-0.04	0.00	0.00	0.053
Ever employed, Q16	Full Sample	0.69	0.02	0.01	0.010*	0.04	0.00	0.00	0.004
	Subgroup:				0.526				0.030
	Male	0.71	0.02	0.01	0.100	0.04	0.00	0.00	0.010
	Female	0.65	0.03	0.02	0.037*	0.06	0.00	0.02	0.015
Ever enrolled in a four-year college program	Full Sample	3.37	-0.09	0.37	0.796	0.00	0.00	0.00	0.012
	Subgroup:				0.876				0.015
	Male	2.50	-0.05	0.41	0.901	0.00	0.00	0.00	0.009
	Female	4.62	-0.17	0.67	0.796	-0.01	0.00	0.01	0.050

Outcome	Sample	Control Mean	Impact	SE	p-value	Effect Size	Absolute Value of Difference Compared to Model-Based Specification		
							Dif. Effect Size	Dif. SE (Effect Size)	Dif. p-value
Ever enrolled in a two-year college program	Full Sample	12.31	-0.82	0.66	0.218	-0.02	0.00	0.00	0.028
	Subgroup:				0.081				0.030
	Male	8.89	0.18	0.75	0.815	0.01	0.00	0.00	0.013
	Female	17.22	-2.28	1.19	0.055	-0.06	0.00	0.01	0.037
Ever enrolled in a vocational program	Full Sample	28.63	-2.52	0.91	0.006*	-0.06	0.00	0.00	0.003
	Subgroup:				0.726				0.023
	Male	24.97	-2.82	1.13	0.012*	-0.07	0.00	0.00	0.001
	Female	33.86	-2.16	1.50	0.151	-0.05	0.00	0.00	0.047
Ever enrolled in education or training program	Full Sample	0.72	0.21	0.01	0.000*	0.47	0.00	0.00	0.000
	Subgroup:				0.000*				0.000
	Male	0.69	0.23	0.01	0.000*	0.50	0.00	0.00	0.000
	Female	0.76	0.17	0.01	0.000*	0.40	0.00	0.00	0.000
Ever took academic classes	Full Sample	0.57	0.24	0.01	0.000*	0.50	0.00	0.00	0.000
	Subgroup:				0.756				0.017
	Male	0.56	0.24	0.02	0.000*	0.50	0.00	0.00	0.000
	Female	0.58	0.23	0.02	0.000*	0.48	0.00	0.00	0.000
Ever took vocational training	Full Sample	0.28	0.46	0.01	0.000*	1.05	0.00	0.00	0.000
	Subgroup:				0.089				0.019
	Male	0.26	0.47	0.02	0.000*	1.11	0.00	0.02	0.000
	Female	0.32	0.43	0.02	0.000*	0.95	0.00	0.00	0.000
Hours per week ever in academic classes	Full Sample	2.50	0.58	0.11	0.000*	0.15	0.00	0.00	0.000
	Subgroup:				0.052				0.014
	Male	2.54	0.40	0.14	0.004*	0.10	0.00	0.00	0.001
	Female	2.42	0.85	0.18	0.000*	0.22	0.00	0.00	0.000

Outcome	Sample	Control Mean	Impact	SE	p-value	Effect Size	Absolute Value of Difference Compared to Model-Based Specification		
							Dif. Effect Size	Dif. SE (Effect Size)	Dif. p-value
Hours per week ever in education or training program	Full Sample	4.12	3.45	0.11	0.000*	0.65	0.00	0.00	0.000
	Subgroup:				0.684				0.011
	Male	3.87	3.49	0.15	0.000*	0.67	0.00	0.00	0.000
	Female	4.48	3.40	0.18	0.000*	0.62	0.00	0.00	0.000
Hours per week ever in vocational training	Full Sample	0.94	2.20	0.09	0.000*	0.92	0.00	0.00	0.000
	Subgroup:				0.633				0.003
	Male	0.82	2.23	0.11	0.000*	1.01	0.00	0.00	0.000
	Female	1.12	2.15	0.14	0.000*	0.82	0.00	0.00	0.000
Hours worked, Q16	Full Sample	26.40	1.46	0.46	0.001*	0.06	0.00	0.00	0.000
	Subgroup:				0.934				0.003
	Male	28.96	1.46	0.61	0.017*	0.06	0.00	0.00	0.008
	Female	22.65	1.54	0.66	0.021*	0.07	0.00	0.00	0.001
Received a college degree	Full Sample	1.54	-0.21	0.24	0.393	-0.02	0.00	0.00	0.038
	Subgroup:				0.275				0.064
	Male	1.08	0.02	0.27	0.942	0.00	0.00	0.00	0.005
	Female	2.21	-0.55	0.45	0.219	-0.04	0.00	0.01	0.102
Received a vocational degree	Full Sample	15.19	22.27	0.81	0.000*	0.62	0.00	0.00	0.000
	Subgroup:				0.251				0.013
	Male	13.21	21.45	1.02	0.000*	0.63	0.00	0.00	0.000
	Female	18.10	23.37	1.33	0.000*	0.61	0.00	0.00	0.000
Served time in jail for convictions	Full Sample	17.95	-2.11	0.75	0.005*	-0.05	0.00	0.00	0.002
	Subgroup:				0.074				0.010
	Male	25.96	-3.02	1.11	0.007*	-0.07	0.00	0.01	0.006
	Female	6.24	-0.61	0.77	0.427	-0.03	0.00	0.01	0.143

Outcome	Sample	Control Mean	Impact	SE	p-value	Effect Size	Absolute Value of Difference Compared to Model-Based Specification: Dif. Effect Size	Dif. SE (Effect Size)	Dif. p-value
Weeks employed, Q16	Full Sample	59.02	2.79	0.90	0.002*	0.06	0.00	0.00	0.001
	Subgroup:				0.958				0.003
	Male	61.71	2.78	1.16	0.016*	0.06	0.00	0.00	0.004
	Female	55.10	2.87	1.43	0.045*	0.06	0.00	0.00	0.014
Weeks in jail for convictions	Full Sample	6.57	-0.58	0.47	0.222	-0.02	0.00	0.00	0.030
	Subgroup:				0.127				0.041
	Male	10.59	-1.04	0.77	0.179	-0.03	0.00	0.01	0.113
	Female	0.71	0.18	0.20	0.371	0.03	0.00	0.08	0.420

Note: The p-value on the subgroup rows tests for differences in impacts across subgroup categories. Effect sizes are calculated by dividing the impact estimate by the standard deviation of the outcome in the control group.
* Difference is statistically significant at the 0.05 level, two-tailed test.

Table II.A.4. Jobs Corps: Additional Information (Individual-Level Randomization)

		Standard Deviation of Outcomes			Standard Deviation of Outcome Residuals		
Outcome	Sample	Full	Treatment Group	Control Group	Full	Treatment Group	Control Group
Amount of benefits received	Full Sample	6998.75	6797.16	7188.78	6994.98	6797.16	7188.78
	Subgroup:						
	Male	4352.96	3994.85	4671.92	4346.83	3994.85	4671.92
	Female	8736.66	8581.54	8886.13	8732.96	8581.54	8886.13
Convicted, pled guilty, or adjudged delinquent	Full Sample	42.48	41.49	43.40	42.45	41.49	43.40
	Subgroup:						
	Male	46.99	46.22	47.67	46.95	46.22	47.67
	Female	30.26	29.22	31.27	30.25	29.22	31.27
Earnings per week, Q16	Full Sample	212.58	221.29	203.07	212.39	221.29	203.07
	Subgroup:						
	Male	228.00	239.77	215.14	227.76	239.77	215.14
	Female	180.80	183.96	177.27	180.65	183.96	177.27
Ever arrested or charged with a delinquency or criminal complaint	Full Sample	46.12	45.30	46.86	46.08	45.30	46.86
	Subgroup:						
	Male	49.19	48.66	49.59	49.12	48.66	49.59
	Female	36.41	35.70	37.10	36.40	35.70	37.10
Ever employed, Q16	Full Sample	0.45	0.45	0.46	0.45	0.45	0.46
	Subgroup:						
	Male	0.44	0.44	0.44	0.44	0.44	0.44
	Female	0.47	0.46	0.47	0.47	0.46	0.47
Ever enrolled in a four-year college program	Full Sample	17.92	17.80	18.04	17.92	17.80	18.04
	Subgroup:						
	Male	15.53	15.45	15.61	15.53	15.45	15.61
	Female	20.80	20.61	20.99	20.80	20.61	20.99

Outcome	Sample	Standard Deviation of Outcomes			Standard Deviation of Outcome Residuals		
		Full	Treatment Group	Control Group	Full	Treatment Group	Control Group
Ever enrolled in a two-year college program	Full Sample	32.39	31.90	32.86	32.38	31.90	32.86
	Subgroup:						
	Male	28.58	28.71	28.46	28.58	28.71	28.46
	Female	36.73	35.65	37.77	36.71	35.65	37.77
Ever enrolled in a vocational program	Full Sample	44.59	43.93	45.21	44.57	43.93	45.21
	Subgroup:						
	Male	42.44	41.53	43.29	42.42	41.53	43.29
	Female	46.95	46.54	47.34	46.93	46.54	47.34
Ever enrolled in education or training program	Full Sample	0.38	0.26	0.44	0.36	0.26	0.44
	Subgroup:						
	Male	0.39	0.27	0.46	0.37	0.27	0.46
	Female	0.36	0.26	0.42	0.35	0.26	0.42
Ever took academic classes	Full Sample	0.46	0.39	0.48	0.44	0.39	0.48
	Subgroup:						
	Male	0.46	0.40	0.48	0.44	0.40	0.48
	Female	0.45	0.39	0.48	0.44	0.39	0.48
Ever took vocational training	Full Sample	0.49	0.44	0.44	0.44	0.44	0.44
	Subgroup:						
	Male	0.49	0.44	0.42	0.43	0.44	0.42
	Female	0.49	0.44	0.45	0.44	0.44	0.45
Hours per week ever in academic classes	Full Sample	3.90	3.89	3.88	3.89	3.89	3.88
	Subgroup:						
	Male	3.84	3.80	3.87	3.84	3.80	3.87
	Female	3.98	4.03	3.89	3.96	4.03	3.89
Hours per week ever in education or training program	Full Sample	6.07	6.29	5.31	5.82	6.29	5.31
	Subgroup:						
	Male	5.97	6.19	5.19	5.71	6.19	5.19
	Female	6.20	6.43	5.45	5.97	6.43	5.45

Outcome	Sample	Standard Deviation of Outcomes			Standard Deviation of Outcome Residuals		
		Full	Treatment Group	Control Group	Full	Treatment Group	Control Group
Hours per week ever in vocational training	Full Sample	3.43	3.95	2.38	3.25	3.95	2.38
	Subgroup:						
	Male	3.36	3.94	2.21	3.17	3.94	2.21
	Female	3.52	3.97	2.62	3.36	3.97	2.62
Hours worked, Q16	Full Sample	22.89	22.97	22.79	22.88	22.97	22.79
	Subgroup:						
	Male	23.69	23.68	23.68	23.68	23.68	23.68
	Female	21.13	21.38	20.86	21.12	21.38	20.86
Received a college degree	Full Sample	11.88	11.45	12.30	11.88	11.45	12.30
	Subgroup:						
	Male	10.37	10.42	10.33	10.37	10.42	10.33
	Female	13.77	12.78	14.71	13.77	12.78	14.71
Received a vocational degree	Full Sample	44.05	48.40	35.89	42.62	48.40	35.89
	Subgroup:						
	Male	42.64	47.59	33.86	41.27	47.59	33.86
	Female	45.78	49.28	38.51	44.26	49.28	38.51
Served time in jail for convictions	Full Sample	37.47	36.51	38.38	37.45	36.51	38.38
	Subgroup:						
	Male	42.99	42.05	43.85	42.96	42.05	43.85
	Female	23.63	23.06	24.20	23.63	23.06	24.20
Weeks employed, Q16	Full Sample	44.48	44.54	44.38	44.46	44.54	44.38
	Subgroup:						
	Male	44.00	44.06	43.91	43.98	44.06	43.91
	Female	44.88	44.95	44.78	44.86	44.95	44.78

Outcome	Sample	Standard Deviation of Outcomes			Standard Deviation of Outcome Residuals		
		Full	Treatment Group	Control Group	Full	Treatment Group	Control Group
Weeks in jail for convictions	Full Sample	23.41	23.06	23.75	23.41	23.06	23.75
	Subgroup:						
	Male	29.28	28.73	29.81	29.27	28.73	29.81
	Female	7.04	8.17	5.67	7.04	8.17	5.67

Note: The residuals are calculated using all covariates included in the model. All calculations include study weights, if applicable.

II.B. Teen Options to Prevent Pregnancy

Original study. Smith et al. (2015).

Sponsor agency. Office of Adolescent Health, U.S. Department of Health & Human Services.

Description of intervention. The original report presents interim findings from an impact evaluation of the Teen Options to Prevent Pregnancy (TOPP) program. TOPP is an 18-month clinic-based intervention that aims to reduce rapid repeat adolescent pregnancies by using three primary components: (1) telephone-based care coordination, (2) facilitated access to contraceptive services, and (3) risk assessment and referrals by a social worker.

Randomization design. Non-clustered, blocked. In the original study, 598 women were randomized to either a treatment or control group. The study used a permuted block design in which blocks were formed based on recruitment location and age group (above or below age 18).

Model-based method. Ordinary least squares.

We use OLS to estimate the following equation:

$$(14)\quad Y_i = \alpha + \beta X_i + \delta T_i + \varepsilon_i,$$

where Y_i is the outcome of interest for person i, T_i is a treatment indicator, X_i is a vector of covariates, and ε_i is an error term. We assume that ε_i is independent between observations. Note that we do not include block fixed effects, because we apply a non-clustered, non-blocked approach for the design-based method.

Design-based specification: RCT-YES model. Non-clustered, non-blocked. A number of blocks have a single observation, so we do not account for the blocked design in our estimation. These blocks would have been excluded from the analysis because of *RCT-YES*'s restriction that each block must have at least one member of the treatment group and one member of the control group.

Design-based specification: RCT-YES parameter (matched settings). Finite population (ATE).

Design-based specification: RCT-YES parameter (default settings). Finite population (ATE).

Covariates. In our analysis, we use the following covariates in each regression model: age, race, participants' self-reported use of a modern, highly effective method of birth control in the three months before they became pregnant, baseline exposure to reproductive health information, and participants' baseline perceptions of their need for birth control. A baseline measure of the outcome variable is included when available. This is consistent with the approach in the original report.

Weights. No. The original study did not use weights, so we do not use them in the model-based method. For the matched settings analyses, we specify equal weights for each observation in *RCT-YES*. For the default settings analyses, we do not specify weights, so the default *RCT-YES* weight scheme is used.

Multiple hypothesis correction. Yes. The original study adjusted for multiple hypotheses by using the Hothorn et al. (2008) procedure within outcome domains. We specify the same outcome domains, but use the Benjamini- Hochberg (1995) procedure instead, as is also done in *RCT-YES*.

Main differences from original analyses. Our model-based methods differ in several ways from the original analysis:

- The original study applied a logistic regression mode for binary outcomes and reported mean marginal effects between the treatment and control groups. For binary outcomes, we estimate a linear specification to be consistent with other studies.

- The original study controlled for groupings of age and location (these groupings were used to define blocks). We do not control for these groupings in the model-based method, because we are treating the study as a non-blocked, non-clustered design. The original study also allowed for clustering of errors at the block level. We do not allow for clustering, because we treat the study as a non-blocked, non-clustered design.

- The original analysis included indicators for missing values. For our model-based methods, we apply the same imputation strategy as *RCT-YES* and do not include any flags for missing variables.

Outcomes. We estimate the model for the following outcomes:

Used LARC method in past 3 months: Participant reported using a long-acting reversible contraception (LARC) method (IUD or implant) in the past 3 months.

Used hormonal method or IUD in past 3 months: Participant reported using one of the following methods in the past 3 months: birth control pills, shot, patch, ring, IUD, or implant.

Had unprotected sex in past 3 months: Participant had sexual intercourse without using an effective birth control method in the past 3 months.

Had sex in past 3 months: Participant reported having sexual intercourse in the past 3 months.

Had unprotected sex without a condom in past 3 months: Participant reported having sexual intercourse without a condom in the past 3 months.

Number of sexual partners in past 3 months: Number of reported sexual partners in the past 3 months.

Received info on relationships in past 12 months: Participant received any information in the past 12 months on relationships, dating, marriage, or family life.

Received info on birth control methods in past 12 months: Participant received any information in the past 12 months on methods of birth control.

Received info on where to get birth control in past 12 months: Participant received any information in the past 12 months on where to get birth control.

Received info on abstinence in past 12 months: Participant received any information in the past 12 months on abstaining from sex.

Received info on STIs in past 12 months: Participant received any information in the past 12 months on STIs.

Received info on talking to partner about sex in past 12 months: Participant received any information in the past 12 months on how to talk to his or her partner about whether to have sex or use birth control.

Received info on saying no to sex in past 12 months: Participant received any information in the past 12 months on how to say no to sex.

Received info on sexual health from professional at facility in past 12 months: Participant received information on sexual and reproductive health topics from a doctor or nurse at a health facility in the past 12 months.

Received info on sexual health from professional at home in past 12 months: Participant received information on sexual and reproductive health topics from a health provider during a home visit in the past 12 months.

Knowledge of efficacy of condoms in preventing pregnancy: Respondent answered a single knowledge question correctly on the efficacy of condoms in preventing pregnancy.

Knowledge of efficacy of condoms in preventing STIs: Respondent answered a single knowledge question correctly on the efficacy of condoms in preventing STIs.

Knowledge of efficacy of birth control pills in preventing STIs: Respondent answered a single knowledge question correctly on the efficacy of birth control pills in preventing STIs.

Perceived access to condoms: Perceived access to condoms on a scale of 1 to 5, with higher values indicating greater perceived access.

Perceived access to birth control other than condoms: Perceived access to birth control other than condoms on a scale of 1 to 5, with higher values indicating greater perceived access.

Perceived trustworthiness of birth control providers: Perceived trustworthiness of birth control providers on a scale of 1 to 5, with higher values indicating greater perceived trustworthiness.

Perceived ease of using birth control: Perceived ease of using birth control on a scale of 1 to 5, with higher values indicating greater perceived ease of use.

Perceived need for condoms: Perceived need for condoms on a scale of 1 to 5, with higher values indicating greater perceived need.

Perceived need for birth control other than condoms: Perceived need for birth control other than condoms on a scale of 1 to 5, with higher values indicating greater perceived need.

Intention to avoid pregnancy in next year: Participant reported "trying to avoid getting pregnant" in the next year.

Received birth control from professional: Participant reported receiving birth control from a doctor or nurse in the past 3 months.

Table II.B.1. Teen Options to Prevent Pregnancy: Evaluation Impact Estimates (Model-Based)

Outcome	Sample	Control Mean	Impact	SE	p-value	Effect Size
Had sex in past 3 months	Full Sample	85.25	-1.50	3.26	0.644	-0.04
	Subgroup:				0.854	
	Not Overage	82.03	-2.02	4.58	0.659	-0.05
	Overage	88.79	-0.82	4.63	0.859	-0.03
Had unprotected sex in past 3 months	Full Sample	22.31	-10.43	3.39	0.002*^	-0.25
	Subgroup:				0.400	
	Not Overage	16.54	-7.55	4.76	0.113	-0.20
	Overage	28.70	-13.24	4.82	0.006*	-0.29
Had unprotected sex without a condom in past 3 months	Full Sample	52.48	-3.36	4.37	0.442	-0.07
	Subgroup:				0.601	
	Not Overage	43.31	-5.68	6.14	0.355	-0.11
	Overage	62.61	-1.13	6.21	0.856	-0.02
Intention to avoid pregnancy in next year	Full Sample	65.83	6.39	4.16	0.125	0.13
	Subgroup:				0.416	
	Not Overage	65.87	2.80	5.81	0.630	0.06
	Overage	65.79	9.54	5.89	0.106	0.20
Knowledge of efficacy of birth control in preventing pregnancy	Full Sample	52.48	-3.46	4.33	0.426	-0.07
	Subgroup:				0.739	
	Not Overage	48.44	-4.84	6.08	0.427	-0.10
	Overage	57.02	-1.96	6.18	0.752	-0.04
Knowledge of efficacy of birth control pills in preventing STIs	Full Sample	63.49	0.16	3.97	0.968	0.00
	Subgroup:				0.711	
	Not Overage	60.94	-1.17	5.56	0.834	-0.02
	Overage	66.37	1.77	5.67	0.755	0.04
Knowledge of efficacy of condoms in preventing STIs	Full Sample	28.63	-2.73	3.99	0.494	-0.06
	Subgroup:				0.505	
	Not Overage	21.88	-0.29	5.58	0.958	-0.01
	Overage	36.28	-5.60	5.69	0.326	-0.12
Knowledge of efficacy of condoms in preventing pregnancy	Full Sample	53.94	-1.56	4.42	0.724	-0.03
	Subgroup:				0.093	
	Not Overage	49.22	5.88	6.16	0.340	0.12
	Overage	59.29	-8.89	6.27	0.157	-0.18

Outcome	Sample	Control Mean	Impact	SE	p-value	Effect Size
Number of sexual partners in past 3 months	Full Sample	1.03	-0.09	0.11	0.448	-0.06
	Subgroup:				0.239	
	Not Overage	1.06	-0.22	0.16	0.172	-0.10
	Overage	1.00	0.05	0.16	0.761	0.10
Perceived access to birth control other than condoms	Full Sample	4.35	-0.06	0.07	0.342	-0.07
	Subgroup:				0.134	
	Not Overage	4.35	-0.16	0.09	0.089	-0.20
	Overage	4.34	0.04	0.10	0.672	0.05
Perceived access to condoms	Full Sample	4.50	-0.05	0.07	0.436	-0.07
	Subgroup:				0.161	
	Not Overage	4.38	0.04	0.09	0.682	0.05
	Overage	4.64	-0.15	0.09	0.119	-0.24
Perceived ease of using birth control	Full Sample	3.49	0.00	0.08	0.961	0.00
	Subgroup:				0.312	
	Not Overage	3.41	0.08	0.11	0.463	0.08
	Overage	3.57	-0.08	0.12	0.489	-0.08
Perceived need for birth control other than condoms	Full Sample	4.31	0.03	0.06	0.599	0.04
	Subgroup:				0.328	
	Not Overage	4.31	0.10	0.09	0.277	0.13
	Overage	4.31	-0.03	0.09	0.765	-0.04
Perceived need for condoms	Full Sample	3.59	0.01	0.06	0.889	0.01
	Subgroup:				0.068	
	Not Overage	3.50	0.11	0.08	0.173	0.15
	Overage	3.68	-0.10	0.08	0.223	-0.17
Perceived trustworthiness of birth control providers	Full Sample	4.04	0.07	0.08	0.383	0.08
	Subgroup:				0.580	
	Not Overage	4.08	0.03	0.11	0.806	0.04
	Overage	3.99	0.11	0.11	0.313	0.12
Received birth control from professional	Full Sample	68.31	10.75	4.00	0.007*^	0.23
	Subgroup:				0.308	
	Not Overage	74.22	6.75	5.62	0.230	0.15
	Overage	61.74	14.90	5.69	0.009*	0.31
Received info on STIs in past 12 months	Full Sample	69.96	8.62	3.93	0.029*	0.19
	Subgroup:				0.331	
	Not Overage	74.80	4.84	5.55	0.383	0.11
	Overage	64.66	12.49	5.59	0.026*	0.26

Outcome	Sample	Control Mean	Impact	SE	p-value	Effect Size
Received info on abstinence in past 12 months	Full Sample	36.25	15.54	4.43	0.000*^	0.32
	Subgroup:				0.858	
	Not Overage	38.89	16.21	6.20	0.009*	0.33
	Overage	33.33	14.63	6.28	0.020*	0.31
Received info on birth control methods in past 12 months	Full Sample	77.78	12.52	3.33	0.000*^	0.30
	Subgroup:				0.540	
	Not Overage	81.10	10.58	4.68	0.024*	0.27
	Overage	74.14	14.67	4.75	0.002*	0.33
Received info on relationships in past 12 months	Full Sample	34.02	7.82	4.39	0.075	0.16
	Subgroup:				0.927	
	Not Overage	35.94	7.39	6.17	0.232	0.15
	Overage	31.90	8.19	6.25	0.190	0.17
Received info on saying no to sex in past 12 months	Full Sample	74.07	4.17	3.69	0.258	0.09
	Subgroup:				0.479	
	Not Overage	78.74	1.54	5.21	0.767	0.04
	Overage	68.97	6.79	5.26	0.197	0.15
Received info on sexual health from professional at facility in past 12 months	Full Sample	83.61	3.96	3.21	0.218	0.11
	Subgroup:				0.608	
	Not Overage	85.16	2.44	4.53	0.590	0.07
	Overage	81.90	5.73	4.55	0.209	0.15
Received info on sexual health from professional at home in past 12 months	Full Sample	38.52	30.24	4.23	0.000*^	0.62
	Subgroup:				0.048*	
	Not Overage	47.66	21.92	5.93	0.000*	0.44
	Overage	28.45	38.63	5.98	0.000*	0.85
Received info on talking to partner about sex in past 12 months	Full Sample	73.66	3.06	3.82	0.422	0.07
	Subgroup:				0.035*	
	Not Overage	79.53	-4.93	5.36	0.358	-0.12
	Overage	67.24	11.11	5.40	0.040*	0.24
Received info on where to get birth control in past 12 months	Full Sample	80.58	11.15	3.14	0.000*^	0.28
	Subgroup:				0.400	
	Not Overage	84.13	8.52	4.42	0.055	0.23
	Overage	76.72	13.81	4.47	0.002*	0.33
Used LARC method in past 3 months	Full Sample	22.55	15.24	4.07	0.000*^	0.36
	Subgroup:				0.975	
	Not Overage	22.40	15.30	5.71	0.008*	0.37
	Overage	22.73	15.05	5.81	0.010*	0.36

Outcome	Sample	Control Mean	Impact	SE	p-value	Effect Size
Used hormonal method or IUD in past 3 months	Full Sample	69.49	8.35	4.04	0.039*^	0.18
	Subgroup:				0.768	
	Not Overage	72.80	9.51	5.67	0.094	0.21
	Overage	65.77	7.13	5.77	0.217	0.15

Note: The p-value on the subgroup rows tests for differences in impacts across subgroup categories. Effect sizes are calculated by dividing the impact estimate by the standard deviation of the outcome in the control group.
* Difference is statistically significant at the 0.05 level, two-tailed test.
^ Difference remains statistically significant at the 0.05 level, two-tailed test, after applying the Benjamini-Hochberg correction for multiple hypothesis testing across all full sample analyses in the same domain.

Table II.B.2. Teen Options to Prevent Pregnancy: Evaluation Impact Estimates (Design-Based, Matching Assumptions)

Outcome	Sample	Control Mean	Impact	SE	p-value	Effect Size	Absolute Value of Difference Compared to Model-Based Specification: Dif. Effect Size	Dif. SE (Effect Size)	Dif. p-value
Had sex in past 3 months	Full Sample	85.25	-1.50	3.22	0.640	-0.04	0.00	0.00	0.004
	Subgroup:				0.802				0.052
	Not Overage	82.03	-2.18	5.00	0.663	-0.06	0.00	0.01	0.004
	Overage	88.79	-0.55	4.13	0.893	-0.02	0.01	0.02	0.034
Had unprotected sex in past 3 months	Full Sample	22.31	-10.43	3.35	0.002*^	-0.25	0.00	0.00	0.000
	Subgroup:				0.445				0.045
	Not Overage	16.54	-7.80	4.32	0.072	-0.21	0.01	0.01	0.041
	Overage	28.70	-12.98	5.22	0.014*	-0.29	0.01	0.01	0.008
Had unprotected sex without a condom in past 3 months	Full Sample	52.48	-3.36	4.31	0.436	-0.07	0.00	0.00	0.006
	Subgroup:				0.542				0.059
	Not Overage	43.31	-6.11	6.22	0.327	-0.12	0.01	0.00	0.028
	Overage	62.61	-0.79	6.10	0.897	-0.02	0.01	0.00	0.041
Intention to avoid pregnancy in next year	Full Sample	65.83	6.39	4.12	0.121	0.13	0.00	0.00	0.004
	Subgroup:				0.447				0.031
	Not Overage	65.87	2.99	5.92	0.614	0.06	0.00	0.00	0.016
	Overage	65.79	9.27	5.75	0.108	0.19	0.01	0.00	0.002
Knowledge of efficacy of birth control in preventing pregnancy	Full Sample	52.48	-3.46	4.28	0.420	-0.07	0.00	0.00	0.006
	Subgroup:				0.712				0.027
	Not Overage	48.44	-4.88	6.02	0.419	-0.10	0.00	0.00	0.008
	Overage	57.02	-1.70	6.16	0.784	-0.03	0.01	0.00	0.032

Outcome	Sample	Control Mean	Impact	SE	p-value	Effect Size	Absolute Value of Difference Compared to Model-Based Specification: Dif. Effect Size	Dif. SE (Effect Size)	Dif. p-value
Knowledge of efficacy of birth control pills in preventing STIs	Full Sample	63.49	0.16	3.92	0.967	0.00	0.00	0.00	0.001
	Subgroup:				0.670				0.041
	Not Overage	60.94	-1.32	5.76	0.819	-0.03	0.00	0.00	0.015
	Overage	66.37	2.05	5.39	0.704	0.04	0.01	0.01	0.051
Knowledge of efficacy of condoms in preventing STIs	Full Sample	28.63	-2.73	3.94	0.489	-0.06	0.00	0.00	0.005
	Subgroup:				0.511				0.006
	Not Overage	21.88	-0.33	5.17	0.950	-0.01	0.00	0.01	0.008
	Overage	36.28	-5.53	5.99	0.357	-0.11	0.00	0.01	0.031
Knowledge of efficacy of condoms in preventing pregnancy	Full Sample	53.94	-1.56	4.36	0.720	-0.03	0.00	0.00	0.004
	Subgroup:				0.104				0.011
	Not Overage	49.22	5.70	5.95	0.339	0.11	0.00	0.00	0.001
	Overage	59.29	-8.58	6.42	0.183	-0.17	0.01	0.00	0.026
Number of sexual partners in past 3 months	Full Sample	1.03	-0.09	0.10	0.405	-0.06	0.00	0.01	0.043
	Subgroup:				0.159				0.080
	Not Overage	1.06	-0.22	0.17	0.204	-0.10	0.00	0.00	0.032
	Overage	1.00	0.05	0.08	0.544	0.10	0.00	0.16	0.217
Perceived access to birth control other than condoms	Full Sample	4.35	-0.06	0.07	0.336	-0.07	0.00	0.00	0.006
	Subgroup:				0.121				0.013
	Not Overage	4.35	-0.16	0.10	0.095	-0.20	0.00	0.01	0.006
	Overage	4.34	0.05	0.09	0.619	0.06	0.01	0.01	0.053

Outcome	Sample	Control Mean	Impact	SE	p-value	Effect Size	Absolute Value of Difference Compared to Model-Based Specification		
							Dif. Effect Size	Dif. SE (Effect Size)	Dif. p-value
Perceived access to condoms	Full Sample	4.50	-0.05	0.07	0.431	-0.07	0.00	0.00	0.005
	Subgroup:				0.127				0.034
	Not Overage	4.38	0.05	0.10	0.632	0.06	0.01	0.01	0.050
	Overage	4.64	-0.15	0.09	0.079	-0.24	0.00	0.00	0.040
Perceived ease of using birth control	Full Sample	3.49	0.00	0.08	0.961	0.00	0.00	0.00	0.000
	Subgroup:				0.307				0.005
	Not Overage	3.41	0.08	0.11	0.453	0.08	0.00	0.00	0.010
	Overage	3.57	-0.08	0.12	0.487	-0.08	0.00	0.00	0.002
Perceived need for birth control other than condoms	Full Sample	4.31	0.03	0.06	0.595	0.04	0.00	0.00	0.004
	Subgroup:				0.332				0.004
	Not Overage	4.31	0.10	0.09	0.261	0.13	0.00	0.00	0.016
	Overage	4.31	-0.03	0.09	0.782	-0.04	0.00	0.00	0.017
Perceived need for condoms	Full Sample	3.59	0.01	0.06	0.888	0.01	0.00	0.00	0.001
	Subgroup:				0.066				0.002
	Not Overage	3.50	0.11	0.09	0.204	0.15	0.00	0.01	0.031
	Overage	3.68	-0.10	0.07	0.181	-0.17	0.00	0.02	0.042
Perceived trustworthiness of birth control providers	Full Sample	4.04	0.07	0.08	0.376	0.08	0.00	0.00	0.007
	Subgroup:				0.584				0.004
	Not Overage	4.08	0.03	0.10	0.787	0.04	0.00	0.01	0.019
	Overage	3.99	0.11	0.12	0.343	0.12	0.00	0.01	0.030

Outcome	Sample	Control Mean	Impact	SE	p-value	Effect Size	Absolute Value of Difference Compared to Model-Based Specification		
							Dif. Effect Size	Dif. SE (Effect Size)	Dif. p-value
Received birth control from professional	Full Sample	68.31	10.75	3.95	0.007*^	0.23	0.00	0.00	0.000
	Subgroup:				0.323				0.015
	Not Overage	74.22	6.85	5.42	0.208	0.16	0.00	0.00	0.022
	Overage	61.74	14.71	5.81	0.012*	0.30	0.00	0.00	0.003
Received info on STIs in past 12 months	Full Sample	69.96	8.62	3.89	0.027*^	0.19	0.00	0.00	0.002
	Subgroup:				0.331				0.000
	Not Overage	74.80	4.86	5.35	0.365	0.11	0.00	0.00	0.018
	Overage	64.66	12.46	5.68	0.029*	0.26	0.00	0.00	0.003
Received info on abstinence in past 12 months	Full Sample	36.25	15.54	4.35	0.000*^	0.32	0.00	0.00	0.000
	Subgroup:				0.837				0.021
	Not Overage	38.89	16.21	6.23	0.010*	0.33	0.00	0.00	0.001
	Overage	33.33	14.42	6.13	0.019*	0.30	0.00	0.00	0.001
Received info on birth control methods in past 12 months	Full Sample	77.78	12.52	3.27	0.000*^	0.30	0.00	0.00	0.000
	Subgroup:				0.531				0.009
	Not Overage	81.10	10.57	4.36	0.016*	0.27	0.00	0.01	0.008
	Overage	74.14	14.68	4.89	0.003*	0.33	0.00	0.00	0.001
Received info on relationships in past 12 months	Full Sample	34.02	7.82	4.34	0.072	0.16	0.00	0.00	0.003
	Subgroup:				0.936				0.009
	Not Overage	35.94	7.42	6.19	0.231	0.15	0.00	0.00	0.001
	Overage	31.90	8.12	6.12	0.186	0.17	0.00	0.00	0.004

Outcome	Sample	Control Mean	Impact	SE	p-value	Effect Size	Absolute Value of Difference Compared to Model-Based Specification: Dif. Effect Size	Dif. SE (Effect Size)	Dif. p-value
Received info on saying no to sex in past 12 months	Full Sample	74.07	4.17	3.65	0.253	0.09	0.00	0.00	0.005
	Subgroup:				0.477				0.002
	Not Overage	78.74	1.51	4.96	0.761	0.04	0.00	0.01	0.006
	Overage	68.97	6.72	5.40	0.214	0.14	0.00	0.00	0.017
Received info on sexual health from professional at facility in past 12 months	Full Sample	83.61	3.96	3.17	0.212	0.11	0.00	0.00	0.006
	Subgroup:				0.585				0.023
	Not Overage	85.16	2.35	4.49	0.602	0.07	0.00	0.00	0.012
	Overage	81.90	5.83	4.52	0.198	0.15	0.00	0.00	0.011
Received info on sexual health from professional at home in past 12 months	Full Sample	38.52	30.24	4.18	0.000*^	0.62	0.00	0.00	0.000
	Subgroup:				0.052				0.004
	Not Overage	47.66	22.11	6.01	0.000*	0.44	0.00	0.00	0.000
	Overage	28.45	38.35	5.79	0.000*	0.85	0.01	0.00	0.000
Received info on talking to partner about sex in past 12 months	Full Sample	73.66	3.06	3.77	0.417	0.07	0.00	0.00	0.005
	Subgroup:				0.035*				0.000
	Not Overage	79.53	-4.93	5.21	0.345	-0.12	0.00	0.00	0.013
	Overage	67.24	11.02	5.46	0.045*	0.23	0.00	0.00	0.005
Received info on where to get birth control in past 12 months	Full Sample	80.58	11.15	3.09	0.000*^	0.28	0.00	0.00	0.000
	Subgroup:				0.404				0.004
	Not Overage	84.13	8.56	4.14	0.040*	0.23	0.00	0.01	0.015
	Overage	76.72	13.73	4.62	0.003*	0.32	0.00	0.00	0.001

Outcome	Sample	Control Mean	Impact	SE	p-value	Effect Size	Absolute Value of Difference Compared to Model-Based Specification		
							Dif. Effect Size	Dif. SE (Effect Size)	Dif. p-value
Used LARC method in past 3 months	Full Sample	22.55	15.24	4.00	0.000*^	0.36	0.00	0.00	0.000
	Subgroup:				0.957				0.018
	Not Overage	22.40	15.38	5.63	0.007*	0.37	0.00	0.00	0.001
	Overage	22.73	14.94	5.69	0.009*	0.35	0.00	0.00	0.001
Used hormonal method or IUD in past 3 months	Full Sample	69.49	8.35	3.99	0.037*^	0.18	0.00	0.00	0.002
	Subgroup:				0.728				0.040
	Not Overage	72.80	9.72	5.42	0.074	0.22	0.00	0.01	0.020
	Overage	65.77	6.93	5.92	0.243	0.15	0.00	0.00	0.026

Note: The p-value on the subgroup rows tests for differences in impacts across subgroup categories. Effect sizes are calculated by dividing the impact estimate by the standard deviation of the outcome in the control group.
* Difference is statistically significant at the 0.05 level, two-tailed test.
^ Difference remains statistically significant at the 0.05 level, two-tailed test, after applying the Benjamini-Hochberg correction for multiple hypothesis testing across all full sample analyses in the same domain.

Table II.B.3. Teen Options to Prevent Pregnancy: Evaluation Impact Estimates (Design-Based, *RCT-YES* Default Assumptions)

Outcome	Sample	Control Mean	Impact	SE	p-value	Effect Size	Absolute Value of Difference Compared to Model-Based Specification: Dif. Effect Size	Dif. SE (Effect Size)	Dif. p-value
Had sex in past 3 months	Full Sample	85.25	-1.50	3.22	0.640	-0.04	0.00	0.00	0.004
	Subgroup:				0.802				0.052
	Not Overage	82.03	-2.18	5.00	0.663	-0.06	0.00	0.01	0.004
	Overage	88.79	-0.55	4.13	0.893	-0.02	0.01	0.02	0.034
Had unprotected sex in past 3 months	Full Sample	22.31	-10.43	3.35	0.002*^	-0.25	0.00	0.00	0.000
	Subgroup:				0.445				0.045
	Not Overage	16.54	-7.80	4.32	0.072	-0.21	0.01	0.01	0.041
	Overage	28.70	-12.98	5.22	0.014*	-0.29	0.01	0.01	0.008
Had unprotected sex without a condom in past 3 months	Full Sample	52.48	-3.36	4.31	0.436	-0.07	0.00	0.00	0.006
	Subgroup:				0.542				0.059
	Not Overage	43.31	-6.11	6.22	0.327	-0.12	0.01	0.00	0.028
	Overage	62.61	-0.79	6.10	0.897	-0.02	0.01	0.00	0.041
Intention to avoid pregnancy in next year	Full Sample	65.83	6.39	4.12	0.121	0.13	0.00	0.00	0.004
	Subgroup:				0.447				0.031
	Not Overage	65.87	2.99	5.92	0.614	0.06	0.00	0.00	0.016
	Overage	65.79	9.27	5.75	0.108	0.19	0.01	0.00	0.002
Knowledge of efficacy of birth control in preventing pregnancy	Full Sample	52.48	-3.46	4.28	0.420	-0.07	0.00	0.00	0.006
	Subgroup:				0.712				0.027
	Not Overage	48.44	-4.88	6.02	0.419	-0.10	0.00	0.00	0.008
	Overage	57.02	-1.70	6.16	0.784	-0.03	0.01	0.00	0.032

Outcome	Sample	Control Mean	Impact	SE	p-value	Effect Size	Absolute Value of Difference Compared to Model-Based Specification		
							Dif. Effect Size	Dif. SE (Effect Size)	Dif. p-value
Knowledge of efficacy of birth control pills in preventing STIs	Full Sample	63.49	0.16	3.92	0.967	0.00	0.00	0.00	0.001
	Subgroup:				0.670				0.041
	Not Overage	60.94	-1.32	5.76	0.819	-0.03	0.00	0.00	0.015
	Overage	66.37	2.05	5.39	0.704	0.04	0.01	0.01	0.051
Knowledge of efficacy of condoms in preventing STIs	Full Sample	28.63	-2.73	3.94	0.489	-0.06	0.00	0.00	0.005
	Subgroup:				0.511				0.006
	Not Overage	21.88	-0.33	5.17	0.950	-0.01	0.00	0.01	0.008
	Overage	36.28	-5.53	5.99	0.357	-0.11	0.00	0.01	0.031
Knowledge of efficacy of condoms in preventing pregnancy	Full Sample	53.94	-1.56	4.36	0.720	-0.03	0.00	0.00	0.004
	Subgroup:				0.104				0.011
	Not Overage	49.22	5.70	5.95	0.339	0.11	0.00	0.00	0.001
	Overage	59.29	-8.58	6.42	0.183	-0.17	0.01	0.00	0.026
Number of sexual partners in past 3 months	Full Sample	1.03	-0.09	0.10	0.405	-0.06	0.00	0.01	0.043
	Subgroup:				0.159				0.080
	Not Overage	1.06	-0.22	0.17	0.204	-0.10	0.00	0.00	0.032
	Overage	1.00	0.05	0.08	0.544	0.10	0.00	0.16	0.217
Perceived access to birth control other than condoms	Full Sample	4.35	-0.06	0.07	0.336	-0.07	0.00	0.00	0.006
	Subgroup:				0.121				0.013
	Not Overage	4.35	-0.16	0.10	0.095	-0.20	0.00	0.01	0.006
	Overage	4.34	0.05	0.09	0.619	0.06	0.01	0.01	0.053

Outcome	Sample	Control Mean	Impact	SE	p-value	Effect Size	Absolute Value of Difference Compared to Model-Based Specification		
							Dif. Effect Size	Dif. SE (Effect Size)	Dif. p-value
Perceived access to condoms	Full Sample	4.50	-0.05	0.07	0.431	-0.07	0.00	0.00	0.005
	Subgroup:				0.127				0.034
	Not Overage	4.38	0.05	0.10	0.632	0.06	0.01	0.01	0.050
	Overage	4.64	-0.15	0.09	0.079	-0.24	0.00	0.00	0.040
Perceived ease of using birth control	Full Sample	3.49	0.00	0.08	0.961	0.00	0.00	0.00	0.000
	Subgroup:				0.307				0.005
	Not Overage	3.41	0.08	0.11	0.453	0.08	0.00	0.00	0.010
	Overage	3.57	-0.08	0.12	0.487	-0.08	0.00	0.00	0.002
Perceived need for birth control other than condoms	Full Sample	4.31	0.03	0.06	0.595	0.04	0.00	0.00	0.004
	Subgroup:				0.332				0.004
	Not Overage	4.31	0.10	0.09	0.261	0.13	0.00	0.00	0.016
	Overage	4.31	-0.03	0.09	0.782	-0.04	0.00	0.00	0.017
Perceived need for condoms	Full Sample	3.59	0.01	0.06	0.888	0.01	0.00	0.00	0.001
	Subgroup:				0.066				0.002
	Not Overage	3.50	0.11	0.09	0.204	0.15	0.00	0.01	0.031
	Overage	3.68	-0.10	0.07	0.181	-0.17	0.00	0.02	0.042
Perceived trustworthiness of birth control providers	Full Sample	4.04	0.07	0.08	0.376	0.08	0.00	0.00	0.007
	Subgroup:				0.584				0.004
	Not Overage	4.08	0.03	0.10	0.787	0.04	0.00	0.01	0.019
	Overage	3.99	0.11	0.12	0.343	0.12	0.00	0.01	0.030

Outcome	Sample	Control Mean	Impact	SE	p-value	Effect Size	Absolute Value of Difference Compared to Model-Based Specification: Dif. Effect Size	Dif. SE (Effect Size)	Dif. p-value
Received birth control from professional	Full Sample	68.31	10.75	3.95	0.007*^	0.23	0.00	0.00	0.000
	Subgroup:				0.323				0.015
	Not Overage	74.22	6.85	5.42	0.208	0.16	0.00	0.00	0.022
	Overage	61.74	14.71	5.81	0.012*	0.30	0.00	0.00	0.003
Received info on STIs in past 12 months	Full Sample	69.96	8.62	3.89	0.027*^	0.19	0.00	0.00	0.002
	Subgroup:				0.331				0.000
	Not Overage	74.80	4.86	5.35	0.365	0.11	0.00	0.00	0.018
	Overage	64.66	12.46	5.68	0.029*	0.26	0.00	0.00	0.003
Received info on abstinence in past 12 months	Full Sample	36.25	15.54	4.35	0.000*^	0.32	0.00	0.00	0.000
	Subgroup:				0.837				0.021
	Not Overage	38.89	16.21	6.23	0.010*	0.33	0.00	0.00	0.001
	Overage	33.33	14.42	6.13	0.019*	0.30	0.00	0.00	0.001
Received info on birth control methods in past 12 months	Full Sample	77.78	12.52	3.27	0.000*^	0.30	0.00	0.00	0.000
	Subgroup:				0.531				0.009
	Not Overage	81.10	10.57	4.36	0.016*	0.27	0.00	0.01	0.008
	Overage	74.14	14.68	4.89	0.003*	0.33	0.00	0.00	0.001
Received info on relationships in past 12 months	Full Sample	34.02	7.82	4.34	0.072	0.16	0.00	0.00	0.003
	Subgroup:				0.936				0.009
	Not Overage	35.94	7.42	6.19	0.231	0.15	0.00	0.00	0.001
	Overage	31.90	8.12	6.12	0.186	0.17	0.00	0.00	0.004

Outcome	Sample	Control Mean	Impact	SE	p-value	Effect Size	Absolute Value of Difference Compared to Model-Based Specification		
							Dif. Effect Size	Dif. SE (Effect Size)	Dif. p-value
Received info on saying no to sex in past 12 months	Full Sample	74.07	4.17	3.65	0.253	0.09	0.00	0.00	0.005
	Subgroup:				0.477				0.002
	Not Overage	78.74	1.51	4.96	0.761	0.04	0.00	0.01	0.006
	Overage	68.97	6.72	5.40	0.214	0.14	0.00	0.00	0.017
Received info on sexual health from professional at facility in past 12 months	Full Sample	83.61	3.96	3.17	0.212	0.11	0.00	0.00	0.006
	Subgroup:				0.585				0.023
	Not Overage	85.16	2.35	4.49	0.602	0.07	0.00	0.00	0.012
	Overage	81.90	5.83	4.52	0.198	0.15	0.00	0.00	0.011
Received info on sexual health from professional at home in past 12 months	Full Sample	38.52	30.24	4.18	0.000*^	0.62	0.00	0.00	0.000
	Subgroup:				0.052				0.004
	Not Overage	47.66	22.11	6.01	0.000*	0.44	0.00	0.00	0.000
	Overage	28.45	38.35	5.79	0.000*	0.85	0.01	0.00	0.000
Received info on talking to partner about sex in past 12 months	Full Sample	73.66	3.06	3.77	0.417	0.07	0.00	0.00	0.005
	Subgroup:				0.035*				0.000
	Not Overage	79.53	-4.93	5.21	0.345	-0.12	0.00	0.00	0.013
	Overage	67.24	11.02	5.46	0.045*	0.23	0.00	0.00	0.005
Received info on where to get birth control in past 12 months	Full Sample	80.58	11.15	3.09	0.000*^	0.28	0.00	0.00	0.000
	Subgroup:				0.404				0.004
	Not Overage	84.13	8.56	4.14	0.040*	0.23	0.00	0.01	0.015
	Overage	76.72	13.73	4.62	0.003*	0.32	0.00	0.00	0.001

Outcome	Sample	Control Mean	Impact	SE	p-value	Effect Size	Absolute Value of Difference Compared to Model-Based Specification		
							Dif. Effect Size	Dif. SE (Effect Size)	Dif. p-value
Used LARC method in past 3 months	Full Sample	22.55	15.24	4.00	0.000*^	0.36	0.00	0.00	0.000
	Subgroup:				0.957				0.018
	Not Overage	22.40	15.38	5.63	0.007*	0.37	0.00	0.00	0.001
	Overage	22.73	14.94	5.69	0.009*	0.35	0.00	0.00	0.001
Used hormonal method or IUD in past 3 months	Full Sample	69.49	8.35	3.99	0.037*^	0.18	0.00	0.00	0.002
	Subgroup:				0.728				0.040
	Not Overage	72.80	9.72	5.42	0.074	0.22	0.00	0.01	0.020
	Overage	65.77	6.93	5.92	0.243	0.15	0.00	0.00	0.026

Note: The p-value on the subgroup rows tests for differences in impacts across subgroup categories. Effect sizes are calculated by dividing the impact estimate by the standard deviation of the outcome in the control group.
* Difference is statistically significant at the 0.05 level, two-tailed test.
^ Difference remains statistically significant at the 0.05 level, two-tailed test, after applying the Benjamini-Hochberg correction for multiple hypothesis testing across all full sample analyses in the same domain.

Table II.B.4. Teen Options to Prevent Pregnancy: Additional Information (Individual-Level Randomization)

		Standard Deviation of Outcomes			Standard Deviation of Outcome Residuals		
Outcome	Sample	Full	Treatment Group	Control Group	Full	Treatment Group	Control Group
Had sex in past 3 months	Full Sample	36.15	36.79	35.54	35.43	36.06	34.84
	Subgroup:						
	Not Overage	39.21	40.04	38.54	38.48	39.16	37.98
	Overage	32.43	33.20	31.68	32.03	32.84	31.25
Had unprotected sex in past 3 months	Full Sample	38.39	34.41	41.72	36.76	33.87	39.59
	Subgroup:						
	Not Overage	34.46	31.10	37.30	33.28	31.10	35.36
	Overage	41.66	37.18	45.43	39.99	36.40	43.81
Had unprotected sex without a condom in past 3 months	Full Sample	50.02	50.09	50.04	47.35	47.98	46.80
	Subgroup:						
	Not Overage	49.38	49.12	49.75	47.54	48.26	47.04
	Overage	48.60	48.80	48.60	47.20	47.77	46.77
Intention to avoid pregnancy in next year	Full Sample	46.52	45.48	47.53	45.04	44.23	45.95
	Subgroup:						
	Not Overage	47.04	46.61	47.60	45.40	45.52	45.47
	Overage	46.05	44.45	47.65	44.21	42.71	46.01
Knowledge of efficacy of birth control in preventing pregnancy	Full Sample	50.05	50.07	50.04	46.99	47.40	46.66
	Subgroup:						
	Not Overage	49.81	49.29	50.17	46.70	47.53	46.08
	Overage	49.72	49.90	49.72	47.37	47.48	47.45
Knowledge of efficacy of birth control pills in preventing STIs	Full Sample	47.89	47.60	48.25	42.89	42.73	43.14
	Subgroup:						
	Not Overage	48.81	48.82	48.98	44.47	44.91	44.23
	Overage	46.78	46.31	47.45	41.20	40.66	41.97

Outcome	Sample	Standard Deviation of Outcomes			Standard Deviation of Outcome Residuals		
		Full	Treatment Group	Control Group	Full	Treatment Group	Control Group
Knowledge of efficacy of condoms in preventing STIs	Full Sample	44.62	44.01	45.30	43.23	43.29	43.25
	Subgroup:						
	Not Overage	41.29	41.24	41.50	40.39	41.42	39.56
	Overage	47.19	46.21	48.30	45.87	45.05	46.97
Knowledge of efficacy of condoms in preventing pregnancy	Full Sample	49.93	50.01	49.95	47.80	48.59	47.08
	Subgroup:						
	Not Overage	50.07	50.00	50.19	46.21	47.61	45.04
	Overage	49.83	50.17	49.35	49.02	49.16	49.09
Number of sexual partners in past 3 months	Full Sample	1.23	0.61	1.64	1.22	0.62	1.62
	Subgroup:						
	Not Overage	1.61	0.45	2.21	1.59	0.50	2.18
	Overage	0.63	0.73	0.52	0.63	0.71	0.53
Perceived access to birth control other than condoms	Full Sample	0.79	0.75	0.83	0.73	0.70	0.76
	Subgroup:						
	Not Overage	0.84	0.86	0.80	0.75	0.79	0.70
	Overage	0.74	0.62	0.86	0.71	0.60	0.82
Perceived access to condoms	Full Sample	0.75	0.73	0.76	0.72	0.70	0.73
	Subgroup:						
	Not Overage	0.80	0.75	0.85	0.77	0.72	0.81
	Overage	0.68	0.72	0.63	0.66	0.68	0.63
Perceived ease of using birth control	Full Sample	0.95	0.91	0.99	0.89	0.87	0.91
	Subgroup:						
	Not Overage	0.90	0.86	0.94	0.88	0.88	0.87
	Overage	0.99	0.95	1.04	0.90	0.85	0.95
Perceived need for birth control other than condoms	Full Sample	0.72	0.70	0.74	0.69	0.69	0.69
	Subgroup:						
	Not Overage	0.69	0.63	0.74	0.67	0.65	0.69
	Overage	0.75	0.76	0.74	0.71	0.71	0.71

Outcome	Sample	Standard Deviation of Outcomes			Standard Deviation of Outcome Residuals		
		Full	Treatment Group	Control Group	Full	Treatment Group	Control Group
Perceived need for condoms	Full Sample	0.64	0.60	0.68	0.62	0.59	0.65
	Subgroup:						
	Not Overage	0.69	0.63	0.74	0.67	0.63	0.70
	Overage	0.58	0.57	0.59	0.56	0.55	0.57
Perceived trustworthiness of birth control providers	Full Sample	0.89	0.89	0.89	0.85	0.85	0.86
	Subgroup:						
	Not Overage	0.83	0.81	0.85	0.79	0.76	0.82
	Overage	0.94	0.96	0.93	0.91	0.93	0.89
Received birth control from professional	Full Sample	44.75	42.53	46.62	43.34	41.96	44.78
	Subgroup:						
	Not Overage	42.73	41.50	43.91	42.17	41.57	42.87
	Overage	46.51	43.59	48.82	44.50	42.45	46.85
Received info on STIs in past 12 months	Full Sample	43.83	41.35	45.94	42.72	41.03	44.47
	Subgroup:						
	Not Overage	42.47	41.37	43.59	41.69	41.07	42.43
	Overage	45.13	41.50	48.01	43.74	41.07	46.69
Received info on abstinence in past 12 months	Full Sample	49.74	50.04	48.17	47.65	48.12	47.26
	Subgroup:						
	Not Overage	50.00	49.96	48.94	48.04	48.53	47.77
	Overage	49.48	50.20	47.35	47.24	47.84	46.77
Received info on birth control methods in past 12 months	Full Sample	36.94	30.64	41.66	36.17	30.81	41.01
	Subgroup:						
	Not Overage	34.89	28.87	39.30	34.64	28.95	39.42
	Overage	38.86	32.29	43.98	37.64	32.32	42.90
Received info on relationships in past 12 months	Full Sample	48.57	49.42	47.47	47.75	48.78	46.78
	Subgroup:						
	Not Overage	48.95	49.71	48.17	48.22	49.89	46.77
	Overage	48.24	49.31	46.81	47.37	47.95	46.92

Outcome	Sample	Standard Deviation of Outcomes			Standard Deviation of Outcome Residuals		
		Full	Treatment Group	Control Group	Full	Treatment Group	Control Group
Received info on saying no to sex in past 12 months	Full Sample	42.62	41.29	43.91	40.08	39.30	40.95
	Subgroup:						
	Not Overage	40.20	39.40	41.08	38.57	39.10	38.21
	Overage	44.72	43.01	46.46	41.59	39.63	43.84
Received info on sexual health from professional at facility in past 12 months	Full Sample	35.47	33.77	37.10	34.82	33.90	35.79
	Subgroup:						
	Not Overage	35.06	34.51	35.69	34.76	34.42	35.20
	Overage	35.94	33.20	38.67	34.87	33.42	36.55
Received info on sexual health from professional at home in past 12 months	Full Sample	49.89	46.25	48.77	45.95	45.10	46.89
	Subgroup:						
	Not Overage	49.34	45.76	50.14	46.61	43.75	49.30
	Overage	50.10	46.84	45.31	44.98	46.10	43.91
Received info on talking to partner about sex in past 12 months	Full Sample	43.11	42.10	44.14	41.43	40.67	42.29
	Subgroup:						
	Not Overage	41.90	43.36	40.51	40.52	42.26	38.97
	Overage	44.30	40.96	47.14	42.02	38.88	45.41
Received info on where to get birth control in past 12 months	Full Sample	35.00	29.01	39.64	34.07	29.14	38.55
	Subgroup:						
	Not Overage	32.73	27.65	36.69	32.56	28.28	36.31
	Overage	37.10	30.33	42.44	35.55	30.03	40.92
Used LARC method in past 3 months	Full Sample	45.47	47.91	41.88	43.70	46.79	40.31
	Subgroup:						
	Not Overage	45.32	47.99	41.86	43.47	46.41	40.66
	Overage	45.72	48.02	42.10	44.01	47.33	40.05

Outcome	Sample	Standard Deviation of Outcomes			Standard Deviation of Outcome Residuals		
		Full	Treatment Group	Control Group	Full	Treatment Group	Control Group
Used hormonal method or IUD in past 3 months	Full Sample	44.64	43.03	46.14	43.37	43.12	43.72
	Subgroup:						
	Not Overage	42.66	40.30	44.68	41.82	41.40	42.39
	Overage	46.38	45.25	47.66	44.97	44.83	45.34

Note: The residuals are calculated using all covariates included in the model. All calculations include study weights, if applicable.

II.C. Teach For America

Original study. Decker et al. (2004).

Sponsor agency. The Smith Richardson Foundation, The William and Flora Hewlett Foundation, The Carnegie Corporation.

Description of intervention. The original study evaluated the effects of Teach For America (TFA). TFA is an educational program designed to help low-income students by expanding the set of candidate teachers. TFA recruits graduating college students to serve as teachers in low-income schools for at least two years—in particular, students who have a high potential to be strong teachers but who might not otherwise serve as teachers. TFA recruits attend an intensive teacher training and continue to receive training after they are placed in a school.

Randomization design. Non-clustered, blocked. In the original study, elementary school students (grades 1 to 5) were randomly assigned to classrooms taught by either a TFA corps member or another teacher. The study featured a block design in which students were randomly assigned within schools and grades.

Model-based method. Ordinary least squares, block fixed effects.

We use OLS to estimate the following equation:

(15) $Y_{ij} = \alpha_j + \beta X_{ij} + \delta_j T_{ij} + \varepsilon_{ij}$,

where Y_{ij} is the outcome of interest for person i in block j, T_{ij} is a treatment indicator, X_{ij} is a vector of covariates, α_j is a block effect that is assumed to be fixed, and ε_{ij} is an error term. We assume that ε_{ij} is independent between observations. This estimation yields a treatment effect for each block, δ_j. We then average these block-specific effects to form an overall treatment effect. We weight the blocks by the sum of the sample weights in each block.

Design-based specification: RCT-YES model. Non-clustered, blocked.

Design-based specification: RCT-YES parameter (matched settings). Finite population (ATE).

Design-based specification: RCT-YES parameter (default settings). Finite population (ATE).

Covariates. In our analysis, we adjust for preprogram math score, preprogram reading score, gender, race/ethnicity, free lunch status, old-for-grade status, and the percentage of students in the classroom not in the research sample. This is consistent with the approach in the original report.

Weights. Yes. We create weights to account for missing outcome data by using the method outlined in the original report:

- Estimate a logit model for the probability of having completed a spring achievement test, using the following covariates: treatment status, black, grade, gender, ethnicity, old-for grade status, free lunch status, percentage of classroom not in research sample, preprogram reading score, and preprogram math score.

- For each observation, predict the probability that the test score was missing.

- Form groups on the basis of 10 equal intervals of the propensity score distribution, compute the average propensity score within each group, and create a nonresponse weight, which is the inverse of the probability of selection within that group.

- To incorporate weighting of the blocks, sum the nonresponse weights (1/group mean of the propensity scores).

Multiple hypothesis correction. None. We do not apply a multiple hypothesis correction, because one was not applied to the main results in the original study. There are two primary outcomes, each of which is in a different domain.

Main differences from original analyses. Our model-based methods differ in several ways from the original analysis:

- The original analysis applied two estimation steps. In the first step, a series of block-by-treatment indicators were estimated. This first step is analogous to our first step. In the second step, a weighted least squares regression was applied, in which the estimated treatment effects for each block were regressed on a set of block-specific covariates, yielding an average treatment effect across blocks. This procedure is not directly comparable to a finite-population approach. Therefore, we adopt a simplified approach in which we estimate a series of treatment effects by block and average over the block-specific effects without additionally adjusting for block-level covariates.

- We impute missing values of the covariates by using the method adopted by *RCT-YES*.

Outcomes. We estimate the model for the following outcomes:

Math total, normal curve equivalent: Math component of the Iowa Test of Basic Skills, presented as a normal curve equivalent, in which scores can range from 0 to 100.

Reading total, normal curve equivalent: Reading component of the Iowa Test of Basic Skills, presented as a normal curve equivalent, in which scores can range from 0 to 100.

Table II.C.1. Teach for America: Evaluation Impact Estimates (Model-Based)

Outcome	Sample	Control Mean	Impact	SE	p-value	Effect Size
Math total, normal curve equivalent	Full Sample	27.84	2.26	0.69	0.001*	0.14
	Subgroup:				0.523	
	Female	27.38	2.71	0.99	0.006*	0.18
	Male	28.26	1.82	0.97	0.062	0.11
Reading total, normal curve equivalent	Full Sample	27.10	0.42	0.66	0.521	0.03
	Subgroup:				0.972	
	Female	27.59	0.48	0.95	0.616	0.03
	Male	26.64	0.43	0.93	0.644	0.03

Note: The p-value on the subgroup rows tests for differences in impacts across subgroup categories. Effect sizes are calculated by dividing the impact estimate by the standard deviation of the outcome in the control group.
* Difference is statistically significant at the 0.05 level, two-tailed test.

Table II.C.2. Teach for America: Evaluation Impact Estimates (Design-Based, Matching Assumptions)

							Absolute Value of Difference Compared to Model-Based Specification		
Outcome	Sample	Control Mean	Impact	SE	p-value	Effect Size	Dif. Effect Size	Dif. SE (Effect Size)	Dif. p-value
Math total, normal curve equivalent	Full Sample	27.84	2.26	0.66	0.001*	0.14	0.00	0.00	0.000
	Subgroup:				0.500				0.023
	Female	27.38	2.71	0.88	0.002*	0.18	0.00	0.01	0.004
	Male	28.26	1.82	0.97	0.061	0.11	0.00	0.00	0.001
Reading total, normal curve equivalent	Full Sample	27.10	0.42	0.66	0.517	0.03	0.00	0.00	0.004
	Subgroup:				0.972				0.000
	Female	27.59	0.48	0.90	0.598	0.03	0.00	0.00	0.018
	Male	26.64	0.43	0.94	0.648	0.03	0.00	0.00	0.004

Note: The p-value on the subgroup rows tests for differences in impacts across subgroup categories. Effect sizes are calculated by dividing the impact estimate by the standard deviation of the outcome in the control group.
* Difference is statistically significant at the 0.05 level, two-tailed test.

Table II.C.3. Teach for America: Evaluation Impact Estimates (Design-Based, *RCT-YES* Default Assumptions)

							Absolute Value of Difference Compared to Model-Based Specification		
Outcome	Sample	Control Mean	Impact	SE	p-value	Effect Size	Dif. Effect Size	Dif. SE (Effect Size)	Dif. p-value
Math total, normal curve equivalent	Full Sample	27.84	2.26	0.66	0.001*	0.14	0.00	0.00	0.000
	Subgroup:				0.500				0.023
	Female	27.38	2.71	0.88	0.002*	0.18	0.00	0.01	0.004
	Male	28.26	1.82	0.97	0.061	0.11	0.00	0.00	0.001
Reading total, normal curve equivalent	Full Sample	27.10	0.42	0.66	0.517	0.03	0.00	0.00	0.004
	Subgroup:				0.972				0.000
	Female	27.59	0.48	0.90	0.598	0.03	0.00	0.00	0.018
	Male	26.64	0.43	0.94	0.648	0.03	0.00	0.00	0.004

Note: The p-value on the subgroup rows tests for differences in impacts across subgroup categories. Effect sizes are calculated by dividing the impact estimate by the standard deviation of the outcome in the control group.
* Difference is statistically significant at the 0.05 level, two-tailed test.

Table II.C.4. Teach for America: Additional Information (Individual-Level Randomization)

Outcome	Sample	Standard Deviation of Outcomes			Standard Deviation of Outcome Residuals		
		Full	Treatment Group	Control Group	Full	Treatment Group	Control Group
Math total, normal curve equivalent	Full Sample	16.58	16.95	16.23	11.91	11.66	12.08
	Subgroup:						
	Female	15.73	16.33	15.08	10.91	11.13	10.77
	Male	17.34	17.51	17.22	12.41	11.78	12.81
Reading total, normal curve equivalent	Full Sample	16.64	17.25	16.19	11.41	11.70	11.21
	Subgroup:						
	Female	16.00	16.47	15.56	10.93	10.89	10.97
	Male	17.17	17.75	16.75	11.39	12.01	10.94

Note: The residuals are calculated using all covariates included in the model. All calculations include study weights, if applicable.

II.D. Charter School Impacts

Original study. Gleason et al. (2010).

Sponsor agency. Institute of Education Sciences, U.S. Department of Education.

Description of intervention. The original study evaluated a group of 36 charter middle schools across 15 states. To participate in the study, each charter school was required to meet two criteria: (1) be in operation for at least two years, and (2) have more applicants to its entry grade than available spots. Not all schools that met these criteria participated in the study.

Randomization design. Non-clustered, blocked. Each charter school in the evaluation offered a lottery in which eligible applicants were then assigned to either a treatment or control group.

Model-based method. Ordinary least squares, block fixed effects.

We use OLS to estimate the following equation:

$$(16)\quad Y_{ij} = \alpha_j + \beta X_{ij} + \delta_j T_{ij} + \varepsilon_{ij},$$

where Y_{ij} is the outcome of interest for person i in block j, T_{ij} is a treatment indicator, X_{ij} is a vector of covariates, α_j is a block effect that is assumed to be fixed, and ε_{ij} is an error term. We assume that ε_{ij} is independent between observations. This estimation yields a treatment effect for each block, δ_j. We then average these block-specific effects to form an overall treatment effect. We weight the blocks by the sum of the sample weights in each block.

Design-based specification: RCT-YES model. Non-clustered, blocked.

Design-based specification: RCT-YES parameter (matched settings). Finite population (ATE).

Design-based specification: RCT-YES parameter (default settings). Finite population (ATE).

Covariates. In our analysis, we include pre-baseline math score (z-score units), pre-baseline math proficiency level, baseline math score (z-score units), baseline math proficiency level, pre-baseline reading score (z-score units), pre-baseline reading proficiency level, baseline reading score (z-score units), baseline reading proficiency level, number of days absent in baseline school year, number of days suspended in baseline school year, race, age at start of school year, young- or old-for-grade status, IEP status, LEP status, income-to-poverty ratio, two-parent family status, two-adult family status (at least one nonparent), English as main language spoken in home, mother's education, born in United States status, family's receipt of TANF or food stamps in past month, free or reduced-price lunch status, number of children in the household, school enrollment at baseline (charter, private, or public), changed schools midyear in baseline school status, applied to other charter schools at baseline status, applied to private school at baseline status, applied to other public school at baseline status, baseline information form collected before lottery status, and second cohort status. This is consistent with the approach in the original report.

Weights. Yes. For both the model-based and design-based estimation, we use the weights specified in the original study, which account for the fact that different students faced different probabilities of being selected into the treatment and control groups.

Multiple hypothesis correction. Yes. We adjust for multiple hypotheses by using the Benjamini-Hochberg procedure, which is the same procedure used in the original report.

Main differences from original analyses. The original study imputed missing values of the control variables as the mean by site and cohort for continuous variables and the mode for discrete variables. The main regression model included indicators for whether the variable was imputed. In contrast, we apply the imputation method used by *RCT-YES*.

Outcomes. We estimate the model for the following outcomes:

Math score (z-score units), follow-up 1: Math score on state assessment in *z*-score units, one year after students applied to study charter schools.

Math score (z-score units), follow-up 2: Math score on state assessment in *z*-score units, two years after students applied to study charter schools.

Reading score (z-score units), follow-up 1: Reading score on state assessment in *z*-score units, one year after students applied to study charter schools.

Reading score (z-score units), follow-up 2: Reading score on state assessment in *z*-score units, two years after students applied to study charter schools.

Table II.D.1. Charter School Impacts: Evaluation Impact Estimates (Model-Based)

Outcome	Sample	Control Mean	Impact	SE	p-value	Effect Size
Math score (z-score units), follow-up 1	Full Sample	0.36	-0.02	0.02	0.493	-0.02
	Subgroup:				0.464	
	Female	0.28	-0.03	0.03	0.301	-0.03
	Male	0.46	0.00	0.04	0.965	0.00
Math score (z-score units), follow-up 2	Full Sample	0.44	-0.05	0.03	0.100	-0.04
	Subgroup:				0.042*	
	Female	0.39	-0.12	0.04	0.008*	-0.11
	Male	0.51	0.02	0.05	0.740	0.02
Reading score (z-score units), follow-up 1	Full Sample	0.44	-0.02	0.03	0.472	-0.02
	Subgroup:				0.868	
	Female	0.44	-0.01	0.04	0.828	-0.01
	Male	0.43	-0.02	0.04	0.673	-0.02
Reading score (z-score units), follow-up 2	Full Sample	0.44	-0.08	0.03	0.003*^	-0.08
	Subgroup:				0.193	
	Female	0.48	-0.11	0.04	0.005*	-0.11
	Male	0.38	-0.03	0.04	0.434	-0.03

Note: The p-value on the subgroup rows tests for differences in impacts across subgroup categories. Effect sizes are calculated by dividing the impact estimate by the standard deviation of the outcome in the control group.
* Difference is statistically significant at the 0.05 level, two-tailed test.
^ Difference remains statistically significant at the 0.05 level, two-tailed test, after applying the Benjamini-Hochberg correction for multiple hypothesis testing across all full sample analyses in the same domain.

Table II.D.2. Charter School Impacts: Evaluation Impact Estimates (Design-Based, Matching Assumptions)

Outcome	Sample	Control Mean	Impact	SE	p-value	Effect Size	Absolute Value of Difference Compared to Model-Based Specification		
							Dif. Effect Size	Dif. SE (Effect Size)	Dif. p-value
Math score (z-score units), follow-up 1	Full Sample	0.36	-0.02	0.03	0.554	-0.02	0.00	0.01	0.061
	Subgroup:				0.488				0.024
	Female	0.28	-0.03	0.04	0.375	-0.03	0.00	0.01	0.074
	Male	0.46	0.00	0.04	0.911	0.00	0.00	0.00	0.054
Math score (z-score units), follow-up 2	Full Sample	0.44	-0.05	0.04	0.164	-0.04	0.00	0.01	0.064
	Subgroup:				0.088				0.046
	Female	0.39	-0.11	0.05	0.019*	-0.10	0.01	0.01	0.011
	Male	0.52	0.02	0.06	0.762	0.02	0.00	0.01	0.022
Reading score (z-score units), follow-up 1	Full Sample	0.44	-0.02	0.03	0.522	-0.02	0.00	0.00	0.050
	Subgroup:				0.975				0.107
	Female	0.44	-0.01	0.04	0.843	-0.01	0.00	0.00	0.015
	Male	0.43	-0.01	0.04	0.821	-0.01	0.01	0.00	0.148
Reading score (z-score units), follow-up 2	Full Sample	0.44	-0.08	0.03	0.009*^	-0.08	0.00	0.00	0.006
	Subgroup:				0.305				0.112
	Female	0.48	-0.11	0.04	0.013*	-0.11	0.00	0.00	0.008
	Male	0.38	-0.03	0.06	0.569	-0.03	0.00	0.02	0.135

Note: The p-value on the subgroup rows tests for differences in impacts across subgroup categories. Effect sizes are calculated by dividing the impact estimate by the standard deviation of the outcome in the control group.
* Difference is statistically significant at the 0.05 level, two-tailed test.
^ Difference remains statistically significant at the 0.05 level, two-tailed test, after applying the Benjamini-Hochberg correction for multiple hypothesis testing across all full sample analyses in the same domain.

Table II.D.3. Charter School Impacts: Evaluation Impact Estimates (Design-Based, *RCT-YES* Default Assumptions)

							Absolute Value of Difference Compared to Model-Based Specification		
Outcome	Sample	Control Mean	Impact	SE	p-value	Effect Size	Dif. Effect Size	Dif. SE (Effect Size)	Dif. p-value
Math score (z-score units), follow-up 1	Full Sample	0.36	-0.02	0.03	0.554	-0.02	0.00	0.01	0.061
	Subgroup:				0.488				0.024
	Female	0.28	-0.03	0.04	0.375	-0.03	0.00	0.01	0.074
	Male	0.46	0.00	0.04	0.911	0.00	0.00	0.00	0.054
Math score (z-score units), follow-up 2	Full Sample	0.44	-0.05	0.04	0.164	-0.04	0.00	0.01	0.064
	Subgroup:				0.088				0.046
	Female	0.39	-0.11	0.05	0.019*	-0.10	0.01	0.01	0.011
	Male	0.52	0.02	0.06	0.762	0.02	0.00	0.01	0.022
Reading score (z-score units), follow-up 1	Full Sample	0.44	-0.02	0.03	0.522	-0.02	0.00	0.00	0.050
	Subgroup:				0.975				0.107
	Female	0.44	-0.01	0.04	0.843	-0.01	0.00	0.00	0.015
	Male	0.43	-0.01	0.04	0.821	-0.01	0.01	0.00	0.148
Reading score (z-score units), follow-up 2	Full Sample	0.44	-0.08	0.03	0.009*^	-0.08	0.00	0.00	0.006
	Subgroup:				0.305				0.112
	Female	0.48	-0.11	0.04	0.013*	-0.11	0.00	0.00	0.008
	Male	0.38	-0.03	0.06	0.569	-0.03	0.00	0.02	0.135

Note: The p-value on the subgroup rows tests for differences in impacts across subgroup categories. Effect sizes are calculated by dividing the impact estimate by the standard deviation of the outcome in the control group.
* Difference is statistically significant at the 0.05 level, two-tailed test.
^ Difference remains statistically significant at the 0.05 level, two-tailed test, after applying the Benjamini-Hochberg correction for multiple hypothesis testing across all full sample analyses in the same domain.

Table II.D.4. Charter School Impacts: Additional Information (Individual-Level Randomization)

		Standard Deviation of Outcomes			Standard Deviation of Outcome Residuals		
Outcome	Sample	Full	Treatment Group	Control Group	Full	Treatment Group	Control Group
Math score (z-score units), follow-up 1	Full Sample	1.03	0.99	1.06	0.54	0.55	0.52
	Subgroup:						
	Female	0.97	0.90	1.04	0.51	0.51	0.51
	Male	1.08	1.09	1.08	0.55	0.58	0.53
Math score (z-score units), follow-up 2	Full Sample	1.14	1.08	1.21	0.69	0.67	0.70
	Subgroup:						
	Female	1.07	1.02	1.12	0.64	0.67	0.62
	Male	1.22	1.14	1.31	0.71	0.67	0.75
Reading score (z-score units), follow-up 1	Full Sample	0.95	0.96	0.93	0.57	0.57	0.57
	Subgroup:						
	Female	0.90	0.91	0.90	0.54	0.55	0.54
	Male	0.99	0.99	0.97	0.58	0.57	0.59
Reading score (z-score units), follow-up 2	Full Sample	0.96	0.92	1.01	0.61	0.59	0.63
	Subgroup:						
	Female	0.91	0.85	0.97	0.58	0.58	0.58
	Male	1.01	0.97	1.06	0.61	0.57	0.66

Note: The residuals are calculated using all covariates included in the model. All calculations include study weights, if applicable.

II.E. Teacher Prep

Original study. Constantine et al. (2009).

Sponsor agency. Institute of Education Sciences, U.S. Department of Education.

Description of intervention. The original study evaluated the effectiveness of teachers who were trained through alternative routes to certification, compared to those trained through traditional routes to certification. The study focuses on 63 alternative training programs and targets schools that serve children in grades K–5 who had at least one teacher trained through an alternative route to certification and one teacher trained through a traditional route to certification.

Randomization design. Non-clustered, blocked. Within each eligible school (block), students were randomly assigned to either a classroom with a teacher trained through an alternative route to certification or a teacher trained through a traditional route to certification.

Model-based method. Ordinary least squares, block fixed effects.

We use OLS to estimate the following equation:

$$(17)\ Y_{ij} = \alpha_j + \beta X_{ij} + \delta_j T_{ij} + \varepsilon_{ij},$$

where Y_{ij} is the outcome of interest for person i in block j, T_{ij} is a treatment indicator, X_{ij} is a vector of covariates, α_j is a block effect that is assumed to be fixed, and ε_{ij} is an error term. We assume that ε_{ij} is independent between observations. This estimation yields a treatment effect for each block, δ_j. We then average these block-specific effects to form an overall treatment effect. We weight the blocks by the sum of the sample weights in each block.

Design-based specification: RCT-YES model. Non-clustered, blocked.

Design-based specification: RCT-YES parameter (matched settings). Finite population (ATE).

Design-based specification: RCT-YES parameter (default settings). Finite population (ATE).

Covariates. In our analysis, we control for baseline test scores (normal curve equivalent) in reading vocabulary, reading comprehension, and math concepts. We additionally control for race, gender, eligibility for free or reduced-price lunch, and teacher's years of experience. This is consistent with the approach in the original report.

Weights. Yes. We specify the weights included in the public use dataset.

Multiple hypothesis correction. None. We do not apply a multiple hypothesis correction, because one was not applied to the main results in the original study.

Main differences from original analyses:

- We exclude 11 observations for which the weights are missing.
- A super-population model (PATE) is used for the subgroup analysis because *RCT-YES* places limits on the number of blocks that may be included in a finite population model in order to maintain degrees of freedom.

Outcomes. We estimate the model for the following outcomes:

Post-test math, normal curve equivalent: Math post-test score on the California Achievement Test, 5th Edition (CAT-5), normal curve equivalent.

Post-test reading, normal curve equivalent: Reading post-test score on the California Achievement Test, 5th Edition (CAT-5), normal curve equivalent.

Table II.E.1. Teacher Prep: Evaluation Impact Estimates (Model-Based)

Outcome	Sample	Control Mean	Impact	SE	p-value	Effect Size
Post-test math, normal curve equivalent	Full Sample	42.52	-1.10	0.68	0.108	-0.05
	Subgroup:				0.554	
	Female	42.84	-1.28	0.97	0.186	-0.06
	Male	43.61	-2.01	0.91	0.027*	-0.09
Post-test reading, normal curve equivalent	Full Sample	38.58	-0.15	0.56	0.787	-0.01
	Subgroup:				0.579	
	Female	39.76	-0.08	0.79	0.918	0.00
	Male	38.61	-0.64	0.75	0.389	-0.03

Note: The p-value on the subgroup rows tests for differences in impacts across subgroup categories. Effect sizes are calculated by dividing the impact estimate by the standard deviation of the outcome in the control group.
* Difference is statistically significant at the 0.05 level, two-tailed test.

Table II.E.2. Teacher Prep: Evaluation Impact Estimates (Design-Based, Matching Assumptions)

Outcome	Sample	Control Mean	Impact	SE	p-value	Effect Size	Absolute Value of Difference Compared to Model-Based Specification: Dif. Effect Size	Dif. SE (Effect Size)	Dif. p-value
Post-test math, normal curve equivalent	Full Sample	42.52	-1.10	0.62	0.078	-0.05	0.00	0.00	0.030
	Subgroup:				0.471				0.083
	Female	42.84	-1.46	1.23	0.238	-0.07	0.01	0.01	0.052
	Male	43.58	-0.22	1.20	0.854	-0.01	0.08	0.01	0.827
Post-test reading, normal curve equivalent	Full Sample	38.58	-0.15	0.52	0.769	-0.01	0.00	0.00	0.018
	Subgroup:				0.685				0.106
	Female	39.70	-0.26	1.01	0.794	-0.01	0.01	0.01	0.124
	Male	38.62	0.35	1.14	0.756	0.02	0.05	0.02	0.367

Note: The p-value on the subgroup rows tests for differences in impacts across subgroup categories. Effect sizes are calculated by dividing the impact estimate by the standard deviation of the outcome in the control group.
* Difference is statistically significant at the 0.05 level, two-tailed test.

Table II.E.3. Teacher Prep: Evaluation Impact Estimates (Design-Based, *RCT-YES* Default Assumptions)

Outcome	Sample	Control Mean	Impact	SE	p-value	Effect Size	Absolute Value of Difference Compared to Model-Based Specification		
							Dif. Effect Size	Dif. SE (Effect Size)	Dif. p-value
Post-test math, normal curve equivalent	Full Sample	42.52	-1.10	0.62	0.078	-0.05	0.00	0.00	0.030
	Subgroup:				0.471				0.083
	Female	42.84	-1.46	1.23	0.238	-0.07	0.01	0.01	0.052
	Male	43.58	-0.22	1.20	0.854	-0.01	0.08	0.01	0.827
Post-test reading, normal curve equivalent	Full Sample	38.58	-0.15	0.52	0.769	-0.01	0.00	0.00	0.018
	Subgroup:				0.685				0.106
	Female	39.70	-0.26	1.01	0.794	-0.01	0.01	0.01	0.124
	Male	38.62	0.35	1.14	0.756	0.02	0.05	0.02	0.367

Note: The p-value on the subgroup rows tests for differences in impacts across subgroup categories. Effect sizes are calculated by dividing the impact estimate by the standard deviation of the outcome in the control group.
* Difference is statistically significant at the 0.05 level, two-tailed test.

Table II.E.4. Teacher Prep: Additional Information (Individual-Level Randomization)

Outcome	Sample	Standard Deviation of Outcomes			Standard Deviation of Outcome Residuals		
		Full	Treatment Group	Control Group	Full	Treatment Group	Control Group
Post-test math, normal curve equivalent	Full Sample	22.09	21.78	22.41	13.73	13.34	14.13
	Subgroup:						
	Female	21.58	21.13	22.06	13.05	12.25	13.90
	Male	22.52	22.35	22.71	13.52	13.42	13.63
Post-test reading, normal curve equivalent	Full Sample	20.05	20.14	19.97	11.53	11.55	11.50
	Subgroup:						
	Female	19.60	19.29	19.92	10.16	9.95	10.40
	Male	20.41	21.00	19.82	11.74	11.95	11.53

Note: The residuals are calculated using all covariates included in the model. All calculations include study weights, if applicable.

II.F. Roads to Success

Original study. Chaplin et al. (2010).

Sponsor agency. Roads to Success.

Description of intervention. *Roads to Success* (RTS) is a school and career planning program that is designed to provide guidance to students regarding their future and to foster engagement in schools. It is a classroom-based program that serves entire cohorts of students, ranging from grade 7 to grade 12.

Randomization design. Clustered, blocked. In the original study, 25 schools were randomly assigned to either a treatment or control group condition. Schools were assigned within blocks determined by their geographic location.

Model-based method. Ordinary least squares, robust cluster standard errors.

We use OLS to estimate the following equation:

(18) $Y_{ij} = \alpha + \beta X_{ij} + \delta T_j + \varepsilon_{ij}$,

where Y_{ij} is the outcome of interest for individual i in cluster j, X_{ij} is a vector of covariates, T_j is a treatment indicator, and ε_{ij} is an error term. As in the original study, we allow for ε_{ij} to be arbitrarily correlated within clusters but uncorrelated across clusters. Note that we do not include block effects in the analysis because the original study did not include block effects.

Design-based specification: RCT-YES model. Clustered, non-blocked. We specify this design because the original study did not include block effects in the analysis.

Design-based specification: RCT-YES parameter (matched settings). Finite population (ATE).

Design-based specification: RCT-YES parameter (default settings). Finite population (ATE).

Covariates. In our analysis, we do not adjust for all covariates listed in the original report because not all of these covariates are accessible in the public use dataset. Further, some covariates used in the original study are only available at the school level in the public use data, so we replace the individual-level covariates with their school-level counterparts. In addition, *RCT-YES* has limits on the number of covariates that can be included relative to the number of clusters to avoid overfitting the model. The number of covariates specified in the original report exceeds the number that *RCT-*

YES allows.

Our final list of covariates includes school-level averages of the following: internet access at home, mother's high school degree status, mother's bachelor's degree status, father's high school degree status, father's bachelor's degree status, overage for grade status, English speaker status, free lunch status, reading test scores, and math test scores. The model includes a student-level indicator for gender.

Weights. Yes. For both the model-based and design-based estimates, we use the weights from the original study that adjust for the imbalance of treatment and control students with respect to the blocks of matched schools. In other words, individual observations are re-weighted so that contributions of the treatment and control groups are equal within a block.

Multiple hypothesis correction. None. We do not apply a multiple hypothesis correction, because one was not applied to the main results in the original study.

Main differences from original analyses. Our methods differ in several ways from the original analysis:

- We include fewer covariates than the original analysis for two reasons. First, we do not have access to all of the individual-level covariates that the original study used because the data have been aggregated. Second, *RCT-YES* restricts the number of covariates that can be included relative to the number of clusters. The number of covariates included in the original study exceeds the number of covariates allowed under *RCT-YES* restrictions, so we exclude several covariates.

Outcomes. We estimate the model for the following outcomes:

Motivation to go to school to learn job skills: Self-report of the extent to which students attend school in order to learn skills for a job on a scale of 1 to 4, where higher values indicate more motivation to learn skills for a job.

Learning and study habits/preparation: Self-report of the extent to which students exhibit good classroom habits, such as sticking with a classroom assignment until it is completed, on a scale of 1 to 5.

School attendance and negative behaviors: Aggregated measure of self-reported school absences and negative behaviors on a scale of 1 to 4, with higher values indicating more school absences and negative behaviors.

Number of times late for school: Self-reported measure of number of times late for school.

Number of times cut/skipped class: Self-reported measure of number of times cut or skipped class.

Number of times absent from school: Self-reported measure of number of times absent from school.

Number of times sent out of class for bad behavior: Self-reported measure of number of times sent out of class for bad behavior.

Number of times received detention: Self-reported measure of number of times received detention.

Career exploration behavior with parents: The frequency at which the student discusses later outcomes (e.g., attending college) with a parent, on a scale of 1 to 5, with higher values indicating more frequency.

Career exploration behavior with teachers/school staff: The frequency at which the student discusses later outcomes (e.g., attending college) with a teacher, on a scale of 1 to 5, with higher values indicating more frequency.

School engagement: Self-report of student engagement, on a scale of 1 to 4, with higher values indicating more engagement.

Importance of grades: Self-report of the extent to which grades are important, on a scale of 1 to 4, with higher values indicating that the student places more importance on grades.

Career exploration efficacy: Self-report of the extent to which students believe they understand how to find a suitable job, on a scale of 1 to 4, with higher values indicating greater understanding.

Knowledge of requirements to succeed in different careers: Self-report of the extent to which students believe they understand what is required to succeed in different careers, on a scale of 1 to 4, with higher values indicating more knowledge.

Knowledge of how to determine what types of jobs are a good fit: Self-report of the extent to which students believe they know how to find careers that are a good fit, on a scale of 1 to 4, with higher values indicating more knowledge.

Knowledge of the types of jobs that are a good fit: Self-report of the extent to which students believe they know which types of jobs would be a good fit for themselves, on a scale of 1 to 4, with higher values indicating more knowledge.

Knowledge of how to overcome barriers to career goals: Self-report of the extent to which students believe they know how to overcome barriers to achieving career goals, on a scale of 1 to 4, with higher values indicating more knowledge.

Table II.F.1. Roads to Success: Evaluation Impact Estimates (Model-Based)

Outcome	Sample	Control Mean	Impact	SE	p-value	Effect Size
Career exploration behavior with parents	Full Sample	3.25	-0.08	0.12	0.516	-0.08
	Subgroup:				0.304	
	Female	3.41	-0.15	0.13	0.270	-0.15
	Male	3.07	0.00	0.15	0.985	0.00
Career exploration behavior with teachers/school staff	Full Sample	2.51	0.21	0.17	0.222	0.19
	Subgroup:				0.883	
	Female	2.61	0.22	0.15	0.153	0.21
	Male	2.38	0.20	0.22	0.390	0.18
Career exploration efficacy	Full Sample	3.24	0.04	0.03	0.214	0.08
	Subgroup:				0.281	
	Female	3.23	0.07	0.04	0.124	0.15
	Male	3.25	0.01	0.04	0.898	0.02
Importance of grades	Full Sample	3.44	-0.10	0.07	0.175	-0.13
	Subgroup:				0.859	
	Female	3.51	-0.11	0.09	0.254	-0.16
	Male	3.36	-0.09	0.08	0.308	-0.11
Knowledge of how to determine what types of jobs are a good fit	Full Sample	3.10	0.15	0.04	0.001*	0.21
	Subgroup:				0.807	
	Female	3.08	0.16	0.07	0.036*	0.23
	Male	3.12	0.14	0.04	0.003*	0.18
Knowledge of how to overcome barriers to career goals	Full Sample	3.29	0.01	0.05	0.912	0.02
	Subgroup:				0.151	
	Female	3.27	0.05	0.06	0.395	0.08
	Male	3.31	-0.05	0.07	0.476	-0.07
Knowledge of requirements to succeed in different careers	Full Sample	3.13	0.02	0.09	0.783	0.03
	Subgroup:				0.312	
	Female	3.10	0.06	0.09	0.510	0.10
	Male	3.15	-0.02	0.10	0.876	-0.03
Knowledge of the types of jobs that are a good fit	Full Sample	3.43	0.00	0.04	0.965	0.00
	Subgroup:				0.617	
	Female	3.45	0.01	0.06	0.819	0.02
	Male	3.40	-0.02	0.05	0.676	-0.03
Learning and study habits/preparation	Full Sample	4.06	-0.13	0.05	0.020*	-0.19
	Subgroup:				0.998	
	Female	4.16	-0.13	0.07	0.072	-0.20
	Male	3.93	-0.13	0.07	0.089	-0.18
Motivation to go to school to learn job skills	Full Sample	3.32	-0.05	0.09	0.582	-0.07
	Subgroup:				0.416	
	Female	3.38	-0.09	0.10	0.390	-0.15
	Male	3.25	0.00	0.10	0.992	0.00

Outcome	Sample	Control Mean	Impact	SE	p-value	Effect Size
Number of times absent from school	Full Sample	2.68	0.04	0.09	0.690	0.05
	Subgroup:				0.552	
	Female	2.70	0.06	0.09	0.497	0.07
	Male	2.66	0.00	0.11	0.990	0.00
Number of times cut/skipped class	Full Sample	1.21	0.12	0.09	0.184	0.22
	Subgroup:				0.625	
	Female	1.19	0.10	0.10	0.308	0.18
	Male	1.24	0.15	0.10	0.169	0.28
Number of times late for school	Full Sample	1.86	-0.02	0.18	0.905	-0.02
	Subgroup:				0.124	
	Female	1.80	0.04	0.19	0.851	0.05
	Male	1.94	-0.09	0.18	0.632	-0.10
Number of times received detention	Full Sample	1.79	-0.20	0.12	0.105	-0.22
	Subgroup:				0.461	
	Female	1.54	-0.14	0.11	0.223	-0.18
	Male	2.08	-0.27	0.18	0.140	-0.26
Number of times sent out of class for bad behavior	Full Sample	1.42	-0.08	0.08	0.308	-0.10
	Subgroup:				0.542	
	Female	1.22	-0.05	0.08	0.529	-0.10
	Male	1.66	-0.12	0.12	0.312	-0.13
School attendance and negative behaviors	Full Sample	1.80	-0.04	0.06	0.534	-0.08
	Subgroup:				0.345	
	Female	1.69	0.00	0.07	0.987	0.00
	Male	1.92	-0.08	0.07	0.273	-0.14
School engagement	Full Sample	2.55	0.04	0.07	0.626	0.05
	Subgroup:				0.454	
	Female	2.63	-0.01	0.10	0.916	-0.02
	Male	2.46	0.09	0.10	0.367	0.11

Note: The p-value on the subgroup rows tests for differences in impacts across subgroup categories. Effect sizes are calculated by dividing the impact estimate by the standard deviation of the outcome in the control group.
* Difference is statistically significant at the 0.05 level, two-tailed test.

Table II.F.2. Roads to Success: Evaluation Impact Estimates (Design-Based, Matching Assumptions)

Outcome	Sample	Control Mean	Impact	SE	p-value	Effect Size	Absolute Value of Difference Compared to Model-Based Specification		
							Dif. Effect Size	Dif. SE (Effect Size)	Dif. p-value
Career exploration behavior with parents	Full Sample	3.25	-0.08	0.10	0.450	-0.08	0.00	0.02	0.066
	Subgroup:				0.361				0.057
	Female	3.41	-0.15	0.07	0.042*	-0.15	0.00	0.06	0.228
	Male	3.07	0.00	0.15	0.985	0.00	0.00	0.00	0.000
Career exploration behavior with teachers/school staff	Full Sample	2.51	0.21	0.15	0.181	0.19	0.00	0.02	0.041
	Subgroup:				0.897				0.014
	Female	2.61	0.22	0.12	0.088	0.21	0.00	0.03	0.065
	Male	2.38	0.20	0.17	0.266	0.18	0.00	0.04	0.124
Career exploration efficacy	Full Sample	3.24	0.04	0.03	0.198	0.08	0.00	0.00	0.016
	Subgroup:				0.335				0.054
	Female	3.23	0.07	0.03	0.060	0.15	0.00	0.02	0.064
	Male	3.25	0.01	0.04	0.905	0.02	0.00	0.00	0.007
Importance of grades	Full Sample	3.44	-0.10	0.07	0.177	-0.13	0.00	0.00	0.002
	Subgroup:				0.870				0.011
	Female	3.51	-0.11	0.08	0.185	-0.16	0.00	0.01	0.069
	Male	3.36	-0.09	0.08	0.271	-0.11	0.00	0.00	0.037
Knowledge of how to determine what types of jobs are a good fit	Full Sample	3.10	0.15	0.03	0.001*	0.21	0.00	0.01	0.000
	Subgroup:				0.824				0.017
	Female	3.08	0.16	0.06	0.016*	0.23	0.00	0.01	0.020
	Male	3.12	0.14	0.06	0.027*	0.18	0.00	0.03	0.024
Knowledge of how to overcome barriers to career goals	Full Sample	3.29	0.01	0.05	0.902	0.02	0.00	0.00	0.010
	Subgroup:				0.206				0.055
	Female	3.27	0.05	0.04	0.246	0.08	0.00	0.03	0.149
	Male	3.31	-0.05	0.06	0.439	-0.07	0.00	0.01	0.037

Outcome	Sample	Control Mean	Impact	SE	p-value	Effect Size	Absolute Value of Difference Compared to Model-Based Specification		
							Dif. Effect Size	Dif. SE (Effect Size)	Dif. p-value
Knowledge of requirements to succeed in different careers	Full Sample	3.13	0.02	0.05	0.644	0.03	0.00	0.06	0.139
	Subgroup:				0.389				0.077
	Female	3.10	0.06	0.05	0.237	0.10	0.00	0.07	0.273
	Male	3.15	-0.02	0.07	0.813	-0.03	0.00	0.04	0.063
Knowledge of the types of jobs that are a good fit	Full Sample	3.43	0.00	0.03	0.950	0.00	0.00	0.02	0.015
	Subgroup:				0.652				0.035
	Female	3.45	0.01	0.05	0.781	0.02	0.00	0.02	0.038
	Male	3.40	-0.02	0.04	0.624	-0.03	0.00	0.02	0.052
Learning and study habits/preparation	Full Sample	4.06	-0.13	0.06	0.066	-0.19	0.00	0.01	0.046
	Subgroup:				0.998				0.000
	Female	4.16	-0.13	0.06	0.049*	-0.20	0.00	0.02	0.023
	Male	3.93	-0.13	0.08	0.121	-0.18	0.00	0.01	0.032
Motivation to go to school to learn job skills	Full Sample	3.32	-0.05	0.08	0.584	-0.07	0.00	0.01	0.002
	Subgroup:				0.476				0.060
	Female	3.38	-0.09	0.07	0.239	-0.15	0.00	0.05	0.151
	Male	3.25	0.00	0.11	0.993	0.00	0.00	0.01	0.001
Number of times absent from school	Full Sample	2.68	0.04	0.06	0.593	0.05	0.00	0.03	0.097
	Subgroup:				0.596				0.044
	Female	2.70	0.06	0.08	0.410	0.07	0.00	0.01	0.087
	Male	2.66	0.00	0.08	0.985	0.00	0.00	0.03	0.005
Number of times cut/skipped class	Full Sample	1.21	0.12	0.07	0.090	0.22	0.00	0.04	0.094
	Subgroup:				0.668				0.043
	Female	1.19	0.10	0.07	0.143	0.18	0.00	0.05	0.165
	Male	1.24	0.15	0.08	0.072	0.28	0.00	0.04	0.097

Outcome	Sample	Control Mean	Impact	SE	p-value	Effect Size	Absolute Value of Difference Compared to Model-Based Specification		
							Dif. Effect Size	Dif. SE (Effect Size)	Dif. p-value
Number of times late for school	Full Sample	1.86	-0.02	0.11	0.840	-0.02	0.00	0.09	0.065
	Subgroup:				0.261				0.137
	Female	1.80	0.04	0.09	0.708	0.05	0.00	0.14	0.143
	Male	1.94	-0.09	0.10	0.379	-0.10	0.00	0.09	0.253
Number of times received detention	Full Sample	1.79	-0.20	0.09	0.057	-0.22	0.00	0.03	0.048
	Subgroup:				0.503				0.042
	Female	1.54	-0.14	0.09	0.110	-0.18	0.00	0.03	0.113
	Male	2.08	-0.27	0.15	0.085	-0.26	0.00	0.03	0.055
Number of times sent out of class for bad behavior	Full Sample	1.42	-0.08	0.08	0.326	-0.10	0.00	0.00	0.018
	Subgroup:				0.586				0.044
	Female	1.22	-0.05	0.06	0.391	-0.10	0.00	0.04	0.138
	Male	1.66	-0.12	0.12	0.307	-0.13	0.00	0.00	0.005
School attendance and negative behaviors	Full Sample	1.80	-0.04	0.05	0.515	-0.08	0.00	0.02	0.019
	Subgroup:				0.398				0.053
	Female	1.69	0.00	0.06	0.984	0.00	0.00	0.02	0.003
	Male	1.92	-0.08	0.07	0.256	-0.14	0.00	0.00	0.017
School engagement	Full Sample	2.55	0.04	0.09	0.704	0.05	0.00	0.03	0.078
	Subgroup:				0.505				0.051
	Female	2.63	-0.01	0.07	0.888	-0.02	0.00	0.05	0.028
	Male	2.46	0.09	0.13	0.488	0.11	0.00	0.04	0.121

Note: The p-value on the subgroup rows tests for differences in impacts across subgroup categories. Effect sizes are calculated by dividing the impact estimate by the standard deviation of the outcome in the control group.
* Difference is statistically significant at the 0.05 level, two-tailed test.

Table II.F.3. Roads to Success: Evaluation Impact Estimates (Design-Based, *RCT-YES* Default Assumptions)

Outcome	Sample	Control Mean	Impact	SE	p-value	Effect Size	Absolute Value of Difference Compared to Model-Based Specification: Dif. Effect Size	Dif. SE (Effect Size)	Dif. p-value
Career exploration behavior with parents	Full Sample	3.25	-0.08	0.10	0.450	-0.08	0.00	0.02	0.066
	Subgroup:				0.361				0.057
	Female	3.41	-0.15	0.07	0.042*	-0.15	0.00	0.06	0.228
	Male	3.07	0.00	0.15	0.985	0.00	0.00	0.00	0.000
Career exploration behavior with teachers/school staff	Full Sample	2.51	0.21	0.15	0.181	0.19	0.00	0.02	0.041
	Subgroup:				0.897				0.014
	Female	2.61	0.22	0.12	0.088	0.21	0.00	0.03	0.065
	Male	2.38	0.20	0.17	0.266	0.18	0.00	0.04	0.124
Career exploration efficacy	Full Sample	3.24	0.04	0.03	0.198	0.08	0.00	0.00	0.016
	Subgroup:				0.335				0.054
	Female	3.23	0.07	0.03	0.060	0.15	0.00	0.02	0.064
	Male	3.25	0.01	0.04	0.905	0.02	0.00	0.00	0.007
Importance of grades	Full Sample	3.44	-0.10	0.07	0.177	-0.13	0.00	0.00	0.002
	Subgroup:				0.870				0.011
	Female	3.51	-0.11	0.08	0.185	-0.16	0.00	0.01	0.069
	Male	3.36	-0.09	0.08	0.271	-0.11	0.00	0.00	0.037
Knowledge of how to determine what types of jobs are a good fit	Full Sample	3.10	0.15	0.03	0.001*	0.21	0.00	0.01	0.000
	Subgroup:				0.824				0.017
	Female	3.08	0.16	0.06	0.016*	0.23	0.00	0.01	0.020
	Male	3.12	0.14	0.06	0.027*	0.18	0.00	0.03	0.024
Knowledge of how to overcome barriers to career goals	Full Sample	3.29	0.01	0.05	0.902	0.02	0.00	0.00	0.010
	Subgroup:				0.206				0.055
	Female	3.27	0.05	0.04	0.246	0.08	0.00	0.03	0.149
	Male	3.31	-0.05	0.06	0.439	-0.07	0.00	0.01	0.037

Outcome	Sample	Control Mean	Impact	SE	p-value	Effect Size	Absolute Value of Difference Compared to Model-Based Specification		
							Dif. Effect Size	Dif. SE (Effect Size)	Dif. p-value
Knowledge of requirements to succeed in different careers	Full Sample	3.13	0.02	0.05	0.644	0.03	0.00	0.06	0.139
	Subgroup:				0.389				0.077
	Female	3.10	0.06	0.05	0.237	0.10	0.00	0.07	0.273
	Male	3.15	-0.02	0.07	0.813	-0.03	0.00	0.04	0.063
Knowledge of the types of jobs that are a good fit	Full Sample	3.43	0.00	0.03	0.950	0.00	0.00	0.02	0.015
	Subgroup:				0.652				0.035
	Female	3.45	0.01	0.05	0.781	0.02	0.00	0.02	0.038
	Male	3.40	-0.02	0.04	0.624	-0.03	0.00	0.02	0.052
Learning and study habits/preparation	Full Sample	4.06	-0.13	0.06	0.066	-0.19	0.00	0.01	0.046
	Subgroup:				0.998				0.000
	Female	4.16	-0.13	0.06	0.049*	-0.20	0.00	0.02	0.023
	Male	3.93	-0.13	0.08	0.121	-0.18	0.00	0.01	0.032
Motivation to go to school to learn job skills	Full Sample	3.32	-0.05	0.08	0.584	-0.07	0.00	0.01	0.002
	Subgroup:				0.476				0.060
	Female	3.38	-0.09	0.07	0.239	-0.15	0.00	0.05	0.151
	Male	3.25	0.00	0.11	0.993	0.00	0.00	0.01	0.001
Number of times absent from school	Full Sample	2.68	0.04	0.06	0.593	0.05	0.00	0.03	0.097
	Subgroup:				0.596				0.044
	Female	2.70	0.06	0.08	0.410	0.07	0.00	0.01	0.087
	Male	2.66	0.00	0.08	0.985	0.00	0.00	0.03	0.005
Number of times cut/skipped class	Full Sample	1.21	0.12	0.07	0.090	0.22	0.00	0.04	0.094
	Subgroup:				0.668				0.043
	Female	1.19	0.10	0.07	0.143	0.18	0.00	0.05	0.165
	Male	1.24	0.15	0.08	0.072	0.28	0.00	0.04	0.097

Outcome	Sample	Control Mean	Impact	SE	p-value	Effect Size	Absolute Value of Difference Compared to Model-Based Specification		
							Dif. Effect Size	Dif. SE (Effect Size)	Dif. p-value
Number of times late for school	Full Sample	1.86	-0.02	0.11	0.840	-0.02	0.00	0.09	0.065
	Subgroup:				0.261				0.137
	Female	1.80	0.04	0.09	0.708	0.05	0.00	0.14	0.143
	Male	1.94	-0.09	0.10	0.379	-0.10	0.00	0.09	0.253
Number of times received detention	Full Sample	1.79	-0.20	0.09	0.057	-0.22	0.00	0.03	0.048
	Subgroup:				0.503				0.042
	Female	1.54	-0.14	0.09	0.110	-0.18	0.00	0.03	0.113
	Male	2.08	-0.27	0.15	0.085	-0.26	0.00	0.03	0.055
Number of times sent out of class for bad behavior	Full Sample	1.42	-0.08	0.08	0.326	-0.10	0.00	0.00	0.018
	Subgroup:				0.586				0.044
	Female	1.22	-0.05	0.06	0.391	-0.10	0.00	0.04	0.138
	Male	1.66	-0.12	0.12	0.307	-0.13	0.00	0.00	0.005
School attendance and negative behaviors	Full Sample	1.80	-0.04	0.05	0.515	-0.08	0.00	0.02	0.019
	Subgroup:				0.398				0.053
	Female	1.69	0.00	0.06	0.984	0.00	0.00	0.02	0.003
	Male	1.92	-0.08	0.07	0.256	-0.14	0.00	0.00	0.017
School engagement	Full Sample	2.55	0.04	0.09	0.704	0.05	0.00	0.03	0.078
	Subgroup:				0.505				0.051
	Female	2.63	-0.01	0.07	0.888	-0.02	0.00	0.05	0.028
	Male	2.46	0.09	0.13	0.488	0.11	0.00	0.04	0.121

Note: The p-value on the subgroup rows tests for differences in impacts across subgroup categories. Effect sizes are calculated by dividing the impact estimate by the standard deviation of the outcome in the control group.
* Difference is statistically significant at the 0.05 level, two-tailed test.

Table II.F.4. Roads to Success: Additional Information (Cluster-Level Randomization)

Outcome	Sample	Standard Deviation of Outcomes			Standard Deviation of Outcome Residuals			Intra-Cluster Correlation		
									Design-Based Assumptions	
		Full	Treatment Group	Control Group	Full	Treatment Group	Control Group	Model-Based	Matching	Default
Career exploration behavior with parents	Full Sample	1.03	1.02	1.04	1.01	1.01	1.00	0.00	0.05	0.05
	Subgroup:									
	Female	0.97	0.96	0.98	0.94	0.94	0.95			
	Male	1.08	1.07	1.09	1.07	1.08	1.06			
Career exploration behavior with teachers/school staff	Full Sample	1.12	1.08	1.10	1.05	1.03	1.07	0.03	0.10	0.10
	Subgroup:									
	Female	1.08	1.04	1.06	1.00	0.94	1.05			
	Male	1.15	1.11	1.13	1.11	1.11	1.11			
Career exploration efficacy	Full Sample	0.52	0.52	0.52	0.51	0.51	0.51	-0.01	0.01	0.01
	Subgroup:									
	Female	0.48	0.48	0.47	0.47	0.47	0.47			
	Male	0.56	0.55	0.57	0.56	0.55	0.57			
Importance of grades	Full Sample	0.74	0.73	0.75	0.73	0.73	0.73	0.00	0.03	0.03
	Subgroup:									
	Female	0.70	0.69	0.70	0.69	0.68	0.69			
	Male	0.78	0.77	0.79	0.78	0.78	0.79			
Knowledge of how to determine what types of jobs are a good fit	Full Sample	0.70	0.66	0.73	0.69	0.66	0.73	-0.01	0.00	0.00
	Subgroup:									
	Female	0.66	0.63	0.68	0.65	0.62	0.68			
	Male	0.74	0.69	0.78	0.73	0.70	0.77			
Knowledge of how to overcome barriers to career goals	Full Sample	0.64	0.63	0.65	0.63	0.63	0.64	0.00	0.02	0.02
	Subgroup:									
	Female	0.61	0.60	0.61	0.60	0.59	0.61			
	Male	0.68	0.66	0.69	0.67	0.66	0.68			

Outcome	Sample	Standard Deviation of Outcomes			Standard Deviation of Outcome Residuals			Intra-Cluster Correlation		
									Design-Based Assumptions	
		Full	Treatment Group	Control Group	Full	Treatment Group	Control Group	Model-Based	Matching	Default
Knowledge of requirements to succeed in different careers	Full Sample	0.65	0.64	0.65	0.64	0.64	0.65	0.00	0.01	0.01
	Subgroup:									
	Female	0.61	0.61	0.61	0.60	0.60	0.61			
	Male	0.68	0.67	0.70	0.68	0.67	0.69			
Knowledge of the types of jobs that are a good fit	Full Sample	0.62	0.62	0.62	0.62	0.61	0.62	-0.01	0.00	0.00
	Subgroup:									
	Female	0.60	0.60	0.61	0.60	0.58	0.61			
	Male	0.64	0.65	0.64	0.64	0.64	0.64			
Learning and study habits/preparation	Full Sample	0.70	0.70	0.70	0.69	0.69	0.68	0.01	0.02	0.02
	Subgroup:									
	Female	0.65	0.64	0.65	0.64	0.63	0.64			
	Male	0.74	0.74	0.74	0.74	0.75	0.72			
Motivation to go to school to learn job skills	Full Sample	0.74	0.78	0.70	0.73	0.78	0.69	0.01	0.04	0.04
	Subgroup:									
	Female	0.68	0.73	0.62	0.67	0.73	0.61			
	Male	0.80	0.82	0.79	0.80	0.82	0.77			
Number of times absent from school	Full Sample	0.88	0.88	0.87	0.86	0.86	0.86	0.00	0.02	0.02
	Subgroup:									
	Female	0.87	0.88	0.86	0.86	0.87	0.85			
	Male	0.88	0.87	0.89	0.86	0.86	0.87			
Number of times cut/skipped class	Full Sample	0.63	0.71	0.54	0.62	0.70	0.54	0.01	0.02	0.02
	Subgroup:									
	Female	0.60	0.65	0.55	0.60	0.64	0.55			
	Male	0.66	0.76	0.53	0.65	0.75	0.53			

		Standard Deviation of Outcomes			Standard Deviation of Outcome Residuals			Intra-Cluster Correlation		
									Design-Based Assumptions	
Outcome	Sample	Full	Treatment Group	Control Group	Full	Treatment Group	Control Group	Model-Based	Matching	Default
Number of times late for school	Full Sample	0.87	0.91	0.82	0.85	0.89	0.81	0.03	0.05	0.05
	Subgroup:									
	Female	0.82	0.90	0.74	0.80	0.87	0.73			
	Male	0.91	0.93	0.90	0.91	0.91	0.90			
Number of times received detention	Full Sample	0.89	0.82	0.93	0.83	0.77	0.88	0.01	0.05	0.05
	Subgroup:									
	Female	0.71	0.63	0.77	0.70	0.62	0.76			
	Male	0.99	0.94	1.02	0.95	0.90	1.00			
Number of times sent out of class for bad behavior	Full Sample	0.77	0.76	0.77	0.73	0.73	0.73	0.01	0.03	0.03
	Subgroup:									
	Female	0.53	0.55	0.50	0.53	0.55	0.51			
	Male	0.91	0.88	0.95	0.91	0.89	0.93			
School attendance and negative behaviors	Full Sample	0.54	0.56	0.52	0.53	0.55	0.51	0.01	0.02	0.02
	Subgroup:									
	Female	0.47	0.49	0.45	0.47	0.49	0.45			
	Male	0.60	0.62	0.57	0.59	0.61	0.57			
School engagement	Full Sample	0.78	0.81	0.74	0.77	0.81	0.72	0.01	0.03	0.03
	Subgroup:									
	Female	0.71	0.76	0.66	0.71	0.76	0.65			
	Male	0.84	0.86	0.81	0.83	0.87	0.79			

Note: The residuals are calculated using all covariates included in the model. All calculations include study weights, if applicable.

II.G. Teach For America and Teaching Fellows Programs—Clustered, Non-Blocked

Original study. Clark et al. (2013).

Sponsor agency. Institute of Education Sciences, U.S. Department of Education.

Description of intervention. The original study evaluated both Teach For America and the Teaching Fellows programs. Through a highly selective process, both programs recruit individuals who have no previous teaching experience but have demonstrated a high level of achievement. The study focused specifically on secondary math teachers who had been recruited through these processes. While these programs share some similarities, they also differ in important ways, including the required commitment and the types of people recruited.

Randomization design. Non-clustered, blocked. Students who enrolled in the same math class were randomly assigned to (1) a Teach For America teacher, (2) a Teaching Fellows teacher, or (3) a comparable class taught by a control teacher. In most cases, there were two possible classrooms to which a participant could be assigned.

Model-based method. Ordinary least squares, robust cluster standard errors.

We use OLS to estimate the following equation:

$$(19)\quad Y_{ij} = \alpha + \beta X_{ij} + \delta T_j + \varepsilon_{ij},$$

where Y_{ij} is the outcome of interest for person i assigned to teacher j, T_j is a treatment indicator, X_{ij} is a vector of covariates, α is an intercept that is assumed to be fixed, and ε_{ij} is an error term. We allow for ε_{ij} to be arbitrarily correlated at the teacher level but independent across observations, as in the original study. We estimate separate models for the Teach For America and Teaching Fellows samples. Note that we model this as a clustered, blocked design as well. Section II.J provides details on a second re-estimation of the study that includes block fixed effects.

Design-based specification: RCT-YES model. Clustered, non-blocked. For this set of estimates, we ignore the blocks that were part of the original design. This was done in order to increase the number of clustered, non-blocked designs included in the study.

Design-based specification: RCT-YES parameter (matched settings). Finite population (ATE).

Design-based specification: RCT-YES parameter (default settings). Finite population (ATE).

Covariates. We adjust for baseline math score (z-score units), baseline reading score (z-score units), old-for-grade status, grade below modal grade in classroom match status, grade above modal grade in classroom match status, student retained in same grade between previous and current year status, gender, race, free lunch status, LEP status, IEP status, and time since baseline math test. This is consistent with the approach in the original report.

Weights. Yes. We use the sample weights specified in the original report, which account for differential probabilities of assignment into the treatment and control groups.

Multiple hypothesis correction. None. We do not adjust for multiple hypothesis testing. The original study adjusted for multiple hypotheses across subgroup analyses and samples, and this approach is not currently supported by *RCT-YES*.

Main differences from original analyses. Our methods differ in several ways from the original analysis:

- The original analysis imputed missing covariates by replacing them with the average value of the research sample in the same classroom match that had non-missing values of the covariate. When the covariate was missing for all students in a classroom matched pair, the value of the covariate was replaced with the sample average. The main analysis included indicators for whether the covariate was missing. In contrast, we impute the values of missing covariates by using the default rule in *RCT-YES*.

- We do not adjust for multiple hypothesis testing, because the original study adjusts only for multiple hypotheses for estimations across subgroups, a feature that is not currently supported by *RCT-YES*.

- For this set of estimates, we ignore the blocks in the original design. Section II.J provides details on a second re-estimation of the study that includes block fixed effects.

Outcomes. We estimate the model for the following outcomes:

Post-test math *z*-score: End-of-year math score on state/NWEA assessment, in *z*-score units. We present separate outcome measures for the (1) TFA sample and (2) TNTP sample.

Table II.G.1. Teach for America and Teaching Fellows Programs—Clustered, Non-Blocked: Evaluation Impact Estimates (Model-Based)

Outcome	Sample	Control Mean	Impact	SE	p-value	Effect Size
Post-test math z-score, TFA sample	Full Sample	-0.60	0.07	0.06	0.220	0.08
	Subgroup:				0.829	
	Male	-0.61	0.07	0.06	0.254	0.07
	Female	-0.58	0.08	0.06	0.232	0.09
Post-test math z-score, TNTP sample	Full Sample	-0.39	0.00	0.07	0.956	0.00
	Subgroup:				0.695	
	Male	-0.31	0.01	0.07	0.866	0.01
	Female	-0.45	-0.01	0.07	0.940	-0.01

Note: The p-value on the subgroup rows tests for differences in impacts across subgroup categories. Effect sizes are calculated by dividing the impact estimate by the standard deviation of the outcome in the control group.
* Difference is statistically significant at the 0.05 level, two-tailed test.

Table II.G.2. Teach for America and Teaching Fellows Programs—Clustered, Non-Blocked: Evaluation Impact Estimates (Design-Based, Matching Assumptions)

							Absolute Value of Difference Compared to Model-Based Specification		
Outcome	Sample	Control Mean	Impact	SE	p-value	Effect Size	Dif. Effect Size	Dif. SE (Effect Size)	Dif. p-value
Post-test math z-score, TFA sample	Full Sample	-0.60	0.07	0.06	0.241	0.08	0.00	0.00	0.021
	Subgroup:				0.825				0.004
	Male	-0.61	0.07	0.06	0.257	0.07	0.00	0.00	0.003
	Female	-0.58	0.08	0.07	0.235	0.09	0.00	0.01	0.003
Post-test math z-score, TNTP sample	Full Sample	-0.39	0.00	0.07	0.958	0.00	0.00	0.00	0.002
	Subgroup:				0.714				0.019
	Male	-0.31	0.01	0.07	0.869	0.01	0.00	0.00	0.003
	Female	-0.45	-0.01	0.07	0.942	-0.01	0.00	0.00	0.002

Note: The p-value on the subgroup rows tests for differences in impacts across subgroup categories. Effect sizes are calculated by dividing the impact estimate by the standard deviation of the outcome in the control group.
* Difference is statistically significant at the 0.05 level, two-tailed test.

Table II.G.3. Teach for America and Teaching Fellows Programs—Clustered, Non-Blocked: Evaluation Impact Estimates (Design-Based, *RCT-YES* Default Assumptions)

							Absolute Value of Difference Compared to Model-Based Specification		
Outcome	Sample	Control Mean	Impact	SE	p-value	Effect Size	Dif. Effect Size	Dif. SE (Effect Size)	Dif. p-value
Post-test math z-score, TFA sample	Full Sample	-0.60	0.07	0.06	0.241	0.08	0.00	0.00	0.021
	Subgroup:				0.825				0.004
	Male	-0.61	0.07	0.06	0.257	0.07	0.00	0.00	0.003
	Female	-0.58	0.08	0.07	0.235	0.09	0.00	0.01	0.003
Post-test math z-score, TNTP sample	Full Sample	-0.39	0.00	0.07	0.958	0.00	0.00	0.00	0.002
	Subgroup:				0.714				0.019
	Male	-0.31	0.01	0.07	0.869	0.01	0.00	0.00	0.003
	Female	-0.45	-0.01	0.07	0.942	-0.01	0.00	0.00	0.002

Note: The p-value on the subgroup rows tests for differences in impacts across subgroup categories. Effect sizes are calculated by dividing the impact estimate by the standard deviation of the outcome in the control group.
* Difference is statistically significant at the 0.05 level, two-tailed test.

Table II.G.4. Teach for America and Teaching Fellows Programs—Clustered, Non-Blocked: Additional Information (Cluster-Level Randomization)

		Standard Deviation of Outcomes			Standard Deviation of Outcome Residuals			Intra-Cluster Correlation		
									Design-Based Assumptions	
Outcome	Sample	Full	Treatment Group	Control Group	Full	Treatment Group	Control Group	Model-Based	Matching	Default
Post-test math z-score, TFA sample	Full Sample	0.93	0.95	0.91	0.68	0.70	0.66	0.14	0.35	0.35
	Subgroup:									
	Male	0.95	0.97	0.94	0.71	0.73	0.68			
	Female	0.90	0.92	0.88	0.64	0.65	0.63			
Post-test math z-score, TNTP sample	Full Sample	1.07	1.12	1.02	0.70	0.71	0.69	0.21	0.93	0.93
	Subgroup:									
	Male	1.09	1.16	1.01	0.70	0.73	0.67			
	Female	1.05	1.08	1.03	0.70	0.69	0.70			

Note: The residuals are calculated using all covariates included in the model. All calculations include study weights, if applicable.

II.H. Health Teacher

Original study. Goesling et al. (2014).

Sponsor agency. Office of Adolescent Health, Department of Health & Human Services.

Description of intervention. The original study evaluated *HealthTeacher*, a curriculum that is designed for children in grades K–12 and that is accessible to teachers online through a subscription. The curriculum consists of lesson plans and materials that teachers can download and use in their classrooms. The evaluation focused on seventh grade students.

Randomization design. Clustered, blocked. Seventeen schools were grouped into matched pairs and randomized into a treatment group and a control group. Fourteen of the original 17 schools were used in the evaluation.

Model-based method. Ordinary least squares, robust cluster standard errors.

We use OLS to estimate the following equation:

$$(20)\quad Y_{ij} = \alpha + \beta X_{ij} + \delta T_j + \varepsilon_{ij},$$

where Y_{ij} is the outcome of interest for individual i in cluster j, X_{ij} is a vector of covariates, T_j is a treatment indicator, and ε_{ij} is an error term. As in the original study, we allow for ε_{ij} to be arbitrarily correlated within clusters but uncorrelated across clusters.

Design-based specification: RCT-YES model. Clustered, non-blocked. Note that we do not include block effects, even though the original study was a blocked design. We do not include the block effects because there are too few clusters to include both block effects and covariates in this model.

Design-based specification: RCT-YES parameter (matched settings). Finite population (ATE).

Design-based specification: RCT-YES parameter (default settings). Finite population (ATE).

Covariates. In our analysis, we control for student age, race, and gender. A baseline measure of the outcome variable is included when available. This is consistent with the approach in the original report.

Weights. Yes. We use the weights specified in the original study to account for nonresponse.

Multiple hypothesis correction. None. We do not apply a multiple hypothesis correction, because one was not applied to the main results in the original study.

Main differences from original analyses.

- The original study included indicators for whether covariates were imputed, which we do not include in our specification.

- The original study adopted a three-step estimation approach. First, the outcomes were regressed on the covariates. Second, the difference in residuals between the treatment and control groups was calculated for each block. Third, the differences were averaged across blocks. We adopt a regression approach that is more consistent with the model-based methods used by other studies so that the results are more comparable across studies. Additionally, due to the low number of clusters we do not include block effects. However, the results are similar if we include block effects and exclude covariates.

Outcomes. We estimate the model for the following outcomes:

Knowledge of contraceptive methods and STD transmission: Average of responses to three knowledge questions of contraceptive methods and STD transmission, with added weight given to student's confidence in his or her response. Scores ranged from −3 to +3, with higher values indicating more confidence in a correct response and lower values representing more confidence in an incorrect response. Follow-up 1 and follow-up 2 were treated as separate outcomes.

General knowledge of teen pregnancy, STDs, and HIV: Average of responses to five knowledge questions of teen pregnancy, STDs, and HIV. Scores ranged from 5 to 25, with higher values indicating greater knowledge. Follow-up 1 and follow-up 2 were treated as separate outcomes.

Knowledge of condoms and risk of pregnancy: Respondent correctly answered question on condoms and risk of pregnancy. Follow-up 1 and follow-up 2 were treated as separate outcomes.

Knowledge of condoms and risk of HIV/AIDS: Respondent correctly answered question on condoms and risk of HIV/AIDS. Follow-up 1 and follow-up 2 were treated as separate outcomes.

Knowledge of birth control pills and risk of pregnancy: Respondent correctly answered the question on birth control pills and risk of pregnancy. Follow-up 1 and follow-up 2 were treated as separate outcomes.

Knowledge of birth control pills and risk of HIV/AIDS: Respondent correctly answered question on birth control pills and risk of HIV/AIDS. Follow-up 1 and follow-up 2 were treated as separate outcomes.

Knowledge of birth control pills and risk of chlamydia/gonorrhea: Respondent correctly answered question on birth control pills and risk of chlamydia and gonorrhea. Follow-up 1 and follow-up 2 were treated as separate outcomes.

Knowledge of transmission of STDs through oral sex: Respondent correctly answered question on transmission of STDs through oral sex. Follow-up 1 and follow-up 2 were treated as separate outcomes.

Perceived refusal skills: Average of responses to two questions on perceived refusal skills on a scale of 1 to 4, with higher values indicating greater perceived skills. Follow-up 1 and follow-up 2 were treated as separate outcomes.

Views on early sexual activity: Average of responses to four questions on views of early sexual activity on a scale of 1 to 4, with higher values indicating less permissive attitudes. Follow-up 1 and follow-up 2 were treated as separate outcomes.

Views on condom use: Sum of responses to three questions on views of condom use on a scale of 0 to 3, with higher values indicating more positive attitudes. Follow-up 1 and follow-up 2 were treated as separate outcomes.

Views on birth control use: Sum of responses to four questions on views of use of birth control on a scale of 0 to 4, with higher values indicating more positive attitudes. Follow-up 1 and follow-up 2 were treated as separate outcomes.

Prevalence of sexual intercourse, follow-up 2: Respondent reported ever having had sexual intercourse at final follow-up.

Prevalence of oral sex, follow-up 2: Respondent reported ever having had oral sex at final follow-up.

Prevalence of sexual intercourse or oral sex, follow-up 2: Respondent reported ever having had sexual intercourse or oral sex at final follow-up.

Table II.H.1. Health Teacher: Evaluation Impact Estimates (Model-Based)

Outcome	Sample	Control Mean	Impact	SE	p-value	Effect Size
General knowledge of teen pregnancy, STDs, and HIV, follow-up 1	Full Sample	17.50	0.30	0.20	0.152	0.12
	Subgroup:				0.173	
	Male	17.56	0.17	0.26	0.530	0.07
	Female	17.45	0.48	0.19	0.027*	0.21
General knowledge of teen pregnancy, STDs, and HIV, follow-up 2	Full Sample	17.86	0.22	0.29	0.458	0.08
	Subgroup:				0.347	
	Male	17.81	0.13	0.34	0.714	0.05
	Female	17.91	0.38	0.27	0.180	0.15
Knowledge of birth control pills and risk of HIV/AIDS, follow-up 1	Full Sample	26.65	-0.54	3.57	0.881	-0.01
	Subgroup:				0.766	
	Male	21.36	0.94	4.46	0.836	0.02
	Female	32.09	-1.09	5.37	0.842	-0.02
Knowledge of birth control pills and risk of HIV/AIDS, follow-up 2	Full Sample	24.72	4.77	3.10	0.148	0.11
	Subgroup:				0.046*	
	Male	20.67	1.86	2.59	0.486	0.05
	Female	28.89	8.92	3.78	0.035*	0.20
Knowledge of birth control pills and risk of chlamydia/gonorrhea, follow-up 1	Full Sample	19.46	0.49	3.29	0.884	0.01
	Subgroup:				0.166	
	Male	18.18	-1.93	3.85	0.625	-0.05
	Female	20.76	4.14	3.66	0.279	0.10
Knowledge of birth control pills and risk of chlamydia/gonorrhea, follow-up 2	Full Sample	21.58	1.30	2.58	0.624	0.03
	Subgroup:				0.215	
	Male	19.18	-1.37	2.71	0.622	-0.03
	Female	24.04	5.07	4.13	0.242	0.12
Knowledge of birth control pills and risk of pregnancy, follow-up 1	Full Sample	34.70	0.73	2.92	0.807	0.02
	Subgroup:				0.191	
	Male	39.79	-3.69	4.66	0.444	-0.08
	Female	29.61	4.77	3.77	0.228	0.10
Knowledge of birth control pills and risk of pregnancy, follow-up 2	Full Sample	32.15	5.41	4.59	0.260	0.12
	Subgroup:				0.674	
	Male	35.16	4.61	4.94	0.368	0.10
	Female	29.10	6.87	4.97	0.190	0.15
Knowledge of condoms and risk of HIV/AIDS, follow-up 1	Full Sample	23.79	2.99	2.99	0.335	0.07
	Subgroup:				0.072	
	Male	32.64	-2.75	4.89	0.583	-0.06
	Female	14.82	8.13	3.14	0.022*	0.23
Knowledge of condoms and risk of HIV/AIDS, follow-up 2	Full Sample	26.13	2.11	4.29	0.631	0.05
	Subgroup:				0.496	
	Male	30.90	4.76	5.69	0.418	0.10
	Female	21.24	0.21	4.87	0.966	0.01

Outcome	Sample	Control Mean	Impact	SE	p-value	Effect Size
Knowledge of condoms and risk of pregnancy, follow-up 1	Full Sample	34.67	7.46	2.90	0.023*	0.16
	Subgroup:				0.724	
	Male	45.97	8.64	4.82	0.096	0.17
	Female	23.20	6.67	3.07	0.049*	0.16
Knowledge of condoms and risk of pregnancy, follow-up 2	Full Sample	42.23	0.88	3.36	0.797	0.02
	Subgroup:				0.631	
	Male	50.27	0.01	4.51	0.998	0.00
	Female	33.99	2.53	3.89	0.527	0.05
Knowledge of contraceptive methods and STD transmission, follow-up 1	Full Sample	0.30	0.19	0.09	0.057	0.16
	Subgroup:				0.406	
	Male	0.47	0.23	0.12	0.076	0.18
	Female	0.13	0.14	0.08	0.094	0.12
Knowledge of contraceptive methods and STD transmission, follow-up 2	Full Sample	0.39	0.16	0.11	0.151	0.13
	Subgroup:				0.541	
	Male	0.50	0.24	0.16	0.165	0.18
	Female	0.27	0.12	0.11	0.278	0.10
Knowledge of transmission of STDs through oral sex, follow-up 1	Full Sample	51.18	12.40	3.45	0.003*	0.25
	Subgroup:				0.060	
	Male	48.87	19.05	4.02	0.000*	0.38
	Female	53.53	6.41	5.00	0.222	0.13
Knowledge of transmission of STDs through oral sex, follow-up 2	Full Sample	56.80	1.21	5.91	0.841	0.02
	Subgroup:				0.043*	
	Male	50.89	8.32	7.00	0.256	0.17
	Female	62.85	-5.10	6.26	0.430	-0.11
Perceived refusal skills, follow-up 1	Full Sample	3.16	0.00	0.07	0.984	0.00
	Subgroup:				0.357	
	Male	2.73	0.06	0.07	0.424	0.06
	Female	3.60	-0.06	0.12	0.626	-0.07
Perceived refusal skills, follow-up 2	Full Sample	3.21	-0.05	0.08	0.549	-0.05
	Subgroup:				0.932	
	Male	2.75	-0.04	0.11	0.703	-0.04
	Female	3.66	-0.05	0.06	0.430	-0.07
Prevalence of oral sex, follow-up 2	Full Sample	8.85	3.21	1.74	0.089	0.11
	Subgroup:				0.453	
	Male	11.13	4.31	2.52	0.111	0.14
	Female	6.45	2.56	1.68	0.151	0.10
Prevalence of sexual intercourse or oral sex, follow-up 2	Full Sample	13.95	-0.32	1.70	0.853	-0.01
	Subgroup:				0.868	
	Male	15.74	0.45	2.70	0.871	0.01
	Female	12.08	-0.20	2.50	0.938	-0.01

Outcome	Sample	Control Mean	Impact	SE	p-value	Effect Size
Prevalence of sexual intercourse, follow-up 2	Full Sample	11.65	-0.28	1.50	0.856	-0.01
	Subgroup:				0.689	
	Male	13.99	-0.88	2.59	0.739	-0.03
	Female	9.20	0.86	2.79	0.762	0.03
Views on birth control use, follow-up 1	Full Sample	1.92	-0.04	0.10	0.720	-0.03
	Subgroup:				0.773	
	Male	1.83	-0.02	0.14	0.871	-0.02
	Female	2.02	-0.06	0.09	0.537	-0.05
Views on birth control use, follow-up 2	Full Sample	1.80	0.08	0.10	0.430	0.07
	Subgroup:				0.368	
	Male	1.66	0.11	0.12	0.346	0.09
	Female	1.94	0.03	0.11	0.784	0.03
Views on condom use, follow-up 1	Full Sample	2.30	-0.02	0.06	0.792	-0.02
	Subgroup:				0.542	
	Male	2.35	0.02	0.07	0.765	0.03
	Female	2.24	-0.05	0.09	0.627	-0.06
Views on condom use, follow-up 2	Full Sample	2.36	0.01	0.07	0.841	0.01
	Subgroup:				0.918	
	Male	2.39	0.01	0.07	0.847	0.01
	Female	2.33	0.02	0.09	0.788	0.02
Views on early sexual activity, follow-up 1	Full Sample	3.17	0.05	0.04	0.280	0.09
	Subgroup:				0.286	
	Male	3.02	0.01	0.05	0.809	0.02
	Female	3.33	0.09	0.06	0.168	0.18
Views on early sexual activity, follow-up 2	Full Sample	3.16	-0.04	0.05	0.483	-0.07
	Subgroup:				0.009*	
	Male	3.04	-0.14	0.06	0.039*	-0.23
	Female	3.29	0.07	0.07	0.308	0.13

Note: The p-value on the subgroup rows tests for differences in impacts across subgroup categories. Effect sizes are calculated by dividing the impact estimate by the standard deviation of the outcome in the control group.
* Difference is statistically significant at the 0.05 level, two-tailed test.

Table II.H.2. Health Teacher: Evaluation Impact Estimates (Design-Based, Matching Assumptions)

Outcome	Sample	Control Mean	Impact	SE	p-value	Effect Size	Absolute Value of Difference Compared to Model-Based Specification		
							Dif. Effect Size	Dif. SE (Effect Size)	Dif. p-value
General knowledge of teen pregnancy, STDs, and HIV, follow-up 1	Full Sample	17.50	0.30	0.25	0.263	0.12	0.00	0.02	0.111
	Subgroup:				0.232				0.059
	Male	17.56	0.14	0.28	0.624	0.06	0.01	0.01	0.094
	Female	17.45	0.45	0.20	0.046*	0.19	0.01	0.00	0.019
General knowledge of teen pregnancy, STDs, and HIV, follow-up 2	Full Sample	17.86	0.22	0.36	0.562	0.08	0.00	0.03	0.104
	Subgroup:				0.454				0.107
	Male	17.81	0.10	0.37	0.800	0.04	0.01	0.01	0.086
	Female	17.91	0.34	0.32	0.307	0.14	0.02	0.02	0.127
Knowledge of birth control pills and risk of HIV/AIDS, follow-up 1	Full Sample	26.65	-0.54	4.78	0.913	-0.01	0.00	0.03	0.032
	Subgroup:				0.800				0.034
	Male	21.36	0.47	4.88	0.925	0.01	0.01	0.01	0.089
	Female	32.09	-1.55	6.12	0.806	-0.03	0.01	0.02	0.036
Knowledge of birth control pills and risk of HIV/AIDS, follow-up 2	Full Sample	24.72	4.77	4.06	0.279	0.11	0.00	0.02	0.131
	Subgroup:				0.100				0.054
	Male	20.67	1.26	3.13	0.696	0.03	0.01	0.01	0.210
	Female	28.89	8.25	4.40	0.091	0.18	0.01	0.01	0.056
Knowledge of birth control pills and risk of chlamydia/gonorrhea, follow-up 1	Full Sample	19.46	0.49	4.51	0.917	0.01	0.00	0.03	0.033
	Subgroup:				0.216				0.050
	Male	18.18	-2.59	4.25	0.557	-0.07	0.02	0.01	0.068
	Female	20.76	3.52	4.02	0.402	0.09	0.02	0.01	0.123
Knowledge of birth control pills and risk of chlamydia/gonorrhea, follow-up 2	Full Sample	21.58	1.30	3.43	0.717	0.03	0.00	0.02	0.093
	Subgroup:				0.281				0.066
	Male	19.18	-1.90	3.10	0.554	-0.05	0.01	0.01	0.068
	Female	24.04	4.45	4.74	0.371	0.10	0.01	0.01	0.129

Outcome	Sample	Control Mean	Impact	SE	p-value	Effect Size	Absolute Value of Difference Compared to Model-Based Specification		
							Dif. Effect Size	Dif. SE (Effect Size)	Dif. p-value
Knowledge of birth control pills and risk of pregnancy, follow-up 1	Full Sample	34.70	0.73	3.69	0.849	0.02	0.00	0.02	0.042
	Subgroup:				0.239				0.048
	Male	39.79	-3.56	4.85	0.480	-0.07	0.00	0.00	0.036
	Female	29.61	4.90	4.23	0.274	0.11	0.00	0.01	0.046
Knowledge of birth control pills and risk of pregnancy, follow-up 2	Full Sample	32.15	5.41	6.15	0.408	0.12	0.00	0.03	0.148
	Subgroup:				0.735				0.061
	Male	35.16	4.32	5.67	0.464	0.09	0.01	0.02	0.096
	Female	29.10	6.49	6.09	0.312	0.14	0.01	0.02	0.122
Knowledge of condoms and risk of HIV/AIDS, follow-up 1	Full Sample	23.79	2.99	4.01	0.481	0.07	0.00	0.02	0.146
	Subgroup:				0.123				0.051
	Male	32.64	-2.48	5.45	0.660	-0.05	0.01	0.01	0.077
	Female	14.82	8.38	3.53	0.039*	0.24	0.01	0.01	0.017
Knowledge of condoms and risk of HIV/AIDS, follow-up 2	Full Sample	26.13	2.11	5.53	0.714	0.05	0.00	0.03	0.083
	Subgroup:				0.532				0.036
	Male	30.90	4.42	6.04	0.481	0.10	0.01	0.01	0.063
	Female	21.24	-0.21	5.47	0.971	-0.01	0.01	0.01	0.005
Knowledge of condoms and risk of pregnancy, follow-up 1	Full Sample	34.67	7.46	3.98	0.103	0.16	0.00	0.02	0.080
	Subgroup:				0.757				0.033
	Male	45.97	8.44	5.50	0.156	0.17	0.00	0.01	0.060
	Female	23.20	6.48	3.14	0.066	0.15	0.00	0.00	0.017
Knowledge of condoms and risk of pregnancy, follow-up 2	Full Sample	42.23	0.88	4.57	0.852	0.02	0.00	0.02	0.055
	Subgroup:				0.678				0.047
	Male	50.27	-0.33	5.11	0.949	-0.01	0.01	0.01	0.049
	Female	33.99	2.10	4.25	0.632	0.04	0.01	0.01	0.105

Outcome	Sample	Control Mean	Impact	SE	p-value	Effect Size	Absolute Value of Difference Compared to Model-Based Specification: Dif. Effect Size	Dif. SE (Effect Size)	Dif. p-value
Knowledge of contraceptive methods and STD transmission, follow-up 1	Full Sample	0.30	0.19	0.12	0.167	0.16	0.00	0.02	0.110
	Subgroup:				0.497				0.091
	Male	0.47	0.23	0.14	0.125	0.18	0.00	0.02	0.049
	Female	0.13	0.14	0.09	0.157	0.12	0.00	0.01	0.063
Knowledge of contraceptive methods and STD transmission, follow-up 2	Full Sample	0.39	0.16	0.15	0.295	0.13	0.00	0.03	0.144
	Subgroup:				0.577				0.036
	Male	0.50	0.22	0.19	0.263	0.17	0.02	0.02	0.098
	Female	0.27	0.11	0.12	0.391	0.09	0.01	0.01	0.113
Knowledge of transmission of STDs through oral sex, follow-up 1	Full Sample	51.18	12.40	4.77	0.036*	0.25	0.00	0.03	0.033
	Subgroup:				0.110				0.050
	Male	48.87	18.76	4.86	0.003*	0.37	0.01	0.02	0.003
	Female	53.53	6.13	5.67	0.305	0.12	0.01	0.01	0.083
Knowledge of transmission of STDs through oral sex, follow-up 2	Full Sample	56.80	1.21	8.04	0.885	0.02	0.00	0.04	0.044
	Subgroup:				0.112				0.069
	Male	50.89	7.99	8.12	0.348	0.16	0.01	0.02	0.092
	Female	62.85	-5.51	7.00	0.449	-0.11	0.01	0.02	0.019
Perceived refusal skills, follow-up 1	Full Sample	3.16	0.00	0.10	0.988	0.00	0.00	0.03	0.004
	Subgroup:				0.418				0.061
	Male	2.73	0.06	0.09	0.498	0.06	0.00	0.02	0.074
	Female	3.60	-0.06	0.13	0.654	-0.07	0.00	0.01	0.028
Perceived refusal skills, follow-up 2	Full Sample	3.21	-0.05	0.11	0.659	-0.05	0.00	0.03	0.110
	Subgroup:				0.944				0.012
	Male	2.75	-0.05	0.13	0.729	-0.05	0.01	0.02	0.026
	Female	3.66	-0.05	0.07	0.467	-0.07	0.00	0.01	0.037

							Absolute Value of Difference Compared to Model-Based Specification		
Outcome	Sample	Control Mean	Impact	SE	p-value	Effect Size	Dif. Effect Size	Dif. SE (Effect Size)	Dif. p-value
Prevalence of oral sex, follow-up 2	Full Sample	8.85	3.21	2.23	0.193	0.11	0.00	0.02	0.104
	Subgroup:				0.513				0.060
	Male	11.13	4.09	2.64	0.153	0.13	0.01	0.00	0.042
	Female	6.45	2.31	1.85	0.239	0.09	0.01	0.01	0.088
Prevalence of sexual intercourse or oral sex, follow-up 2	Full Sample	13.95	-0.32	2.24	0.890	-0.01	0.00	0.02	0.037
	Subgroup:				0.872				0.004
	Male	15.74	0.01	3.05	0.998	0.00	0.01	0.01	0.127
	Female	12.08	-0.65	2.32	0.784	-0.02	0.01	0.01	0.154
Prevalence of sexual intercourse, follow-up 2	Full Sample	11.65	-0.24	2.13	0.912	-0.01	0.00	0.02	0.056
	Subgroup:				0.831				0.142
	Male	13.99	-0.76	2.87	0.796	-0.02	0.00	0.01	0.057
	Female	9.20	0.27	3.07	0.931	0.01	0.02	0.01	0.169
Views on birth control use, follow-up 1	Full Sample	1.92	-0.04	0.12	0.764	-0.03	0.00	0.02	0.044
	Subgroup:				0.788				0.015
	Male	1.83	-0.02	0.14	0.897	-0.02	0.00	0.00	0.026
	Female	2.02	-0.06	0.10	0.606	-0.05	0.00	0.01	0.069
Views on birth control use, follow-up 2	Full Sample	1.80	0.08	0.13	0.540	0.07	0.00	0.03	0.110
	Subgroup:				0.481				0.113
	Male	1.66	0.12	0.13	0.371	0.10	0.01	0.01	0.025
	Female	1.94	0.04	0.11	0.717	0.03	0.01	0.00	0.067
Views on condom use, follow-up 1	Full Sample	2.30	-0.02	0.08	0.837	-0.02	0.00	0.02	0.045
	Subgroup:				0.563				0.021
	Male	2.35	0.02	0.08	0.831	0.03	0.00	0.01	0.066
	Female	2.24	-0.05	0.10	0.602	-0.06	0.00	0.01	0.025

Outcome	Sample	Control Mean	Impact	SE	p-value	Effect Size	Absolute Value of Difference Compared to Model-Based Specification: Dif. Effect Size	Dif. SE (Effect Size)	Dif. p-value
Views on condom use, follow-up 2	Full Sample	2.36	0.01	0.09	0.876	0.01	0.00	0.02	0.035
	Subgroup:				0.937				0.019
	Male	2.39	0.01	0.08	0.905	0.01	0.00	0.01	0.058
	Female	2.33	0.02	0.10	0.855	0.02	0.00	0.01	0.067
Views on early sexual activity, follow-up 1	Full Sample	3.17	0.05	0.05	0.389	0.09	0.00	0.02	0.109
	Subgroup:				0.314				0.028
	Male	3.02	0.01	0.05	0.824	0.02	0.00	0.00	0.015
	Female	3.33	0.09	0.07	0.220	0.18	0.00	0.02	0.052
Views on early sexual activity, follow-up 2	Full Sample	3.16	-0.04	0.07	0.587	-0.07	0.00	0.03	0.104
	Subgroup:				0.019*				0.010
	Male	3.04	-0.15	0.07	0.055	-0.24	0.02	0.02	0.016
	Female	3.29	0.07	0.07	0.352	0.13	0.00	0.00	0.044

Note: The p-value on the subgroup rows tests for differences in impacts across subgroup categories. Effect sizes are calculated by dividing the impact estimate by the standard deviation of the outcome in the control group.
* Difference is statistically significant at the 0.05 level, two-tailed test.

Table II.H.3. Health Teacher: Evaluation Impact Estimates (Design-Based, *RCT-YES* Default Assumptions)

							Absolute Value of Difference Compared to Model-Based Specification		
Outcome	Sample	Control Mean	Impact	SE	p-value	Effect Size	Dif. Effect Size	Dif. SE (Effect Size)	Dif. p-value
General knowledge of teen pregnancy, STDs, and HIV, follow-up 1	Full Sample	17.50	0.30	0.25	0.263	0.12	0.00	0.02	0.111
	Subgroup:				0.232				0.059
	Male	17.56	0.14	0.28	0.624	0.06	0.01	0.01	0.094
	Female	17.45	0.45	0.20	0.046*	0.19	0.01	0.00	0.019
General knowledge of teen pregnancy, STDs, and HIV, follow-up 2	Full Sample	17.86	0.22	0.36	0.562	0.08	0.00	0.03	0.104
	Subgroup:				0.454				0.107
	Male	17.81	0.10	0.37	0.800	0.04	0.01	0.01	0.086
	Female	17.91	0.34	0.32	0.307	0.14	0.02	0.02	0.127
Knowledge of birth control pills and risk of HIV/AIDS, follow-up 1	Full Sample	26.65	-0.54	4.78	0.913	-0.01	0.00	0.03	0.032
	Subgroup:				0.800				0.034
	Male	21.36	0.47	4.88	0.925	0.01	0.01	0.01	0.089
	Female	32.09	-1.55	6.12	0.806	-0.03	0.01	0.02	0.036
Knowledge of birth control pills and risk of HIV/AIDS, follow-up 2	Full Sample	24.72	4.77	4.06	0.279	0.11	0.00	0.02	0.131
	Subgroup:				0.100				0.054
	Male	20.67	1.26	3.13	0.696	0.03	0.01	0.01	0.210
	Female	28.89	8.25	4.40	0.091	0.18	0.01	0.01	0.056
Knowledge of birth control pills and risk of chlamydia/gonorrhea, follow-up 1	Full Sample	19.46	0.49	4.51	0.917	0.01	0.00	0.03	0.033
	Subgroup:				0.216				0.050
	Male	18.18	-2.59	4.25	0.557	-0.07	0.02	0.01	0.068
	Female	20.76	3.52	4.02	0.402	0.09	0.02	0.01	0.123
Knowledge of birth control pills and risk of chlamydia/gonorrhea, follow-up 2	Full Sample	21.58	1.30	3.43	0.717	0.03	0.00	0.02	0.093
	Subgroup:				0.281				0.066
	Male	19.18	-1.90	3.10	0.554	-0.05	0.01	0.01	0.068
	Female	24.04	4.45	4.74	0.371	0.10	0.01	0.01	0.129

Outcome	Sample	Control Mean	Impact	SE	p-value	Effect Size	Absolute Value of Difference Compared to Model-Based Specification		
							Dif. Effect Size	Dif. SE (Effect Size)	Dif. p-value
Knowledge of birth control pills and risk of pregnancy, follow-up 1	Full Sample	34.70	0.73	3.69	0.849	0.02	0.00	0.02	0.042
	Subgroup:				0.239				0.048
	Male	39.79	-3.56	4.85	0.480	-0.07	0.00	0.00	0.036
	Female	29.61	4.90	4.23	0.274	0.11	0.00	0.01	0.046
Knowledge of birth control pills and risk of pregnancy, follow-up 2	Full Sample	32.15	5.41	6.15	0.408	0.12	0.00	0.03	0.148
	Subgroup:				0.735				0.061
	Male	35.16	4.32	5.67	0.464	0.09	0.01	0.02	0.096
	Female	29.10	6.49	6.09	0.312	0.14	0.01	0.02	0.122
Knowledge of condoms and risk of HIV/AIDS, follow-up 1	Full Sample	23.79	2.99	4.01	0.481	0.07	0.00	0.02	0.146
	Subgroup:				0.123				0.051
	Male	32.64	-2.48	5.45	0.660	-0.05	0.01	0.01	0.077
	Female	14.82	8.38	3.53	0.039*	0.24	0.01	0.01	0.017
Knowledge of condoms and risk of HIV/AIDS, follow-up 2	Full Sample	26.13	2.11	5.53	0.714	0.05	0.00	0.03	0.083
	Subgroup:				0.532				0.036
	Male	30.90	4.42	6.04	0.481	0.10	0.01	0.01	0.063
	Female	21.24	-0.21	5.47	0.971	-0.01	0.01	0.01	0.005
Knowledge of condoms and risk of pregnancy, follow-up 1	Full Sample	34.67	7.46	3.98	0.103	0.16	0.00	0.02	0.080
	Subgroup:				0.757				0.033
	Male	45.97	8.44	5.50	0.156	0.17	0.00	0.01	0.060
	Female	23.20	6.48	3.14	0.066	0.15	0.00	0.00	0.017
Knowledge of condoms and risk of pregnancy, follow-up 2	Full Sample	42.23	0.88	4.57	0.852	0.02	0.00	0.02	0.055
	Subgroup:				0.678				0.047
	Male	50.27	-0.33	5.11	0.949	-0.01	0.01	0.01	0.049
	Female	33.99	2.10	4.25	0.632	0.04	0.01	0.01	0.105

Outcome	Sample	Control Mean	Impact	SE	p-value	Effect Size	Absolute Value of Difference Compared to Model-Based Specification		
							Dif. Effect Size	Dif. SE (Effect Size)	Dif. p-value
Knowledge of contraceptive methods and STD transmission, follow-up 1	Full Sample	0.30	0.19	0.12	0.167	0.16	0.00	0.02	0.110
	Subgroup:				0.497				0.091
	Male	0.47	0.23	0.14	0.125	0.18	0.00	0.02	0.049
	Female	0.13	0.14	0.09	0.157	0.12	0.00	0.01	0.063
Knowledge of contraceptive methods and STD transmission, follow-up 2	Full Sample	0.39	0.16	0.15	0.295	0.13	0.00	0.03	0.144
	Subgroup:				0.577				0.036
	Male	0.50	0.22	0.19	0.263	0.17	0.02	0.02	0.098
	Female	0.27	0.11	0.12	0.391	0.09	0.01	0.01	0.113
Knowledge of transmission of STDs through oral sex, follow-up 1	Full Sample	51.18	12.40	4.77	0.036*	0.25	0.00	0.03	0.033
	Subgroup:				0.110				0.050
	Male	48.87	18.76	4.86	0.003*	0.37	0.01	0.02	0.003
	Female	53.53	6.13	5.67	0.305	0.12	0.01	0.01	0.083
Knowledge of transmission of STDs through oral sex, follow-up 2	Full Sample	56.80	1.21	8.04	0.885	0.02	0.00	0.04	0.044
	Subgroup:				0.112				0.069
	Male	50.89	7.99	8.12	0.348	0.16	0.01	0.02	0.092
	Female	62.85	-5.51	7.00	0.449	-0.11	0.01	0.02	0.019
Perceived refusal skills, follow-up 1	Full Sample	3.16	0.00	0.10	0.988	0.00	0.00	0.03	0.004
	Subgroup:				0.418				0.061
	Male	2.73	0.06	0.09	0.498	0.06	0.00	0.02	0.074
	Female	3.60	-0.06	0.13	0.654	-0.07	0.00	0.01	0.028
Perceived refusal skills, follow-up 2	Full Sample	3.21	-0.05	0.11	0.659	-0.05	0.00	0.03	0.110
	Subgroup:				0.944				0.012
	Male	2.75	-0.05	0.13	0.729	-0.05	0.01	0.02	0.026
	Female	3.66	-0.05	0.07	0.467	-0.07	0.00	0.01	0.037

Outcome	Sample	Control Mean	Impact	SE	p-value	Effect Size	Absolute Value of Difference Compared to Model-Based Specification		
							Dif. Effect Size	Dif. SE (Effect Size)	Dif. p-value
Prevalence of oral sex, follow-up 2	Full Sample	8.85	3.21	2.23	0.193	0.11	0.00	0.02	0.104
	Subgroup:				0.513				0.060
	Male	11.13	4.09	2.64	0.153	0.13	0.01	0.00	0.042
	Female	6.45	2.31	1.85	0.239	0.09	0.01	0.01	0.088
Prevalence of sexual intercourse or oral sex, follow-up 2	Full Sample	13.95	-0.32	2.24	0.890	-0.01	0.00	0.02	0.037
	Subgroup:				0.872				0.004
	Male	15.74	0.01	3.05	0.998	0.00	0.01	0.01	0.127
	Female	12.08	-0.65	2.32	0.784	-0.02	0.01	0.01	0.154
Prevalence of sexual intercourse, follow-up 2	Full Sample	11.65	-0.24	2.13	0.912	-0.01	0.00	0.02	0.056
	Subgroup:				0.831				0.142
	Male	13.99	-0.76	2.87	0.796	-0.02	0.00	0.01	0.057
	Female	9.20	0.27	3.07	0.931	0.01	0.02	0.01	0.169
Views on birth control use, follow-up 1	Full Sample	1.92	-0.04	0.12	0.764	-0.03	0.00	0.02	0.044
	Subgroup:				0.788				0.015
	Male	1.83	-0.02	0.14	0.897	-0.02	0.00	0.00	0.026
	Female	2.02	-0.06	0.10	0.606	-0.05	0.00	0.01	0.069
Views on birth control use, follow-up 2	Full Sample	1.80	0.08	0.13	0.540	0.07	0.00	0.03	0.110
	Subgroup:				0.481				0.113
	Male	1.66	0.12	0.13	0.371	0.10	0.01	0.01	0.025
	Female	1.94	0.04	0.11	0.717	0.03	0.01	0.00	0.067
Views on condom use, follow-up 1	Full Sample	2.30	-0.02	0.08	0.837	-0.02	0.00	0.02	0.045
	Subgroup:				0.563				0.021
	Male	2.35	0.02	0.08	0.831	0.03	0.00	0.01	0.066
	Female	2.24	-0.05	0.10	0.602	-0.06	0.00	0.01	0.025

Outcome	Sample	Control Mean	Impact	SE	p-value	Effect Size	Absolute Value of Difference Compared to Model-Based Specification		
							Dif. Effect Size	Dif. SE (Effect Size)	Dif. p-value
Views on condom use, follow-up 2	Full Sample	2.36	0.01	0.09	0.876	0.01	0.00	0.02	0.035
	Subgroup:				0.937				0.019
	Male	2.39	0.01	0.08	0.905	0.01	0.00	0.01	0.058
	Female	2.33	0.02	0.10	0.855	0.02	0.00	0.01	0.067
Views on early sexual activity, follow-up 1	Full Sample	3.17	0.05	0.05	0.389	0.09	0.00	0.02	0.109
	Subgroup:				0.314				0.028
	Male	3.02	0.01	0.05	0.824	0.02	0.00	0.00	0.015
	Female	3.33	0.09	0.07	0.220	0.18	0.00	0.02	0.052
Views on early sexual activity, follow-up 2	Full Sample	3.16	-0.04	0.07	0.587	-0.07	0.00	0.03	0.104
	Subgroup:				0.019*				0.010
	Male	3.04	-0.15	0.07	0.055	-0.24	0.02	0.02	0.016
	Female	3.29	0.07	0.07	0.352	0.13	0.00	0.00	0.044

Note: The p-value on the subgroup rows tests for differences in impacts across subgroup categories. Effect sizes are calculated by dividing the impact estimate by the standard deviation of the outcome in the control group.
* Difference is statistically significant at the 0.05 level, two-tailed test.

Table II.H.4. Health Teacher: Additional Information (Cluster-Level Randomization)

Outcome	Sample	Standard Deviation of Outcomes: Full	Standard Deviation of Outcomes: Treatment Group	Standard Deviation of Outcomes: Control Group	Standard Deviation of Outcome Residuals: Full	Standard Deviation of Outcome Residuals: Treatment Group	Standard Deviation of Outcome Residuals: Control Group	Intra-Cluster Correlation: Model-Based	Intra-Cluster Correlation: Design-Based Assumptions: Matching	Intra-Cluster Correlation: Design-Based Assumptions: Default
General knowledge of teen pregnancy, STDs, and HIV, follow-up 1	Full Sample	2.48	2.54	2.42	2.47	2.54	2.41	0.01	0.02	0.02
	Subgroup:									
	Male	2.53	2.56	2.51	2.52	2.54	2.51			
	Female	2.44	2.53	2.32	2.41	2.52	2.29			
General knowledge of teen pregnancy, STDs, and HIV, follow-up 2	Full Sample	2.63	2.58	2.68	2.61	2.57	2.65	0.04	0.03	0.03
	Subgroup:									
	Male	2.73	2.61	2.85	2.71	2.61	2.81			
	Female	2.53	2.55	2.49	2.50	2.52	2.47			
Knowledge of birth control pills and risk of HIV/AIDS, follow-up 1	Full Sample	44.14	44.07	44.25	42.25	42.11	42.43	0.00	0.01	0.01
	Subgroup:									
	Male	41.35	41.72	41.05	40.09	39.79	40.45			
	Female	46.32	45.95	46.76	44.15	44.18	44.18			
Knowledge of birth control pills and risk of HIV/AIDS, follow-up 2	Full Sample	44.47	45.63	43.18	42.57	43.98	41.16	0.01	0.00	0.00
	Subgroup:									
	Male	41.05	41.62	40.56	39.77	40.43	39.19			
	Female	46.97	48.21	45.41	44.87	46.73	42.93			
Knowledge of birth control pills and risk of chlamydia/gonorrhea, follow-up 1	Full Sample	39.88	40.17	39.62	38.87	39.19	38.57	0.00	0.01	0.01
	Subgroup:									
	Male	37.71	36.77	38.63	37.24	36.78	37.74			
	Female	41.78	42.86	40.62	40.00	41.24	38.74			
Knowledge of birth control pills and risk of chlamydia/gonorrhea, follow-up 2	Full Sample	41.61	42.06	41.17	40.91	41.22	40.62	0.00	0.00	0.00
	Subgroup:									
	Male	38.78	38.14	39.44	38.61	38.40	38.88			
	Female	43.92	44.96	42.81	42.79	43.54	42.08			

Outcome	Sample	Standard Deviation of Outcomes: Full	Standard Deviation of Outcomes: Treatment Group	Standard Deviation of Outcomes: Control Group	Standard Deviation of Outcome Residuals: Full	Standard Deviation of Outcome Residuals: Treatment Group	Standard Deviation of Outcome Residuals: Control Group	Intra-Cluster Correlation: Model-Based	Intra-Cluster Correlation: Design-Based Assumptions: Matching	Intra-Cluster Correlation: Design-Based Assumptions: Default
Knowledge of birth control pills and risk of pregnancy, follow-up 1	Full Sample	47.54	47.48	47.64	46.49	47.19	45.81	0.00	0.01	0.01
	Subgroup:									
	Male	48.42	47.73	49.03	47.06	47.25	46.96			
	Female	46.55	47.32	45.72	45.85	47.07	44.65			
Knowledge of birth control pills and risk of pregnancy, follow-up 2	Full Sample	47.48	48.15	46.75	46.61	47.39	45.85	0.01	0.02	0.02
	Subgroup:									
	Male	48.28	48.75	47.83	47.13	47.99	46.35			
	Female	46.59	47.56	45.50	46.04	46.81	45.32			
Knowledge of condoms and risk of HIV/AIDS, follow-up 1	Full Sample	43.14	43.67	42.61	42.27	43.40	41.14	0.01	0.01	0.01
	Subgroup:									
	Male	46.27	45.56	46.96	45.99	45.37	46.65			
	Female	38.87	41.56	35.58	38.03	41.26	34.45			
Knowledge of condoms and risk of HIV/AIDS, follow-up 2	Full Sample	44.08	44.22	43.97	43.10	43.10	43.14	0.02	0.03	0.03
	Subgroup:									
	Male	46.84	47.43	46.29	45.96	47.02	44.98			
	Female	40.29	39.67	40.97	39.94	38.97	40.98			
Knowledge of condoms and risk of pregnancy, follow-up 1	Full Sample	48.43	49.09	47.63	45.11	45.33	44.92	0.01	0.02	0.02
	Subgroup:									
	Male	50.04	49.96	49.92	47.94	47.65	48.29			
	Female	43.62	44.83	42.28	42.13	43.14	41.14			
Knowledge of condoms and risk of pregnancy, follow-up 2	Full Sample	49.34	49.28	49.44	46.93	46.71	47.19	0.01	0.01	0.01
	Subgroup:									
	Male	50.04	50.08	50.08	47.87	48.33	47.50			
	Female	47.36	47.34	47.45	45.91	45.16	46.76			

Outcome	Sample	Standard Deviation of Outcomes			Standard Deviation of Outcome Residuals			Intra-Cluster Correlation		
									Design-Based Assumptions	
		Full	Treatment Group	Control Group	Full	Treatment Group	Control Group	Model-Based	Matching	Default
Knowledge of contraceptive methods and STD transmission, follow-up 1	Full Sample	1.21	1.19	1.22	1.14	1.12	1.16	0.01	0.02	0.02
	Subgroup:									
	Male	1.27	1.27	1.27	1.21	1.22	1.20			
	Female	1.10	1.06	1.15	1.07	1.02	1.12			
Knowledge of contraceptive methods and STD transmission, follow-up 2	Full Sample	1.25	1.25	1.25	1.21	1.19	1.23	0.02	0.02	0.02
	Subgroup:									
	Male	1.31	1.30	1.30	1.26	1.26	1.25			
	Female	1.18	1.17	1.19	1.15	1.11	1.20			
Knowledge of transmission of STDs through oral sex, follow-up 1	Full Sample	49.50	48.26	50.03	47.74	47.13	48.38	0.00	0.01	0.01
	Subgroup:									
	Male	49.39	46.92	50.07	46.97	45.90	48.06			
	Female	49.63	49.24	49.95	48.25	48.05	48.54			
Knowledge of transmission of STDs through oral sex, follow-up 2	Full Sample	49.49	49.45	49.58	48.23	48.14	48.37	0.02	0.04	0.04
	Subgroup:									
	Male	49.82	49.31	50.08	48.05	47.28	48.86			
	Female	49.09	49.64	48.41	48.14	48.55	47.80			
Perceived refusal skills, follow-up 1	Full Sample	0.99	0.97	1.01	0.84	0.82	0.87	0.00	0.01	0.01
	Subgroup:									
	Male	0.99	0.97	1.00	0.92	0.91	0.93			
	Female	0.81	0.80	0.82	0.76	0.72	0.80			
Perceived refusal skills, follow-up 2	Full Sample	0.98	0.98	0.98	0.83	0.83	0.83	0.01	0.02	0.02
	Subgroup:									
	Male	1.02	1.04	1.00	0.98	1.00	0.96			
	Female	0.69	0.68	0.70	0.66	0.64	0.68			

Outcome	Sample	Standard Deviation of Outcomes			Standard Deviation of Outcome Residuals			Intra-Cluster Correlation		
									Design-Based Assumptions	
		Full	Treatment Group	Control Group	Full	Treatment Group	Control Group	Model-Based	Matching	Default
Prevalence of oral sex, follow-up 2	Full Sample	30.51	32.41	28.43	28.50	30.03	26.90	0.00	0.00	0.00
	Subgroup:									
	Male	34.06	36.43	31.50	30.88	32.66	29.11			
	Female	26.12	27.51	24.60	25.81	27.15	24.38			
Prevalence of sexual intercourse or oral sex, follow-up 2	Full Sample	34.52	34.40	34.67	32.23	31.62	32.84	0.00	0.00	0.00
	Subgroup:									
	Male	36.83	37.26	36.48	33.73	33.26	34.23			
	Female	31.88	31.17	32.64	30.47	29.84	31.16			
Prevalence of sexual intercourse, follow-up 2	Full Sample	31.83	31.57	32.10	29.57	29.36	29.81	0.00	0.00	0.00
	Subgroup:									
	Male	34.17	33.61	34.75	31.62	30.87	32.37			
	Female	29.18	29.45	28.96	27.33	27.80	26.89			
Views on birth control use, follow-up 1	Full Sample	1.19	1.22	1.17	1.18	1.19	1.17	0.01	0.02	0.02
	Subgroup:									
	Male	1.23	1.25	1.21	1.23	1.23	1.22			
	Female	1.15	1.18	1.13	1.14	1.16	1.11			
Views on birth control use, follow-up 2	Full Sample	1.18	1.17	1.19	1.17	1.15	1.19	0.01	0.02	0.02
	Subgroup:									
	Male	1.18	1.15	1.20	1.17	1.15	1.20			
	Female	1.17	1.18	1.17	1.17	1.16	1.17			
Views on condom use, follow-up 1	Full Sample	0.84	0.85	0.83	0.83	0.84	0.82	0.01	0.01	0.01
	Subgroup:									
	Male	0.76	0.74	0.78	0.76	0.74	0.78			
	Female	0.90	0.93	0.88	0.89	0.92	0.87			

Outcome	Sample	Standard Deviation of Outcomes			Standard Deviation of Outcome Residuals			Intra-Cluster Correlation		
									Design-Based Assumptions	
		Full	Treatment Group	Control Group	Full	Treatment Group	Control Group	Model-Based	Matching	Default
Views on condom use, follow-up 2	Full Sample	0.84	0.85	0.83	0.84	0.84	0.83	0.02	0.01	0.01
	Subgroup:									
	Male	0.82	0.82	0.82	0.82	0.82	0.83			
	Female	0.86	0.88	0.84	0.85	0.87	0.83			
Views on early sexual activity, follow-up 1	Full Sample	0.58	0.59	0.57	0.54	0.55	0.54	0.01	0.01	0.01
	Subgroup:									
	Male	0.60	0.62	0.58	0.59	0.61	0.57			
	Female	0.50	0.49	0.51	0.50	0.49	0.51			
Views on early sexual activity, follow-up 2	Full Sample	0.61	0.63	0.59	0.58	0.58	0.58	0.02	0.03	0.03
	Subgroup:									
	Male	0.64	0.65	0.62	0.63	0.64	0.61			
	Female	0.53	0.52	0.53	0.52	0.51	0.53			

Note: The residuals are calculated using all covariates included in the model. All calculations include study weights, if applicable.

II.I. POWER Through Choices—Clustered, Non-Blocked

Original study. Goesling et al. (2015).

Sponsor agency. Office of Adolescent Health, U.S. Department of Health & Human Services.

Description of intervention. The original report estimated the effects of the *POWER Through Choices* (PTC) program, which is a sexual health education curriculum designed for youths in foster care or other similar settings. The program is delivered by a trained facilitator in a classroom setting to small groups of youths. It differs from other programs in that part of the curriculum is tailored to students in the foster care setting.

Randomization design. Clustered, blocked. In the original study, 80 dormitories or living facilities were randomly assigned to either a treatment condition that offered PTC or a control group that did not. The clusters were divided into 39 separate blocks, and assignment occurred at the block level. Most of the blocks contained one treatment group and one control group.

Model-based method. Ordinary least squares, robust cluster standard errors.

We use OLS to estimate the following equation:

(21) $Y_{ij} = \alpha + \beta X_{ij} + \delta T_j + \varepsilon_{ij}$,

where Y_{ij} is the outcome of interest for individual i in cluster j, X_{ij} is a vector of covariates, T_j is a treatment indicator, α is an intercept, and ε_{ij} is an error term. As in the original study, we calculate standard errors that allow for ε_{ij} to be arbitrarily correlated within clusters but uncorrelated across clusters. Note that we model this as a clustered, blocked design as well. Section II.K provides details on a second re-estimation of the study that includes block fixed effects.

Design-based specification: RCT-YES model. Clustered, non-blocked. For this set of estimates, we ignore the blocks that were part of the original design. This was done in order to increase the number of clustered, non-blocked designs included in the study.

Design-based specification: RCT-YES parameter (matched settings). Finite population (ATE).

Design-based specification: RCT-YES parameter (default settings). Finite population (ATE).

Covariates. In our analysis, we control for gender, age, and race. A baseline measure of the outcome variable is included when available. This is consistent with the approach in the original report.

Weights. No. The original study did not use weights so we do not use them in the model-based method. For the matched settings analyses, we specify equal weights for each observation in *RCT-YES*. For the default settings analyses, we did not specify weights, so the default *RCT-YES* weight scheme is employed.

Multiple hypothesis correction. Yes. The original study adjusted for multiple hypotheses by using the Hothorn et al. (2008) procedure within outcome domains. We specify the same outcome domains and use the Benjamini-Hochberg (1995) procedure, as is also done in *RCT-YES*.

Main differences from original analyses. Our methods differ in several ways from the original analysis:

- The original study applied a logistic regression model for binary outcomes and reported mean marginal effects between the treatment and control groups. For binary outcomes, we estimate a linear specification to be more consistent with other studies.
- The original analysis included indicators for missing values. For our model-based methods, we apply the same imputation strategy as *RCT-YES* and do not include any flags for missing variables.
- The original study accounted for blocks. Section II.K provides details on a second re-estimation of the study that includes block fixed effects.

Outcomes. We estimate the model for the following outcomes:

Received info on relationships in past 12 months: Participant received any information in the past 12 months on relationships, dating, marriage, or family life.

Received info on abstinence in past 12 months: Participant received any information in the past 12 months on abstaining from sex.

Received info on birth control methods in past 12 months: Participant received any information in the past 12 months on methods of birth control.

Received info on where to get birth control in past 12 months: Participant received any information in the past 12 months on where to get birth control.

Received info on STIs in past 12 months: Participant received any information in the past 12 months on STIs.

Received info on talking to partner about sex in past 12 months: Participant received any information in the past 12 months on how to talk to his or her partner about whether to have sex or use birth control.

Received info on saying no to sex in past 12 months: Participant received any information in the past 12 months on how to say no to sex.

Number of times received info about sex from school class: Number of times participant received any information from a school class about reproductive health, pregnancy and STI prevention, and methods of protection. The survey asked youths to respond in one of the following four categories: "never," "1–3 times," "4–9 times," or "10 or more times." Numerical values were assigned as follows: "never" = 0; "1–3 times" = 2 (the midpoint value); "4–9 times" = 6.5 (the midpoint value); "10 or more times" = 10.

Number of times received info about sex from community center: Number of times participant received from a community center or youth organization any information about reproductive health, pregnancy and STI prevention, and methods of protection. Same numerical value scheme as above.

Number of times received info about sex from doctor or nurse: Number of times participant received any information from a doctor, nurse, or clinic about reproductive health, pregnancy and STI prevention, and methods of protection. Same numerical value scheme as above.

Number of times received info about sex from group home: Number of times participant received any information from a group home about reproductive health, pregnancy and STI prevention, and methods of protection from a group home. Same numerical value scheme as above.

Knowledge of reproductive anatomy and fertility: Sum of responses to four questions based on knowledge of reproductive anatomy and fertility. Scores range from 0 to 4, with higher scores indicating greater knowledge.

Knowledge of HIV and STIs: Sum of responses to seven questions based on knowledge of HIV and STIs. Scores range from 0 to 7, with higher scores indicating greater knowledge.

Knowledge of methods of protection: Sum of responses to 10 questions based on knowledge of methods of protection. Scores range from 0 to 10, with higher scores indicating greater knowledge.

Ability to find methods of protection: Participant reported feeling "very sure" he or she could find a place to obtain methods of protection.

Perceived access to condoms: Participant reported that he or she "strongly agrees" that condoms are pretty easy to get.

Perceived access to birth control other than condoms: Participant reported that he or she "strongly agrees" that birth control is pretty easy to get.

General support for methods of protection: Average of responses to six survey questions on general support for methods of protection, on a scale of 1 to 4, with higher values indicating stronger support.

Perceived barriers to methods of protection: Average of responses to five survey questions on perceived barriers to methods of protection, on a scale of 1 to 4, with higher values indicating fewer perceived barriers.

Perceived ability to communicate with partner: Average of responses to three survey questions on perceived ability to communicate with a partner, on a scale of 1 to 4, with higher values indicating greater perceived ability.

Perceived ability to plan for and avoid unprotected sex: Average of responses to four survey questions on perceived ability to plan for and avoid unprotected sex, on a scale of 1 to 4, with higher values indicating greater perceived ability.

Intentions to have sexual intercourse: Participant reported definitely intending to have sexual intercourse in the next year.

Intentions to use a condom: Participant reported definitely intending to use a condom if he or she has sexual intercourse.

Intentions to use birth control: Participant reported definitely intending to use birth control if he or she has sexual intercourse.

Table II.I.1. POWER Through Choices—Clustered, Non-Blocked: Evaluation Impact Estimates (Model-Based)

Outcome	Sample	Control Mean	Impact	SE	p-value	Effect Size
Ability to find methods of protection	Full Sample	51.05	14.75	3.65	0.000*^	0.29
	Subgroup:				0.010*	
	Female	62.26	-3.17	7.50	0.674	-0.07
	Male	47.83	19.35	3.92	0.000*	0.39
General support for methods of protection	Full Sample	3.33	0.16	0.04	0.000*^	0.31
	Subgroup:				0.066	
	Female	3.50	0.04	0.07	0.499	0.08
	Male	3.28	0.19	0.04	0.000*	0.36
Intentions to have sexual intercourse	Full Sample	68.92	1.57	2.56	0.543	0.03
	Subgroup:				0.554	
	Female	37.86	-1.25	5.23	0.812	-0.03
	Male	77.57	2.29	2.92	0.436	0.05
Intentions to use a condom	Full Sample	43.74	13.70	3.95	0.001*^	0.28
	Subgroup:				0.642	
	Female	55.77	10.89	6.21	0.083	0.22
	Male	40.33	14.42	4.64	0.003*	0.29
Intentions to use birth control	Full Sample	39.61	5.25	4.34	0.230	0.11
	Subgroup:				0.246	
	Female	58.82	-2.63	7.00	0.708	-0.05
	Male	34.25	7.28	5.02	0.151	0.15
Knowledge of HIV and STIs	Full Sample	4.50	0.82	0.15	0.000*^	0.40
	Subgroup:				0.052	
	Female	4.70	0.32	0.27	0.250	0.18
	Male	4.44	0.95	0.17	0.000*	0.44
Knowledge of methods of protection	Full Sample	6.18	1.71	0.20	0.000*^	0.66
	Subgroup:				0.022*	
	Female	6.43	0.94	0.35	0.008*	0.41
	Male	6.11	1.91	0.23	0.000*	0.72
Knowledge of reproductive anatomy and fertility	Full Sample	2.39	0.35	0.08	0.000*^	0.29
	Subgroup:				0.025*	
	Female	2.56	0.07	0.12	0.551	0.06
	Male	2.34	0.42	0.09	0.000*	0.35
Number of times received info about sex from community center	Full Sample	16.56	12.46	3.86	0.002*^	0.33
	Subgroup:				0.743	
	Female	25.23	14.87	7.96	0.066	0.34
	Male	14.01	11.85	4.42	0.009*	0.34

Outcome	Sample	Control Mean	Impact	SE	p-value	Effect Size
Number of times received info about sex from doctor or nurse	Full Sample	25.16	0.25	3.04	0.935	0.01
	Subgroup:				0.065	
	Female	42.06	-11.53	7.21	0.114	-0.23
	Male	20.27	3.23	3.19	0.315	0.08
Number of times received info about sex from group home	Full Sample	26.08	28.35	5.48	0.000*^	0.64
	Subgroup:				0.032*	
	Female	43.08	3.84	11.14	0.732	0.08
	Male	22.73	32.10	6.10	0.000*	0.76
Number of times received info about sex from school class	Full Sample	21.09	17.31	5.18	0.001*^	0.42
	Subgroup:				0.380	
	Female	18.52	10.35	7.77	0.187	0.27
	Male	21.83	19.11	6.12	0.003*	0.46
Perceived ability to communicate with partner	Full Sample	3.28	0.26	0.04	0.000*^	0.31
	Subgroup:				0.349	
	Female	3.47	0.19	0.07	0.004*	0.24
	Male	3.23	0.28	0.05	0.000*	0.32
Perceived ability to plan for and avoid unprotected sex	Full Sample	3.09	0.27	0.05	0.000*^	0.32
	Subgroup:				0.702	
	Female	3.29	0.23	0.10	0.027*	0.27
	Male	3.03	0.28	0.06	0.000*	0.33
Perceived access to birth control other than condoms	Full Sample	28.38	3.58	2.94	0.227	0.08
	Subgroup:				0.084	
	Female	43.81	-6.48	6.48	0.321	-0.13
	Male	23.80	6.18	3.20	0.057	0.14
Perceived access to condoms	Full Sample	57.85	5.84	3.62	0.111	0.12
	Subgroup:				0.221	
	Female	50.96	13.94	7.17	0.055	0.28
	Male	59.83	3.74	4.12	0.367	0.08
Perceived barriers to methods of protection	Full Sample	2.44	0.06	0.04	0.152	0.09
	Subgroup:				0.293	
	Female	2.69	0.15	0.10	0.120	0.21
	Male	2.37	0.04	0.05	0.404	0.07
Received info on STIs in past 12 months	Full Sample	65.14	23.81	3.23	0.000*^	0.50
	Subgroup:				0.253	
	Female	71.96	18.03	4.74	0.000*	0.40
	Male	63.17	25.29	3.94	0.000*	0.52
Received info on abstinence in past 12 months	Full Sample	60.33	29.02	3.52	0.000*^	0.59
	Subgroup:				0.247	
	Female	69.44	23.31	4.18	0.000*	0.50
	Male	57.68	30.49	4.31	0.000*	0.62

Outcome	Sample	Control Mean	Impact	SE	p-value	Effect Size
Received info on birth control methods in past 12 months	Full Sample	57.74	32.97	3.94	0.000*^	0.67
	Subgroup:				0.000*	
	Female	78.70	12.27	5.16	0.020*	0.30
	Male	51.62	38.31	4.50	0.000*	0.77
Received info on relationships in past 12 months	Full Sample	77.39	14.21	2.72	0.000*^	0.34
	Subgroup:				0.553	
	Female	81.48	11.28	5.37	0.039*	0.29
	Male	76.20	14.95	3.12	0.000*	0.35
Received info on saying no to sex in past 12 months	Full Sample	60.08	31.94	4.04	0.000*^	0.65
	Subgroup:				0.010*	
	Female	77.78	16.63	5.62	0.004*	0.40
	Male	54.96	35.87	4.63	0.000*	0.72
Received info on talking to partner about sex in past 12 months	Full Sample	62.97	30.56	3.18	0.000*^	0.63
	Subgroup:				0.220	
	Female	69.16	24.10	5.26	0.000*	0.52
	Male	61.19	32.21	3.75	0.000*	0.66
Received info on where to get birth control in past 12 months	Full Sample	55.46	38.97	4.17	0.000*^	0.78
	Subgroup:				0.000*	
	Female	76.64	16.96	5.23	0.002*	0.40
	Male	49.32	44.63	4.70	0.000*	0.89

Note: The p-value on the subgroup rows tests for differences in impacts across subgroup categories. Effect sizes are calculated by dividing the impact estimate by the standard deviation of the outcome in the control group.
* Difference is statistically significant at the 0.05 level, two-tailed test.
^ Difference remains statistically significant at the 0.05 level, two-tailed test, after applying the Benjamini-Hochberg correction for multiple hypothesis testing across all full sample analyses in the same domain.

Table II.I.2. POWER Through Choices—Clustered, Non-Blocked: Evaluation Impact Estimates (Design-Based, Matching Assumptions)

Outcome	Sample	Control Mean	Impact	SE	p-value	Effect Size	Absolute Value of Difference Compared to Model-Based Specification		
							Dif. Effect Size	Dif. SE (Effect Size)	Dif. p-value
Ability to find methods of protection	Full Sample	51.05	14.75	3.84	0.000*^	0.29	0.00	0.00	0.000
	Subgroup:				0.017*				0.007
	Female	62.26	-3.17	8.19	0.703	-0.07	0.00	0.01	0.029
	Male	47.83	19.35	4.11	0.000*	0.39	0.00	0.00	0.000
General support for methods of protection	Full Sample	3.33	0.16	0.04	0.000*^	0.31	0.00	0.00	0.000
	Subgroup:				0.085				0.019
	Female	3.50	0.04	0.07	0.536	0.08	0.00	0.00	0.037
	Male	3.28	0.19	0.04	0.000*	0.36	0.00	0.00	0.000
Intentions to have sexual intercourse	Full Sample	68.92	1.57	2.75	0.571	0.03	0.00	0.00	0.028
	Subgroup:				0.580				0.026
	Female	37.86	-1.25	5.57	0.825	-0.03	0.00	0.01	0.013
	Male	77.57	2.29	3.11	0.465	0.05	0.00	0.00	0.029
Intentions to use a condom	Full Sample	43.74	13.70	4.19	0.002*^	0.28	0.00	0.00	0.001
	Subgroup:				0.660				0.018
	Female	55.77	10.89	6.51	0.111	0.22	0.00	0.01	0.028
	Male	40.33	14.42	4.91	0.005*	0.29	0.00	0.01	0.002
Intentions to use birth control	Full Sample	39.61	5.25	4.57	0.254	0.11	0.00	0.00	0.024
	Subgroup:				0.272				0.026
	Female	58.82	-2.63	7.41	0.727	-0.05	0.00	0.01	0.019
	Male	34.25	7.28	5.26	0.172	0.15	0.00	0.01	0.021
Knowledge of HIV and STIs	Full Sample	4.50	0.82	0.15	0.000*^	0.40	0.00	0.00	0.000
	Subgroup:				0.071				0.019
	Female	4.70	0.32	0.30	0.304	0.18	0.00	0.02	0.054
	Male	4.44	0.95	0.17	0.000*	0.44	0.00	0.00	0.000

Outcome	Sample	Control Mean	Impact	SE	p-value	Effect Size	Absolute Value of Difference Compared to Model-Based Specification		
							Dif. Effect Size	Dif. SE (Effect Size)	Dif. p-value
Knowledge of methods of protection	Full Sample	6.18	1.71	0.20	0.000*^	0.66	0.00	0.00	0.000
	Subgroup:				0.028*				0.006
	Female	6.43	0.94	0.37	0.020*	0.41	0.00	0.01	0.012
	Male	6.11	1.91	0.23	0.000*	0.72	0.00	0.00	0.000
Knowledge of reproductive anatomy and fertility	Full Sample	2.39	0.35	0.08	0.000*^	0.29	0.00	0.00	0.000
	Subgroup:				0.033*				0.008
	Female	2.56	0.07	0.13	0.576	0.06	0.00	0.01	0.025
	Male	2.34	0.42	0.09	0.000*	0.35	0.00	0.00	0.000
Number of times received info about sex from community center	Full Sample	16.56	12.46	4.00	0.003*^	0.33	0.00	0.00	0.001
	Subgroup:				0.756				0.013
	Female	25.23	14.87	8.55	0.098	0.34	0.00	0.01	0.032
	Male	14.01	11.85	4.37	0.009*	0.34	0.00	0.00	0.000
Number of times received info about sex from doctor or nurse	Full Sample	25.16	0.25	3.22	0.938	0.01	0.00	0.00	0.003
	Subgroup:				0.080				0.015
	Female	42.06	-11.53	7.59	0.145	-0.23	0.00	0.01	0.031
	Male	20.27	3.23	3.32	0.336	0.08	0.00	0.00	0.021
Number of times received info about sex from group home	Full Sample	26.08	28.35	5.83	0.000*^	0.64	0.00	0.01	0.000
	Subgroup:				0.055				0.023
	Female	43.08	3.84	12.82	0.770	0.08	0.00	0.03	0.038
	Male	22.73	32.10	6.30	0.000*	0.76	0.00	0.00	0.000
Number of times received info about sex from school class	Full Sample	21.09	17.31	5.34	0.002*^	0.42	0.00	0.00	0.001
	Subgroup:				0.397				0.017
	Female	18.52	10.35	8.05	0.214	0.27	0.00	0.01	0.027
	Male	21.83	19.11	6.30	0.004*	0.46	0.00	0.00	0.001

Outcome	Sample	Control Mean	Impact	SE	p-value	Effect Size	Absolute Value of Difference Compared to Model-Based Specification: Dif. Effect Size	Dif. SE (Effect Size)	Dif. p-value
Perceived ability to communicate with partner	Full Sample	3.28	0.26	0.05	0.000*^	0.31	0.00	0.01	0.000
	Subgroup:				0.345				0.004
	Female	3.47	0.19	0.07	0.010*	0.24	0.00	0.00	0.006
	Male	3.23	0.28	0.05	0.000*	0.32	0.00	0.00	0.000
Perceived ability to plan for and avoid unprotected sex	Full Sample	3.09	0.27	0.05	0.000*^	0.32	0.00	0.00	0.000
	Subgroup:				0.718				0.016
	Female	3.29	0.23	0.11	0.050*	0.27	0.00	0.01	0.023
	Male	3.03	0.28	0.06	0.000*	0.33	0.00	0.00	0.000
Perceived access to birth control other than condoms	Full Sample	28.38	3.58	3.12	0.255	0.08	0.00	0.00	0.028
	Subgroup:				0.100				0.016
	Female	43.81	-6.48	6.80	0.353	-0.13	0.00	0.01	0.032
	Male	23.80	6.18	3.37	0.072	0.14	0.00	0.00	0.015
Perceived access to condoms	Full Sample	57.85	5.84	3.82	0.131	0.12	0.00	0.00	0.020
	Subgroup:				0.246				0.025
	Female	50.96	13.94	7.56	0.081	0.28	0.00	0.01	0.026
	Male	59.83	3.74	4.36	0.395	0.08	0.00	0.00	0.028
Perceived barriers to methods of protection	Full Sample	2.44	0.06	0.05	0.172	0.09	0.00	0.02	0.020
	Subgroup:				0.325				0.032
	Female	2.69	0.15	0.10	0.155	0.21	0.00	0.00	0.035
	Male	2.37	0.04	0.05	0.429	0.07	0.00	0.00	0.025
Received info on STIs in past 12 months	Full Sample	65.14	23.81	3.25	0.000*^	0.50	0.00	0.00	0.000
	Subgroup:				0.269				0.016
	Female	71.96	18.03	5.16	0.002*	0.40	0.00	0.01	0.002
	Male	63.17	25.29	3.91	0.000*	0.52	0.00	0.00	0.000

Outcome	Sample	Control Mean	Impact	SE	p-value	Effect Size	Absolute Value of Difference Compared to Model-Based Specification		
							Dif. Effect Size	Dif. SE (Effect Size)	Dif. p-value
Received info on abstinence in past 12 months	Full Sample	60.33	29.02	3.43	0.000*^	0.59	0.00	0.00	0.000
	Subgroup:				0.251				0.004
	Female	69.44	23.31	4.47	0.000*	0.50	0.00	0.01	0.000
	Male	57.68	30.49	4.19	0.000*	0.62	0.00	0.00	0.000
Received info on birth control methods in past 12 months	Full Sample	57.74	32.97	3.81	0.000*^	0.67	0.00	0.00	0.000
	Subgroup:				0.000*				0.000
	Female	78.70	12.27	5.56	0.040*	0.30	0.00	0.01	0.020
	Male	51.62	38.31	4.29	0.000*	0.77	0.00	0.00	0.000
Received info on relationships in past 12 months	Full Sample	77.39	14.21	2.77	0.000*^	0.34	0.00	0.00	0.000
	Subgroup:				0.573				0.020
	Female	81.48	11.28	5.70	0.062	0.29	0.00	0.01	0.023
	Male	76.20	14.95	3.15	0.000*	0.35	0.00	0.00	0.000
Received info on saying no to sex in past 12 months	Full Sample	60.08	31.94	3.95	0.000*^	0.65	0.00	0.00	0.000
	Subgroup:				0.010*				0.000
	Female	77.78	16.63	5.62	0.008*	0.40	0.00	0.00	0.004
	Male	54.96	35.87	4.54	0.000*	0.72	0.00	0.00	0.000
Received info on talking to partner about sex in past 12 months	Full Sample	62.97	30.56	3.05	0.000*^	0.63	0.00	0.00	0.000
	Subgroup:				0.226				0.006
	Female	69.16	24.10	5.56	0.000*	0.52	0.00	0.01	0.000
	Male	61.19	32.21	3.52	0.000*	0.66	0.00	0.00	0.000

Outcome	Sample	Control Mean	Impact	SE	p-value	Effect Size	Absolute Value of Difference Compared to Model-Based Specification		
							Dif. Effect Size	Dif. SE (Effect Size)	Dif. p-value
Received info on where to get birth control in past 12 months	Full Sample	55.46	38.97	3.93	0.000*^	0.78	0.00	0.00	0.000
	Subgroup:				0.000*				0.000
	Female	76.64	16.96	5.36	0.005*	0.40	0.00	0.00	0.003
	Male	49.32	44.63	4.30	0.000*	0.89	0.00	0.01	0.000

Note: The p-value on the subgroup rows tests for differences in impacts across subgroup categories. Effect sizes are calculated by dividing the impact estimate by the standard deviation of the outcome in the control group.
* Difference is statistically significant at the 0.05 level, two-tailed test.
^ Difference remains statistically significant at the 0.05 level, two-tailed test, after applying the Benjamini-Hochberg correction for multiple hypothesis testing across all full sample analyses in the same domain.

Table II.I.3. POWER Through Choices—Clustered, Non-Blocked: Evaluation Impact Estimates (Design-Based, *RCT-YES* Default Assumptions)

							Absolute Value of Difference Compared to Model-Based Specification		
Outcome	Sample	Control Mean	Impact	SE	p-value	Effect Size	Dif. Effect Size	Dif. SE (Effect Size)	Dif. p-value
Ability to find methods of protection	Full Sample	53.45	10.82	4.25	0.013*^	0.22	0.08	0.01	0.013
	Subgroup:				0.432				0.422
	Female	62.81	6.66	11.70	0.576	0.14	0.20	0.09	0.098
	Male	48.10	16.82	4.75	0.001*	0.34	0.05	0.02	0.001
General support for methods of protection	Full Sample	3.31	0.19	0.06	0.002*^	0.36	0.06	0.04	0.002
	Subgroup:				0.235				0.169
	Female	3.54	0.08	0.08	0.330	0.16	0.08	0.02	0.169
	Male	3.26	0.21	0.07	0.006*	0.40	0.04	0.06	0.006
Intentions to have sexual intercourse	Full Sample	63.79	2.69	3.70	0.469	0.06	0.02	0.02	0.074
	Subgroup:				0.835				0.281
	Female	41.01	2.49	8.52	0.773	0.05	0.08	0.07	0.039
	Male	69.71	4.43	4.01	0.274	0.11	0.05	0.03	0.162
Intentions to use a condom	Full Sample	48.08	9.37	5.31	0.082	0.19	0.09	0.03	0.081
	Subgroup:				0.331				0.311
	Female	61.24	-4.34	11.26	0.704	-0.09	0.31	0.10	0.621
	Male	47.72	8.08	6.50	0.219	0.16	0.13	0.04	0.216
Intentions to use birth control	Full Sample	43.30	0.79	5.31	0.882	0.02	0.09	0.02	0.652
	Subgroup:				0.535				0.289
	Female	60.39	-8.65	10.53	0.422	-0.17	0.12	0.07	0.286
	Male	41.94	-1.30	6.51	0.843	-0.03	0.18	0.03	0.692
Knowledge of HIV and STIs	Full Sample	4.27	0.94	0.20	0.000*^	0.46	0.06	0.02	0.000
	Subgroup:				0.085				0.033
	Female	4.72	0.45	0.31	0.166	0.25	0.07	0.02	0.084
	Male	4.07	1.14	0.24	0.000*	0.53	0.09	0.03	0.000

Outcome	Sample	Control Mean	Impact	SE	p-value	Effect Size	Absolute Value of Difference Compared to Model-Based Specification		
							Dif. Effect Size	Dif. SE (Effect Size)	Dif. p-value
Knowledge of methods of protection	Full Sample	5.83	1.86	0.28	0.000*^	0.72	0.06	0.03	0.000
	Subgroup:				0.140				0.118
	Female	6.31	1.32	0.40	0.004*	0.57	0.16	0.02	0.004
	Male	5.60	2.11	0.33	0.000*	0.80	0.08	0.04	0.000
Knowledge of reproductive anatomy and fertility	Full Sample	2.31	0.31	0.12	0.012*^	0.26	0.03	0.03	0.012
	Subgroup:				0.430				0.405
	Female	2.53	0.16	0.20	0.427	0.13	0.07	0.06	0.124
	Male	2.23	0.37	0.15	0.017*	0.31	0.04	0.05	0.017
Number of times received info about sex from community center	Full Sample	16.12	17.40	3.86	0.000*^	0.47	0.13	0.00	0.002
	Subgroup:				0.611				0.132
	Female	22.34	23.35	8.47	0.013*	0.54	0.19	0.01	0.053
	Male	12.32	18.17	4.53	0.000*	0.52	0.18	0.00	0.009
Number of times received info about sex from doctor or nurse	Full Sample	24.65	0.12	3.65	0.975	0.00	0.00	0.01	0.040
	Subgroup:				0.445				0.380
	Female	34.17	-2.90	9.54	0.765	-0.06	0.17	0.05	0.651
	Male	18.77	5.14	4.04	0.208	0.13	0.05	0.02	0.107
Number of times received info about sex from group home	Full Sample	27.19	29.55	6.11	0.000*^	0.67	0.03	0.01	0.000
	Subgroup:				0.008*				0.024
	Female	48.84	-7.18	14.18	0.622	-0.14	0.22	0.06	0.110
	Male	22.07	36.71	6.73	0.000*	0.87	0.11	0.02	0.000
Number of times received info about sex from school class	Full Sample	19.93	11.89	4.55	0.011*^	0.29	0.13	0.02	0.010
	Subgroup:				0.993				0.613
	Female	16.30	14.05	9.34	0.149	0.36	0.09	0.04	0.038
	Male	19.49	14.15	5.29	0.010*	0.34	0.12	0.02	0.007

Outcome	Sample	Control Mean	Impact	SE	p-value	Effect Size	Absolute Value of Difference Compared to Model-Based Specification		
							Dif. Effect Size	Dif. SE (Effect Size)	Dif. p-value
Perceived ability to communicate with partner	Full Sample	3.31	0.22	0.06	0.001*^	0.26	0.05	0.02	0.001
	Subgroup:				0.691				0.342
	Female	3.52	0.17	0.10	0.126	0.22	0.03	0.04	0.122
	Male	3.26	0.22	0.07	0.005*	0.26	0.07	0.02	0.005
Perceived ability to plan for and avoid unprotected sex	Full Sample	3.11	0.27	0.07	0.000*^	0.32	0.00	0.02	0.000
	Subgroup:				0.359				0.343
	Female	3.43	0.11	0.15	0.491	0.13	0.14	0.06	0.464
	Male	3.05	0.26	0.08	0.001*	0.31	0.02	0.02	0.001
Perceived access to birth control other than condoms	Full Sample	29.06	3.08	3.69	0.406	0.07	0.01	0.02	0.179
	Subgroup:				0.200				0.116
	Female	44.88	-6.21	8.96	0.496	-0.12	0.01	0.05	0.175
	Male	23.40	6.50	4.10	0.119	0.15	0.01	0.02	0.062
Perceived access to condoms	Full Sample	54.03	7.21	4.76	0.135	0.15	0.03	0.02	0.024
	Subgroup:				0.428				0.207
	Female	58.57	14.21	9.29	0.143	0.28	0.01	0.04	0.088
	Male	53.46	5.71	5.68	0.319	0.12	0.04	0.03	0.048
Perceived barriers to methods of protection	Full Sample	2.52	0.05	0.05	0.378	0.08	0.02	0.02	0.226
	Subgroup:				0.657				0.364
	Female	2.85	0.05	0.11	0.637	0.07	0.14	0.01	0.517
	Male	2.48	0.00	0.06	0.941	0.00	0.07	0.02	0.537
Received info on STIs in past 12 months	Full Sample	63.06	25.54	3.83	0.000*^	0.54	0.04	0.01	0.000
	Subgroup:				0.397				0.144
	Female	65.97	17.19	12.30	0.178	0.38	0.02	0.17	0.178
	Male	59.33	28.73	4.84	0.000*	0.59	0.07	0.02	0.000

Outcome	Sample	Control Mean	Impact	SE	p-value	Effect Size	Absolute Value of Difference Compared to Model-Based Specification		
							Dif. Effect Size	Dif. SE (Effect Size)	Dif. p-value
Received info on abstinence in past 12 months	Full Sample	61.30	28.39	3.67	0.000*^	0.58	0.01	0.00	0.000
	Subgroup:				0.481				0.234
	Female	64.30	21.38	11.96	0.090	0.46	0.04	0.17	0.090
	Male	57.75	30.76	4.71	0.000*	0.62	0.01	0.01	0.000
Received info on birth control methods in past 12 months	Full Sample	58.91	31.19	3.94	0.000*^	0.63	0.04	0.00	0.000
	Subgroup:				0.009*				0.009
	Female	78.57	6.87	10.17	0.507	0.17	0.13	0.12	0.487
	Male	52.09	37.27	4.77	0.000*	0.74	0.02	0.01	0.000
Received info on relationships in past 12 months	Full Sample	74.48	18.02	3.50	0.000*^	0.43	0.09	0.02	0.000
	Subgroup:				0.291				0.262
	Female	81.40	12.87	6.67	0.068	0.33	0.04	0.03	0.029
	Male	71.14	21.03	4.26	0.000*	0.49	0.14	0.03	0.000
Received info on saying no to sex in past 12 months	Full Sample	62.23	30.70	3.85	0.000*^	0.63	0.03	0.00	0.000
	Subgroup:				0.379				0.369
	Female	70.70	24.63	7.90	0.006*	0.59	0.19	0.05	0.002
	Male	59.46	32.72	4.54	0.000*	0.66	0.06	0.00	0.000
Received info on talking to partner about sex in past 12 months	Full Sample	63.56	28.57	3.58	0.000*^	0.59	0.04	0.01	0.000
	Subgroup:				0.722				0.502
	Female	64.69	28.04	7.89	0.002*	0.60	0.08	0.06	0.002
	Male	60.48	31.47	4.47	0.000*	0.64	0.02	0.01	0.000

Outcome	Sample	Control Mean	Impact	SE	p-value	Effect Size	Absolute Value of Difference Compared to Model-Based Specification: Dif. Effect Size	Dif. SE (Effect Size)	Dif. p-value
Received info on where to get birth control in past 12 months	Full Sample	56.70	36.39	3.87	0.000*^	0.73	0.05	0.01	0.000
	Subgroup:				0.059				0.059
	Female	69.42	24.63	7.87	0.006*	0.58	0.18	0.06	0.004
	Male	49.76	43.25	4.49	0.000*	0.86	0.03	0.00	0.000

Note: The p-value on the subgroup rows tests for differences in impacts across subgroup categories. Effect sizes are calculated by dividing the impact estimate by the standard deviation of the outcome in the control group.
* Difference is statistically significant at the 0.05 level, two-tailed test.
^ Difference remains statistically significant at the 0.05 level, two-tailed test, after applying the Benjamini-Hochberg correction for multiple hypothesis testing across all full sample analyses in the same domain.

Table II.I.4. POWER Through Choices—Clustered, Non-Blocked: Additional Information (Cluster-Level Randomization)

								Intra-Cluster Correlation		
		Standard Deviation of Outcomes			Standard Deviation of Outcome Residuals				Design-Based Assumptions	
Outcome	Sample	Full	Treatment Group	Control Group	Full	Treatment Group	Control Group	Model-Based	Matching	Default
Ability to find methods of protection	Full Sample	49.31	47.42	50.04	46.08	46.14	46.06	0.03	0.06	0.10
	Subgroup:									
	Female	48.80	49.18	48.70	45.67	47.74	44.16			
	Male	49.45	46.98	50.02	45.93	45.51	46.43			
General support for methods of protection	Full Sample	0.52	0.50	0.52	0.44	0.43	0.44	0.04	0.13	0.29
	Subgroup:									
	Female	0.50	0.52	0.49	0.43	0.40	0.45			
	Male	0.52	0.50	0.52	0.44	0.43	0.44			
Intentions to have sexual intercourse	Full Sample	46.19	46.09	46.33	37.38	37.54	37.26	0.01	0.28	0.28
	Subgroup:									
	Female	47.85	46.84	48.74	42.47	40.56	44.24			
	Male	41.45	41.18	41.77	35.97	36.83	35.13			
Intentions to use a condom	Full Sample	50.02	49.45	49.66	48.01	47.75	48.31	0.03	0.09	0.14
	Subgroup:									
	Female	48.96	47.41	49.91	47.01	45.71	48.29			
	Male	50.00	49.75	49.12	48.29	48.27	48.38			
Intentions to use birth control	Full Sample	49.43	49.81	48.96	48.26	48.67	47.90	0.05	0.13	0.14
	Subgroup:									
	Female	49.54	49.89	49.46	49.32	49.92	49.05			
	Male	48.65	49.48	47.52	47.96	48.38	47.60			
Knowledge of HIV and STIs	Full Sample	1.93	1.69	2.07	1.71	1.57	1.83	0.04	0.10	0.28
	Subgroup:									
	Female	1.75	1.71	1.77	1.67	1.60	1.73			
	Male	1.98	1.68	2.14	1.71	1.56	1.86			

Outcome	Sample	Standard Deviation of Outcomes			Standard Deviation of Outcome Residuals			Intra-Cluster Correlation		
									Design-Based Assumptions	
		Full	Treatment Group	Control Group	Full	Treatment Group	Control Group	Model-Based	Matching	Default
Knowledge of methods of protection	Full Sample	2.54	2.18	2.58	2.12	2.06	2.18	0.06	0.13	0.34
	Subgroup:									
	Female	2.42	2.45	2.31	2.20	2.33	2.10			
	Male	2.57	2.10	2.65	2.09	1.98	2.20			
Knowledge of reproductive anatomy and fertility	Full Sample	1.17	1.11	1.21	0.98	0.98	0.98	0.03	0.09	0.28
	Subgroup:									
	Female	1.19	1.09	1.26	1.00	0.90	1.07			
	Male	1.17	1.11	1.19	0.97	0.99	0.95			
Number of times received info about sex from community center	Full Sample	41.70	45.21	37.21	40.88	44.68	36.98	0.06	0.11	0.10
	Subgroup:									
	Female	46.67	49.34	43.64	45.70	48.89	43.34			
	Male	39.99	43.87	34.76	39.56	43.75	34.96			
Number of times received info about sex from doctor or nurse	Full Sample	43.31	43.22	43.44	42.63	42.94	42.39	0.01	0.04	0.07
	Subgroup:									
	Female	48.56	46.82	49.60	48.38	47.20	49.49			
	Male	41.28	42.28	40.26	40.93	41.85	40.06			
Number of times received info about sex from group home	Full Sample	48.72	49.96	43.96	46.35	49.77	43.31	0.08	0.16	0.15
	Subgroup:									
	Female	50.00	50.67	49.90	49.65	48.86	50.47			
	Male	48.46	49.92	41.97	45.54	49.61	41.64			
Number of times received info about sex from school class	Full Sample	45.73	48.80	40.83	44.12	47.60	40.62	0.13	0.24	0.13
	Subgroup:									
	Female	42.06	45.28	39.03	41.39	45.40	38.14			
	Male	46.49	49.31	41.37	44.79	48.10	41.32			

Outcome	Sample	Standard Deviation of Outcomes			Standard Deviation of Outcome Residuals			Intra-Cluster Correlation		
									Design-Based Assumptions	
		Full	Treatment Group	Control Group	Full	Treatment Group	Control Group	Model-Based	Matching	Default
Perceived ability to communicate with partner	Full Sample	0.76	0.63	0.85	0.67	0.53	0.78	0.00	0.04	0.08
	Subgroup:									
	Female	0.73	0.63	0.79	0.65	0.49	0.76			
	Male	0.77	0.63	0.86	0.68	0.54	0.79			
Perceived ability to plan for and avoid unprotected sex	Full Sample	0.77	0.65	0.85	0.68	0.57	0.76	0.02	0.08	0.13
	Subgroup:									
	Female	0.76	0.62	0.85	0.68	0.55	0.77			
	Male	0.77	0.66	0.84	0.67	0.57	0.76			
Perceived access to birth control other than condoms	Full Sample	45.71	46.31	45.14	42.55	43.25	41.91	0.00	0.03	0.05
	Subgroup:									
	Female	49.31	48.67	49.85	47.55	46.51	48.59			
	Male	44.29	45.71	42.64	41.08	42.43	39.71			
Perceived access to condoms	Full Sample	48.77	47.93	49.43	45.43	43.57	47.24	0.03	0.06	0.13
	Subgroup:									
	Female	49.63	48.17	50.23	47.08	46.76	47.58			
	Male	48.53	47.93	49.09	44.96	42.80	47.16			
Perceived barriers to methods of protection	Full Sample	0.60	0.55	0.65	0.52	0.49	0.55	0.03	0.14	0.17
	Subgroup:									
	Female	0.67	0.59	0.72	0.57	0.51	0.63			
	Male	0.57	0.52	0.61	0.50	0.48	0.53			
Received info on STIs in past 12 months	Full Sample	42.28	31.78	47.70	39.96	31.63	46.70	0.03	0.07	0.13
	Subgroup:									
	Female	39.79	29.06	45.13	38.04	28.35	44.53			
	Male	42.88	32.39	48.30	40.44	32.33	47.34			

Outcome	Sample	Standard Deviation of Outcomes: Full	Standard Deviation of Outcomes: Treatment Group	Standard Deviation of Outcomes: Control Group	Standard Deviation of Outcome Residuals: Full	Standard Deviation of Outcome Residuals: Treatment Group	Standard Deviation of Outcome Residuals: Control Group	Intra-Cluster Correlation: Model-Based	Intra-Cluster Correlation: Design-Based Assumptions: Matching	Intra-Cluster Correlation: Design-Based Assumptions: Default
Received info on abstinence in past 12 months	Full Sample	43.64	31.31	48.97	40.84	31.03	48.52	0.04	0.08	0.10
	Subgroup:									
	Female	40.10	25.49	46.28	37.76	24.75	45.75			
	Male	44.43	32.46	49.47	41.60	32.30	49.32			
Received info on birth control methods in past 12 months	Full Sample	44.06	29.92	49.45	40.09	30.01	47.91	0.07	0.14	0.14
	Subgroup:									
	Female	36.17	27.36	41.13	34.56	26.36	40.08			
	Male	45.49	30.49	50.04	41.01	30.24	49.69			
Received info on relationships in past 12 months	Full Sample	36.42	28.04	41.88	35.50	28.09	41.45	0.02	0.05	0.13
	Subgroup:									
	Female	34.08	25.49	39.03	33.31	25.70	38.48			
	Male	36.99	28.61	42.64	36.05	28.63	42.31			
Received info on saying no to sex in past 12 months	Full Sample	43.03	28.04	49.02	39.51	28.22	48.01	0.09	0.16	0.12
	Subgroup:									
	Female	35.16	21.06	41.77	33.94	20.85	41.71			
	Male	44.49	29.36	49.82	40.62	29.36	49.55			
Received info on talking to partner about sex in past 12 months	Full Sample	41.47	24.95	48.34	38.35	25.21	47.84	0.04	0.06	0.12
	Subgroup:									
	Female	39.79	23.41	46.40	37.42	23.38	45.90			
	Male	41.90	25.31	48.80	38.57	25.59	48.40			
Received info on where to get birth control in past 12 months	Full Sample	43.63	23.84	49.75	38.34	24.34	48.26	0.10	0.16	0.14
	Subgroup:									
	Female	36.25	23.41	42.51	34.23	22.66	41.43			
	Male	45.00	23.97	50.06	38.87	23.92	49.74			

Note: The residuals are calculated using all covariates included in the model. All calculations include study weights, if applicable.

II.J. Teach For America and Teaching Fellows Programs—Clustered, Blocked

Original study. Clark et al. (2013).

Sponsor agency. Institute of Education Sciences, U.S. Department of Education.

Description of intervention. See Section II.G.

Randomization design. Non-clustered, blocked. See Section II.G.

Model-based method. Ordinary least squares, block fixed effects, robust cluster standard errors.

We use OLS to estimate the following equation:

$$(22)\ Y_{ijk} = \alpha_k + \beta X_{ijk} + \delta T_{jk} + \varepsilon_{ijk},$$

where Y_{ijk} is the outcome of interest for person i assigned to teacher j in block k, T_{jk} is a treatment indicator, X_{ijk} is a vector of covariates, α_k is a block effect that is assumed to be fixed, and ε_{ijk} is an error term. We allow for ε_{ijk} to be arbitrarily correlated at the teacher level but independent across observations, as in the original study. We estimate separate models for the Teach For America and Teaching Fellows samples.

Design-based specification: RCT-YES model. Clustered, blocked. Note that we model this as a clustered, non-blocked design as well. Section II.G provides details on this re-estimation of the study.

Design-based specification: RCT-YES parameter (matched settings). Finite population (ATE).

Design-based specification: RCT-YES parameter (default settings). Super population (PATE).

Covariates. See Section II.G.

Weights. Yes. See Section II.G.

Multiple hypothesis correction. None. See Section II.G.

Main differences from original analyses. See Section II.G.

We randomly drop a small number of blocks because *RCT-YES* places limits on the number of blocks that may be included in a finite-population model in order to maintain degrees of freedom. We drop only as many blocks as needed to produce full sample estimates. A super-population model (PATE) is used for the sub-group analysis in order to preserve as much original study data as possible.

Outcomes. See Section II.G.

Table II.J.1. Teach for America and Teaching Fellows Programs—Clustered, Blocked: Evaluation Impact Estimates (Model-Based)

Outcome	Sample	Control Mean	Impact	SE	p-value	Effect Size
Post-test math z-score, TFA sample	Full Sample	-0.57	0.08	0.02	0.000*	0.09
	Subgroup:				0.673	
	Male	-0.59	0.07	0.03	0.023*	0.08
	Female	-0.54	0.09	0.03	0.001*	0.11
Post-test math z-score, TNTP sample	Full Sample	-0.43	-0.02	0.03	0.536	-0.02
	Subgroup:				0.719	
	Male	-0.33	-0.01	0.04	0.847	-0.01
	Female	-0.50	-0.02	0.03	0.501	-0.02

Note: The p-value on the subgroup rows tests for differences in impacts across subgroup categories. Effect sizes are calculated by dividing the impact estimate by the standard deviation of the outcome in the control group.
* Difference is statistically significant at the 0.05 level, two-tailed test.

Table II.J.2. Teach for America and Teaching Fellows Programs—Clustered, Blocked: Evaluation Impact Estimates (Design-Based, Matching Assumptions)

Outcome	Sample	Control Mean	Impact	SE	p-value	Effect Size	Absolute Value of Difference Compared to Model-Based Specification: Dif. Effect Size	Dif. SE (Effect Size)	Dif. p-value
Post-test math z-score, TFA sample	Full Sample	-0.57	0.08	0.03	0.006*	0.09	0.00	0.01	0.006
	Subgroup:				0.711				0.038
	Male	-0.58	0.07	0.04	0.085	0.08	0.00	0.01	0.062
	Female	-0.53	0.08	0.03	0.015*	0.09	0.01	0.00	0.014
Post-test math z-score, TNTP sample	Full Sample	-0.43	-0.02	0.03	0.628	-0.02	0.00	0.00	0.092
	Subgroup:				0.323				0.396
	Male	-0.33	-0.01	0.05	0.856	-0.01	0.00	0.01	0.009
	Female	-0.50	-0.03	0.04	0.508	-0.03	0.01	0.01	0.007

Note: The p-value on the subgroup rows tests for differences in impacts across subgroup categories. Effect sizes are calculated by dividing the impact estimate by the standard deviation of the outcome in the control group.
* Difference is statistically significant at the 0.05 level, two-tailed test.

Table II.J.3. Teach for America and Teaching Fellows Programs—Clustered, Blocked: Evaluation Impact Estimates (Design-Based, *RCT-YES* Default Assumptions)

Outcome	Sample	Control Mean	Impact	SE	p-value	Effect Size	Absolute Value of Difference Compared to Model-Based Specification		
							Dif. Effect Size	Dif. SE (Effect Size)	Dif. p-value
Post-test math z-score, TFA sample	Full Sample	-0.57	0.08	0.03	0.010*	0.09	0.00	0.01	0.010
	Subgroup:				0.525				0.148
	Male	-0.58	0.05	0.05	0.332	0.05	0.02	0.02	0.309
	Female	-0.53	0.09	0.04	0.044*	0.11	0.00	0.01	0.043
Post-test math z-score, TNTP sample	Full Sample	-0.43	-0.02	0.04	0.614	-0.02	0.00	0.01	0.078
	Subgroup:				0.644				0.075
	Male	-0.33	0.00	0.05	0.916	0.00	0.01	0.01	0.069
	Female	-0.50	-0.03	0.05	0.490	-0.03	0.01	0.02	0.011

Note: The p-value on the subgroup rows tests for differences in impacts across subgroup categories. Effect sizes are calculated by dividing the impact estimate by the standard deviation of the outcome in the control group.
* Difference is statistically significant at the 0.05 level, two-tailed test.

Table II.J.4. Teach for America and Teaching Fellows Programs—Clustered, Blocked: Additional Information (Cluster-Level Randomization)

		Standard Deviation of Outcomes			Standard Deviation of Outcome Residuals			Intra-Cluster Correlation		
									Design-Based Assumptions	
Outcome	Sample	Full	Treatment Group	Control Group	Full	Treatment Group	Control Group	Model-Based	Matching	Default
Post-test math z-score, TFA sample	Full Sample	0.89	0.90	0.89	0.60	0.62	0.58	0.00	0.20	0.20
	Subgroup:									
	Male	0.92	0.93	0.91	0.62	0.64	0.61			
	Female	0.85	0.85	0.86	0.57	0.59	0.55			
Post-test math z-score, TNTP sample	Full Sample	1.02	1.08	0.96	0.59	0.59	0.58	0.02	0.38	0.38
	Subgroup:									
	Male	1.03	1.11	0.93	0.59	0.63	0.54			
	Female	1.01	1.04	0.97	0.58	0.55	0.60			

Note: The residuals are calculated using all covariates included in the model. All calculations include study weights, if applicable.

II.K. POWER Through Choices—Clustered, Blocked

Original study. Goesling et al. (2015).

Sponsor agency. Office of Adolescent Health, Department of Health & Human Services.

Description of intervention. See Section II.I.

Randomization design. Clustered, blocked. See Section II.I.

Model-based method. Ordinary least squares, block fixed effects, robust cluster standard errors.

We use OLS to estimate the following equation:

(23) $$Y_{ijk} = \alpha_k + \beta X_{ijk} + \delta T_{jk} + \varepsilon_{ijk},$$

where Y_{ijk} is the outcome of interest for individual i in cluster j in block k, X_{ijk} is a vector of covariates, T_{jk} is a treatment indicator, α_k is a block fixed effect, and ε_{ijk} is an error term. As in the original study, we calculate standard errors that allow for ε_{ijk} to be arbitrarily correlated within clusters but uncorrelated across clusters.

Design-based specification: RCT-YES model. Clustered, blocked. Note that we model this as a clustered, non-blocked design as well. Section II.I provides details on this re-estimation of the study.

Design-based specification: RCT-YES parameter (matched settings). Finite population (ATE).

Design-based specification: RCT-YES parameter (default settings). Super population (PATE).

Covariates. See Section II.I.

Weights. No. See Section II.I.

Multiple hypothesis correction. Yes. See Section II.I.

Main differences from original analyses. See Section II.I.

Outcomes. See Section II.I.

Table II.K.1. POWER Through Choices—Clustered, Blocked: Evaluation Impact Estimates (Model-Based)

Outcome	Sample	Control Mean	Impact	SE	p-value	Effect Size
Ability to find methods of protection	Full Sample	51.05	15.42	3.01	0.000*^	0.31
	Subgroup:				0.003*	
	Female	62.26	-1.92	6.36	0.764	-0.04
	Male	47.83	19.87	3.13	0.000*	0.40
General support for methods of protection	Full Sample	3.33	0.17	0.03	0.000*^	0.32
	Subgroup:				0.002*	
	Female	3.50	0.03	0.05	0.514	0.06
	Male	3.28	0.21	0.03	0.000*	0.40
Intentions to have sexual intercourse	Full Sample	68.92	1.63	2.11	0.442	0.04
	Subgroup:				0.357	
	Female	37.86	-2.05	4.50	0.650	-0.04
	Male	77.57	2.57	2.34	0.276	0.06
Intentions to use a condom	Full Sample	43.74	14.07	2.55	0.000*^	0.28
	Subgroup:				0.395	
	Female	55.77	10.48	4.45	0.021*	0.21
	Male	40.33	15.00	2.94	0.000*	0.31
Intentions to use birth control	Full Sample	39.61	5.84	2.87	0.045*	0.12
	Subgroup:				0.145	
	Female	58.82	-1.97	5.92	0.739	-0.04
	Male	34.25	7.84	3.13	0.014*	0.16
Knowledge of HIV and STIs	Full Sample	4.50	0.81	0.11	0.000*^	0.39
	Subgroup:				0.020*	
	Female	4.70	0.32	0.23	0.164	0.18
	Male	4.44	0.94	0.12	0.000*	0.44
Knowledge of methods of protection	Full Sample	6.18	1.70	0.16	0.000*^	0.66
	Subgroup:				0.004*	
	Female	6.43	0.91	0.29	0.002*	0.39
	Male	6.11	1.90	0.17	0.000*	0.72
Knowledge of reproductive anatomy and fertility	Full Sample	2.39	0.36	0.05	0.000*^	0.30
	Subgroup:				0.005*	
	Female	2.56	0.08	0.11	0.446	0.06
	Male	2.34	0.44	0.05	0.000*	0.37
Number of times received info about sex from community center	Full Sample	16.56	13.23	2.64	0.000*^	0.36
	Subgroup:				0.524	
	Female	25.23	16.70	6.09	0.008*	0.38
	Male	14.01	12.34	2.91	0.000*	0.36

Outcome	Sample	Control Mean	Impact	SE	p-value	Effect Size
Number of times received info about sex from doctor or nurse	Full Sample	25.16	0.68	2.35	0.773	0.02
	Subgroup:				0.008*	
	Female	42.06	-11.58	5.07	0.025*	-0.23
	Male	20.27	3.76	2.41	0.122	0.09
Number of times received info about sex from group home	Full Sample	25.61	29.01	3.10	0.000*^	0.66
	Subgroup:				0.000*	
	Female	40.38	4.58	3.31	0.173	0.09
	Male	23.17	32.51	3.42	0.000*	0.77
Number of times received info about sex from school class	Full Sample	21.09	16.98	3.66	0.000*^	0.42
	Subgroup:				0.121	
	Female	18.52	9.82	3.44	0.005*	0.25
	Male	21.83	18.82	4.42	0.000*	0.45
Perceived ability to communicate with partner	Full Sample	3.28	0.26	0.03	0.000*^	0.31
	Subgroup:				0.315	
	Female	3.47	0.20	0.06	0.001*	0.25
	Male	3.23	0.27	0.04	0.000*	0.31
Perceived ability to plan for and avoid unprotected sex	Full Sample	3.09	0.27	0.03	0.000*^	0.32
	Subgroup:				0.702	
	Female	3.29	0.24	0.08	0.003*	0.28
	Male	3.03	0.27	0.04	0.000*	0.32
Perceived access to birth control other than condoms	Full Sample	28.38	4.47	2.23	0.048*^	0.10
	Subgroup:				0.001*	
	Female	43.81	-7.44	3.86	0.058	-0.15
	Male	23.80	7.57	2.40	0.002*	0.18
Perceived access to condoms	Full Sample	57.85	6.35	2.78	0.025*^	0.13
	Subgroup:				0.212	
	Female	50.96	12.75	5.62	0.026*	0.25
	Male	59.83	4.69	3.15	0.140	0.10
Perceived barriers to methods of protection	Full Sample	2.44	0.06	0.03	0.074	0.09
	Subgroup:				0.109	
	Female	2.69	0.16	0.07	0.022*	0.22
	Male	2.37	0.04	0.04	0.345	0.07
Received info on STIs in past 12 months	Full Sample	65.14	23.89	2.36	0.000*^	0.50
	Subgroup:				0.109	
	Female	71.96	18.00	3.34	0.000*	0.40
	Male	63.17	25.39	2.90	0.000*	0.53
Received info on abstinence in past 12 months	Full Sample	60.33	28.76	2.29	0.000*^	0.59
	Subgroup:				0.077	
	Female	69.44	22.71	3.00	0.000*	0.49
	Male	57.68	30.32	2.77	0.000*	0.61

Outcome	Sample	Control Mean	Impact	SE	p-value	Effect Size
Received info on birth control methods in past 12 months	Full Sample	57.74	32.63	2.76	0.000*^	0.66
	Subgroup:				0.000*	
	Female	78.70	12.75	3.73	0.001*	0.31
	Male	51.62	37.76	2.95	0.000*	0.75
Received info on relationships in past 12 months	Full Sample	77.39	14.24	2.20	0.000*^	0.34
	Subgroup:				0.456	
	Female	81.48	11.24	4.36	0.012*	0.29
	Male	76.20	15.00	2.54	0.000*	0.35
Received info on saying no to sex in past 12 months	Full Sample	60.08	32.45	3.00	0.000*^	0.66
	Subgroup:				0.001*	
	Female	77.78	16.48	4.96	0.001*	0.39
	Male	54.96	36.54	3.17	0.000*	0.73
Received info on talking to partner about sex in past 12 months	Full Sample	62.97	30.42	2.26	0.000*^	0.63
	Subgroup:				0.067	
	Female	69.16	23.10	4.21	0.000*	0.50
	Male	61.19	32.29	2.53	0.000*	0.66
Received info on where to get birth control in past 12 months	Full Sample	55.46	38.73	3.11	0.000*^	0.78
	Subgroup:				0.000*	
	Female	76.64	16.74	3.81	0.000*	0.39
	Male	49.32	44.38	3.22	0.000*	0.89

Note: The p-value on the subgroup rows tests for differences in impacts across subgroup categories. Effect sizes are calculated by dividing the impact estimate by the standard deviation of the outcome in the control group.
* Difference is statistically significant at the 0.05 level, two-tailed test.
^ Difference remains statistically significant at the 0.05 level, two-tailed test, after applying the Benjamini-Hochberg correction for multiple hypothesis testing across all full sample analyses in the same domain.

Table II.K.2. POWER Through Choices—Clustered, Blocked: Evaluation Impact Estimates (Design-Based, Matching Assumptions)

Outcome	Sample	Control Mean	Impact	SE	p-value	Effect Size	Absolute Value of Difference Compared to Model-Based Specification		
							Dif. Effect Size	Dif. SE (Effect Size)	Dif. p-value
Ability to find methods of protection	Full Sample	51.05	15.42	4.46	0.002*^	0.31	0.00	0.03	0.002
	Subgroup:				0.012*				0.009
	Female	62.50	-3.63	18.63	0.851	-0.07	0.04	0.25	0.087
	Male	47.96	19.91	4.75	0.000*	0.40	0.00	0.03	0.000
General support for methods of protection	Full Sample	3.33	0.17	0.04	0.000*^	0.32	0.00	0.02	0.000
	Subgroup:				0.094				0.092
	Female	3.49	0.03	0.14	0.837	0.06	0.00	0.18	0.323
	Male	3.28	0.21	0.04	0.000*	0.40	0.00	0.02	0.000
Intentions to have sexual intercourse	Full Sample	68.92	1.63	3.16	0.610	0.04	0.00	0.02	0.168
	Subgroup:				0.580				0.223
	Female	37.62	-2.38	13.45	0.864	-0.05	0.01	0.18	0.214
	Male	77.78	2.50	3.65	0.499	0.06	0.00	0.03	0.223
Intentions to use a condom	Full Sample	43.74	14.07	3.80	0.001*^	0.28	0.00	0.03	0.001
	Subgroup:				0.799				0.404
	Female	54.90	11.70	12.72	0.388	0.23	0.02	0.17	0.367
	Male	40.16	15.10	4.51	0.003*	0.31	0.00	0.03	0.003
Intentions to use birth control	Full Sample	39.61	5.84	4.25	0.179	0.12	0.00	0.03	0.134
	Subgroup:				0.346				0.201
	Female	58.00	-1.04	17.58	0.954	-0.02	0.02	0.24	0.215
	Male	34.07	7.91	4.71	0.105	0.17	0.00	0.03	0.091
Knowledge of HIV and STIs	Full Sample	4.50	0.81	0.16	0.000*^	0.39	0.00	0.02	0.000
	Subgroup:				0.080				0.060
	Female	4.69	0.32	0.69	0.660	0.18	0.00	0.26	0.496
	Male	4.45	0.94	0.18	0.000*	0.44	0.00	0.03	0.000

Outcome	Sample	Control Mean	Impact	SE	p-value	Effect Size	Absolute Value of Difference Compared to Model-Based Specification		
							Dif. Effect Size	Dif. SE (Effect Size)	Dif. p-value
Knowledge of methods of protection	Full Sample	6.18	1.70	0.23	0.000*^	0.66	0.00	0.03	0.000
	Subgroup:				0.030*				0.026
	Female	6.42	0.88	0.81	0.313	0.38	0.01	0.23	0.311
	Male	6.12	1.90	0.26	0.000*	0.72	0.00	0.03	0.000
Knowledge of reproductive anatomy and fertility	Full Sample	2.39	0.36	0.07	0.000*^	0.30	0.00	0.02	0.000
	Subgroup:				0.029*				0.024
	Female	2.56	0.08	0.32	0.820	0.06	0.00	0.17	0.374
	Male	2.34	0.43	0.08	0.000*	0.36	0.01	0.03	0.000
Number of times received info about sex from community center	Full Sample	16.56	13.23	3.83	0.002*^	0.36	0.00	0.03	0.002
	Subgroup:				0.847				0.323
	Female	25.71	15.12	14.06	0.314	0.35	0.04	0.18	0.306
	Male	14.05	12.37	4.32	0.008*	0.36	0.00	0.04	0.008
Number of times received info about sex from doctor or nurse	Full Sample	25.16	0.68	3.44	0.845	0.02	0.00	0.03	0.072
	Subgroup:				0.059				0.051
	Female	42.86	-13.29	12.09	0.304	-0.27	0.03	0.14	0.279
	Male	20.33	3.81	3.57	0.296	0.09	0.00	0.03	0.174
Number of times received info about sex from group home	Full Sample	25.61	29.01	4.89	0.000*^	0.66	0.00	0.04	0.000
	Subgroup:								
	Female	39.22	7.08			0.14	0.05		
	Male	23.25	32.34	5.56	0.000*	0.77	0.00	0.05	0.000
Number of times received info about sex from school class	Full Sample	21.09	16.98	5.36	0.003*^	0.42	0.00	0.04	0.003
	Subgroup:				0.353				0.232
	Female	18.87	8.68	7.67	0.291	0.22	0.03	0.11	0.286
	Male	21.89	18.91	6.57	0.008*	0.46	0.00	0.05	0.008

Outcome	Sample	Control Mean	Impact	SE	p-value	Effect Size	Absolute Value of Difference Compared to Model-Based Specification		
							Dif. Effect Size	Dif. SE (Effect Size)	Dif. p-value
Perceived ability to communicate with partner	Full Sample	3.28	0.26	0.05	0.000*^	0.31	0.00	0.02	0.000
	Subgroup:				0.358				0.043
	Female	3.47	0.20	0.17	0.268	0.25	0.00	0.14	0.267
	Male	3.23	0.27	0.06	0.000*	0.31	0.00	0.02	0.000
Perceived ability to plan for and avoid unprotected sex	Full Sample	3.09	0.27	0.05	0.000*^	0.32	0.00	0.02	0.000
	Subgroup:				0.828				0.126
	Female	3.28	0.26	0.22	0.293	0.31	0.02	0.16	0.290
	Male	3.03	0.27	0.05	0.000*	0.32	0.00	0.01	0.000
Perceived access to birth control other than condoms	Full Sample	28.38	4.47	3.26	0.180	0.10	0.00	0.02	0.132
	Subgroup:				0.105				0.104
	Female	43.69	-7.86	11.26	0.507	-0.16	0.01	0.15	0.449
	Male	23.86	7.55	3.68	0.051	0.18	0.00	0.03	0.049
Perceived access to condoms	Full Sample	57.85	6.35	4.11	0.132	0.13	0.00	0.03	0.107
	Subgroup:				0.246				0.034
	Female	50.00	12.22	16.52	0.484	0.24	0.01	0.22	0.458
	Male	60.00	4.70	4.81	0.337	0.10	0.00	0.03	0.197
Perceived barriers to methods of protection	Full Sample	2.44	0.06	0.05	0.227	0.09	0.00	0.03	0.153
	Subgroup:				0.283				0.174
	Female	2.67	0.18	0.20	0.407	0.25	0.03	0.18	0.385
	Male	2.37	0.04	0.06	0.517	0.07	0.00	0.03	0.172
Received info on STIs in past 12 months	Full Sample	65.14	23.89	3.44	0.000*^	0.50	0.00	0.02	0.000
	Subgroup:				0.312				0.203
	Female	72.38	18.49	7.64	0.042*	0.41	0.01	0.10	0.042
	Male	63.34	25.11	4.24	0.000*	0.52	0.01	0.03	0.000

Outcome	Sample	Control Mean	Impact	SE	p-value	Effect Size	Absolute Value of Difference Compared to Model-Based Specification: Dif. Effect Size	Dif. SE (Effect Size)	Dif. p-value
Received info on abstinence in past 12 months	Full Sample	60.33	28.76	3.34	0.000*^	0.59	0.00	0.02	0.000
	Subgroup:				0.296				0.219
	Female	69.81	23.22	6.54	0.007*	0.50	0.01	0.08	0.007
	Male	57.84	30.04	4.08	0.000*	0.61	0.01	0.03	0.000
Received info on birth control methods in past 12 months	Full Sample	57.74	32.63	4.05	0.000*^	0.66	0.00	0.03	0.000
	Subgroup:				0.001*				0.001
	Female	78.30	13.91	8.73	0.150	0.34	0.03	0.12	0.149
	Male	51.76	37.57	4.39	0.000*	0.75	0.00	0.03	0.000
Received info on relationships in past 12 months	Full Sample	77.39	14.24	3.22	0.000*^	0.34	0.00	0.02	0.000
	Subgroup:				0.626				0.170
	Female	81.13	11.15	10.33	0.312	0.29	0.00	0.15	0.300
	Male	76.41	14.89	3.78	0.001*	0.35	0.00	0.03	0.001
Received info on saying no to sex in past 12 months	Full Sample	60.08	32.45	4.41	0.000*^	0.66	0.00	0.03	0.000
	Subgroup:				0.008*				0.007
	Female	78.30	15.83	11.62	0.210	0.38	0.02	0.16	0.209
	Male	54.84	36.54	4.78	0.000*	0.73	0.00	0.03	0.000
Received info on talking to partner about sex in past 12 months	Full Sample	62.97	30.42	3.32	0.000*^	0.63	0.00	0.02	0.000
	Subgroup:				0.217				0.150
	Female	69.52	22.58	10.01	0.054	0.49	0.01	0.12	0.054
	Male	61.35	32.10	3.79	0.000*	0.66	0.00	0.03	0.000

Outcome	Sample	Control Mean	Impact	SE	p-value	Effect Size	Absolute Value of Difference Compared to Model-Based Specification		
							Dif. Effect Size	Dif. SE (Effect Size)	Dif. p-value
Received info on where to get birth control in past 12 months	Full Sample	55.46	38.73	4.57	0.000*^	0.78	0.00	0.03	0.000
	Subgroup:				0.000*				0.000
	Female	77.14	16.09	9.07	0.114	0.38	0.02	0.12	0.114
	Male	49.46	44.21	4.87	0.000*	0.88	0.00	0.03	0.000

Note: The p-value on the subgroup rows tests for differences in impacts across subgroup categories. Effect sizes are calculated by dividing the impact estimate by the standard deviation of the outcome in the control group.
* Difference is statistically significant at the 0.05 level, two-tailed test.
^ Difference remains statistically significant at the 0.05 level, two-tailed test, after applying the Benjamini-Hochberg correction for multiple hypothesis testing across all full sample analyses in the same domain.

Table II.K.3. POWER Through Choices—Clustered, Blocked: Evaluation Impact Estimates (Design-Based, *RCT-YES* Default Assumptions)

							Absolute Value of Difference Compared to Model-Based Specification		
Outcome	Sample	Control Mean	Impact	SE	p-value	Effect Size	Dif. Effect Size	Dif. SE (Effect Size)	Dif. p-value
Ability to find methods of protection	Full Sample	53.45	11.08	4.16	0.012*^	0.22	0.09	0.02	0.012
	Subgroup:				0.042*				0.039
	Female	65.37	-5.08	8.68	0.577	-0.10	0.06	0.05	0.187
	Male	49.71	15.94	4.83	0.003*	0.32	0.08	0.03	0.003
General support for methods of protection	Full Sample	3.31	0.19	0.05	0.001*^	0.36	0.04	0.04	0.001
	Subgroup:				0.215				0.213
	Female	3.46	0.10	0.07	0.221	0.21	0.14	0.04	0.293
	Male	3.25	0.22	0.06	0.002*	0.42	0.02	0.06	0.002
Intentions to have sexual intercourse	Full Sample	63.79	2.82	4.12	0.499	0.06	0.03	0.04	0.057
	Subgroup:				0.833				0.476
	Female	39.21	1.23	8.59	0.890	0.03	0.07	0.08	0.240
	Male	72.03	3.31	4.66	0.484	0.08	0.02	0.06	0.208
Intentions to use a condom	Full Sample	48.08	8.84	4.47	0.057	0.18	0.11	0.04	0.057
	Subgroup:				0.731				0.336
	Female	53.49	6.43	9.31	0.512	0.13	0.08	0.10	0.491
	Male	45.98	10.10	5.01	0.055	0.21	0.10	0.04	0.055
Intentions to use birth control	Full Sample	43.30	0.02	4.88	0.996	0.00	0.12	0.04	0.951
	Subgroup:				0.991				0.846
	Female	52.47	0.38	11.08	0.974	0.01	0.05	0.10	0.235
	Male	40.00	0.23	5.26	0.966	0.00	0.16	0.04	0.952
Knowledge of HIV and STIs	Full Sample	4.27	0.94	0.19	0.000*^	0.46	0.06	0.04	0.000
	Subgroup:				0.143				0.123
	Female	4.57	0.48	0.34	0.201	0.27	0.09	0.06	0.037
	Male	4.18	1.10	0.23	0.000*	0.51	0.07	0.05	0.000

Outcome	Sample	Control Mean	Impact	SE	p-value	Effect Size	Absolute Value of Difference Compared to Model-Based Specification		
							Dif. Effect Size	Dif. SE (Effect Size)	Dif. p-value
Knowledge of methods of protection	Full Sample	5.83	1.87	0.28	0.000*^	0.73	0.07	0.05	0.000
	Subgroup:				0.198				0.194
	Female	6.17	1.26	0.52	0.045*	0.55	0.15	0.10	0.043
	Male	5.72	2.07	0.33	0.000*	0.78	0.06	0.06	0.000
Knowledge of reproductive anatomy and fertility	Full Sample	2.31	0.31	0.10	0.006*^	0.26	0.04	0.04	0.006
	Subgroup:				0.169				0.164
	Female	2.53	0.08	0.17	0.658	0.06	0.00	0.05	0.212
	Male	2.23	0.38	0.12	0.006*	0.32	0.05	0.06	0.006
Number of times received info about sex from community center	Full Sample	16.12	17.65	4.39	0.000*^	0.47	0.12	0.05	0.000
	Subgroup:				0.822				0.298
	Female	26.81	15.82	7.32	0.063	0.36	0.02	0.03	0.055
	Male	12.73	17.87	5.29	0.003*	0.51	0.16	0.07	0.003
Number of times received info about sex from doctor or nurse	Full Sample	24.65	0.50	3.77	0.895	0.01	0.00	0.03	0.122
	Subgroup:				0.021*				0.013
	Female	41.00	-14.52	6.83	0.066	-0.29	0.06	0.04	0.041
	Male	19.39	5.19	4.44	0.253	0.13	0.04	0.05	0.131
Number of times received info about sex from group home	Full Sample	26.89	29.19	4.94	0.000*^	0.67	0.00	0.04	0.000
	Subgroup:				0.029*				0.029
	Female	37.64	10.16	8.22	0.304	0.21	0.11	0.10	0.131
	Male	24.43	33.90	5.86	0.000*	0.80	0.03	0.06	0.000
Number of times received info about sex from school class	Full Sample	19.93	12.30	3.82	0.003*^	0.30	0.11	0.00	0.003
	Subgroup:				0.262				0.141
	Female	19.57	5.10	6.75	0.472	0.13	0.12	0.08	0.467
	Male	20.14	14.41	4.57	0.004*	0.35	0.11	0.00	0.004

Outcome	Sample	Control Mean	Impact	SE	p-value	Effect Size	Absolute Value of Difference Compared to Model-Based Specification		
							Dif. Effect Size	Dif. SE (Effect Size)	Dif. p-value
Perceived ability to communicate with partner	Full Sample	3.31	0.23	0.05	0.000*^	0.27	0.04	0.02	0.000
	Subgroup:				0.569				0.254
	Female	3.52	0.18	0.09	0.071	0.23	0.03	0.04	0.070
	Male	3.24	0.24	0.06	0.001*	0.28	0.03	0.02	0.001
Perceived ability to plan for and avoid unprotected sex	Full Sample	3.11	0.27	0.06	0.000*^	0.32	0.00	0.04	0.000
	Subgroup:				0.788				0.086
	Female	3.31	0.25	0.11	0.055	0.29	0.01	0.04	0.052
	Male	3.04	0.29	0.07	0.000*	0.35	0.02	0.04	0.000
Perceived access to birth control other than condoms	Full Sample	29.06	2.84	3.06	0.360	0.06	0.04	0.02	0.312
	Subgroup:				0.112				0.111
	Female	43.86	-6.98	7.20	0.364	-0.14	0.01	0.07	0.306
	Male	24.18	5.97	3.29	0.082	0.14	0.04	0.02	0.080
Perceived access to condoms	Full Sample	54.03	7.35	3.78	0.061	0.15	0.02	0.02	0.036
	Subgroup:				0.327				0.115
	Female	50.28	14.40	8.73	0.143	0.29	0.03	0.06	0.117
	Male	55.24	4.78	4.12	0.259	0.10	0.00	0.02	0.119
Perceived barriers to methods of protection	Full Sample	2.52	0.04	0.05	0.424	0.06	0.03	0.03	0.350
	Subgroup:				0.205				0.096
	Female	2.68	0.16	0.09	0.130	0.22	0.00	0.03	0.108
	Male	2.47	0.01	0.06	0.857	0.02	0.05	0.03	0.512
Received info on STIs in past 12 months	Full Sample	63.06	25.51	4.01	0.000*^	0.53	0.03	0.03	0.000
	Subgroup:				0.575				0.466
	Female	69.16	21.77	7.18	0.016*	0.48	0.08	0.09	0.016
	Male	61.31	26.67	4.80	0.000*	0.55	0.03	0.04	0.000

Outcome	Sample	Control Mean	Impact	SE	p-value	Effect Size	Absolute Value of Difference Compared to Model-Based Specification		
							Dif. Effect Size	Dif. SE (Effect Size)	Dif. p-value
Received info on abstinence in past 12 months	Full Sample	61.30	28.39	3.97	0.000*^	0.58	0.01	0.03	0.000
	Subgroup:				0.807				0.730
	Female	67.16	26.83	6.22	0.003*	0.58	0.09	0.07	0.003
	Male	59.67	28.78	4.87	0.000*	0.58	0.03	0.04	0.000
Received info on birth control methods in past 12 months	Full Sample	58.91	31.10	3.66	0.000*^	0.63	0.03	0.02	0.000
	Subgroup:				0.033*				0.033
	Female	74.29	17.16	7.28	0.046*	0.42	0.11	0.09	0.045
	Male	53.83	35.94	4.26	0.000*	0.72	0.04	0.03	0.000
Received info on relationships in past 12 months	Full Sample	74.48	18.22	4.12	0.000*^	0.44	0.10	0.05	0.000
	Subgroup:				0.731				0.275
	Female	77.68	15.71	7.96	0.084	0.40	0.11	0.09	0.072
	Male	73.51	18.93	4.79	0.001*	0.44	0.09	0.05	0.001
Received info on saying no to sex in past 12 months	Full Sample	62.23	30.53	4.05	0.000*^	0.62	0.04	0.02	0.000
	Subgroup:				0.181				0.180
	Female	74.84	20.65	8.40	0.039*	0.49	0.10	0.08	0.038
	Male	58.11	33.73	4.57	0.000*	0.68	0.06	0.03	0.000
Received info on talking to partner about sex in past 12 months	Full Sample	63.56	28.64	3.59	0.000*^	0.59	0.04	0.03	0.000
	Subgroup:				0.577				0.510
	Female	67.63	24.96	6.99	0.007*	0.54	0.04	0.06	0.007
	Male	62.49	29.55	4.19	0.000*	0.61	0.06	0.03	0.000

Outcome	Sample	Control Mean	Impact	SE	p-value	Effect Size	Absolute Value of Difference Compared to Model-Based Specification		
							Dif. Effect Size	Dif. SE (Effect Size)	Dif. p-value
Received info on where to get birth control in past 12 months	Full Sample	56.70	36.28	3.73	0.000*^	0.73	0.05	0.01	0.000
	Subgroup:				0.014*				0.014
	Female	73.31	19.04	7.51	0.035*	0.45	0.05	0.09	0.035
	Male	51.42	41.71	4.34	0.000*	0.83	0.05	0.02	0.000

Note: The p-value on the subgroup rows tests for differences in impacts across subgroup categories. Effect sizes are calculated by dividing the impact estimate by the standard deviation of the outcome in the control group.
* Difference is statistically significant at the 0.05 level, two-tailed test.
^ Difference remains statistically significant at the 0.05 level, two-tailed test, after applying the Benjamini-Hochberg correction for multiple hypothesis testing across all full sample analyses in the same domain.

Table II.K.4. POWER Through Choices—Clustered, Blocked: Additional Information (Cluster-Level Randomization)

		Standard Deviation of Outcomes			Standard Deviation of Outcome Residuals			Intra-Cluster Correlation		
									Design-Based Assumptions	
Outcome	Sample	Full	Treatment Group	Control Group	Full	Treatment Group	Control Group	Model-Based	Matching	Default
Ability to find methods of protection	Full Sample	49.31	47.42	50.04	44.93	45.39	44.53	0.00	0.06	0.11
	Subgroup:									
	Female	48.80	49.18	48.70	44.48	46.58	42.93			
	Male	49.45	46.98	50.02	44.81	44.89	44.79			
General support for methods of protection	Full Sample	0.52	0.50	0.52	0.41	0.41	0.42	-0.01	0.13	0.33
	Subgroup:									
	Female	0.50	0.52	0.49	0.42	0.39	0.43			
	Male	0.52	0.50	0.52	0.41	0.41	0.41			
Intentions to have sexual intercourse	Full Sample	46.19	46.09	46.33	36.43	36.53	36.37	-0.02	0.28	0.28
	Subgroup:									
	Female	47.85	46.84	48.74	41.33	39.96	42.66			
	Male	41.45	41.18	41.77	35.07	35.72	34.46			
Intentions to use a condom	Full Sample	50.02	49.45	49.66	45.82	45.84	45.84	-0.02	0.09	0.14
	Subgroup:									
	Female	48.96	47.41	49.91	45.12	41.69	48.01			
	Male	50.00	49.75	49.12	46.01	46.80	45.27			
Intentions to use birth control	Full Sample	49.43	49.81	48.96	45.88	47.44	44.35	-0.01	0.13	0.14
	Subgroup:									
	Female	49.54	49.89	49.46	47.58	49.30	46.30			
	Male	48.65	49.48	47.52	45.41	46.99	43.83			
Knowledge of HIV and STIs	Full Sample	1.93	1.69	2.07	1.62	1.53	1.70	-0.01	0.10	0.29
	Subgroup:									
	Female	1.75	1.71	1.77	1.60	1.57	1.64			
	Male	1.98	1.68	2.14	1.62	1.52	1.71			

Outcome	Sample	Standard Deviation of Outcomes			Standard Deviation of Outcome Residuals			Intra-Cluster Correlation		
									Design-Based Assumptions	
		Full	Treatment Group	Control Group	Full	Treatment Group	Control Group	Model-Based	Matching	Default
Knowledge of methods of protection	Full Sample	2.54	2.18	2.58	2.00	2.03	1.98	0.02	0.13	0.36
	Subgroup:									
	Female	2.42	2.45	2.31	2.14	2.33	1.98			
	Male	2.57	2.10	2.65	1.95	1.95	1.96			
Knowledge of reproductive anatomy and fertility	Full Sample	1.17	1.11	1.21	0.92	0.93	0.91	-0.02	0.09	0.28
	Subgroup:									
	Female	1.19	1.09	1.26	0.98	0.91	1.04			
	Male	1.17	1.11	1.19	0.90	0.94	0.87			
Number of times received info about sex from community center	Full Sample	41.70	45.21	37.21	39.12	42.13	36.08	-0.01	0.11	0.10
	Subgroup:									
	Female	46.67	49.34	43.64	44.20	48.48	40.86			
	Male	39.99	43.87	34.76	37.70	40.62	34.60			
Number of times received info about sex from doctor or nurse	Full Sample	43.31	43.22	43.44	41.64	42.26	41.09	-0.02	0.04	0.07
	Subgroup:									
	Female	48.56	46.82	49.60	47.06	47.40	47.02			
	Male	41.28	42.28	40.26	40.02	40.92	39.16			
Number of times received info about sex from group home	Full Sample	48.78	49.96	43.71	43.52	45.40	41.80	-0.01	0.16	0.14
	Subgroup:									
	Female	49.90	50.67	49.55	45.81	46.07	46.07			
	Male	48.61	49.92	42.26	42.94	45.09	40.89			
Number of times received info about sex from school class	Full Sample	45.73	48.80	40.83	41.07	43.21	38.97	0.04	0.24	0.14
	Subgroup:									
	Female	42.06	45.28	39.03	38.46	39.31	37.96			
	Male	46.49	49.31	41.37	41.70	44.05	39.27			

Outcome	Sample	Standard Deviation of Outcomes			Standard Deviation of Outcome Residuals			Intra-Cluster Correlation		
									Design-Based Assumptions	
		Full	Treatment Group	Control Group	Full	Treatment Group	Control Group	Model-Based	Matching	Default
Perceived ability to communicate with partner	Full Sample	0.76	0.63	0.85	0.65	0.53	0.75	-0.03	0.04	0.09
	Subgroup:									
	Female	0.73	0.63	0.79	0.63	0.48	0.74			
	Male	0.77	0.63	0.86	0.66	0.54	0.76			
Perceived ability to plan for and avoid unprotected sex	Full Sample	0.77	0.65	0.85	0.65	0.55	0.73	-0.03	0.08	0.14
	Subgroup:									
	Female	0.76	0.62	0.85	0.66	0.53	0.75			
	Male	0.77	0.66	0.84	0.65	0.56	0.73			
Perceived access to birth control other than condoms	Full Sample	45.71	46.31	45.14	41.39	41.88	40.94	-0.03	0.03	0.05
	Subgroup:									
	Female	49.31	48.67	49.85	45.85	45.64	46.23			
	Male	44.29	45.71	42.64	40.03	40.89	39.18			
Perceived access to condoms	Full Sample	48.77	47.93	49.43	43.80	42.29	45.29	-0.01	0.06	0.13
	Subgroup:									
	Female	49.63	48.17	50.23	45.68	44.76	46.65			
	Male	48.53	47.93	49.09	43.30	41.72	44.93			
Perceived barriers to methods of protection	Full Sample	0.60	0.55	0.65	0.50	0.48	0.53	-0.01	0.14	0.18
	Subgroup:									
	Female	0.67	0.59	0.72	0.55	0.49	0.60			
	Male	0.57	0.52	0.61	0.49	0.48	0.50			
Received info on STIs in past 12 months	Full Sample	42.28	31.78	47.70	38.48	31.65	44.16	-0.01	0.07	0.15
	Subgroup:									
	Female	39.79	29.06	45.13	36.53	27.66	42.53			
	Male	42.88	32.39	48.30	38.96	32.48	44.64			

Outcome	Sample	Standard Deviation of Outcomes			Standard Deviation of Outcome Residuals			Intra-Cluster Correlation		
									Design-Based Assumptions	
		Full	Treatment Group	Control Group	Full	Treatment Group	Control Group	Model-Based	Matching	Default
Received info on abstinence in past 12 months	Full Sample	43.64	31.31	48.97	39.22	31.13	45.74	-0.02	0.08	0.11
	Subgroup:									
	Female	40.10	25.49	46.28	36.60	23.82	44.41			
	Male	44.43	32.46	49.47	39.86	32.57	46.14			
Received info on birth control methods in past 12 months	Full Sample	44.06	29.92	49.45	38.21	30.31	44.59	0.01	0.14	0.15
	Subgroup:									
	Female	36.17	27.36	41.13	33.22	26.27	38.02			
	Male	45.49	30.49	50.04	39.03	30.63	46.09			
Received info on relationships in past 12 months	Full Sample	36.42	28.04	41.88	34.51	28.72	39.33	-0.01	0.05	0.17
	Subgroup:									
	Female	34.08	25.49	39.03	32.03	26.71	35.87			
	Male	36.99	28.61	42.64	35.13	29.18	40.31			
Received info on saying no to sex in past 12 months	Full Sample	43.03	28.04	49.02	37.77	29.21	44.55	0.02	0.16	0.15
	Subgroup:									
	Female	35.16	21.06	41.77	32.76	23.58	38.72			
	Male	44.49	29.36	49.82	38.73	30.05	45.94			
Received info on talking to partner about sex in past 12 months	Full Sample	41.47	24.95	48.34	36.91	25.57	45.34	-0.01	0.06	0.12
	Subgroup:									
	Female	39.79	23.41	46.40	36.53	23.95	44.32			
	Male	41.90	25.31	48.80	36.97	25.90	45.63			
Received info on where to get birth control in past 12 months	Full Sample	43.63	23.84	49.75	36.47	25.65	44.59	0.03	0.16	0.17
	Subgroup:									
	Female	36.25	23.41	42.51	32.80	23.95	38.65			
	Male	45.00	23.97	50.06	36.88	25.32	45.81			

Note: The residuals are calculated using all covariates included in the model. All calculations include study weights, if applicable.

(This page intentionally blank)