



## Frequently Asked Questions

May, 2016

*RCT-YES* was developed by Mathematica Policy Research, Inc.  
with funding from the Institute of Education Sciences  
under Contract ED-IES-12-C-0057 with  
Decision Information Resources, Inc.

© Mathematica Policy Research, Inc.



## Contents

<b>Download, System, and Software Issues</b> .....	<b>1</b>
1. Is <i>RCT-YES</i> supported for Macintosh or Linux? .....	1
2. When installing the <i>RCT-YES</i> software, why does a yellow window titled “User Account Control” appear requiring administrative privileges? .....	1
3. To which versions of Windows can <i>RCT-YES</i> be installed? .....	1
4. Which version of the .NET Framework is required? .....	1
5. Is <i>RCT-YES</i> compliant with the 1998 Section 508 amendment to the Rehabilitation Act? .....	1
6. How do I get R and Stata? .....	1
7. Which versions of R and Stata do I need? .....	1
8. How do I get the software needed for graphing? .....	2
9. I don’t have permission to install R and/or Stata on my computer. What should I do? .....	2
10. Is there an advantage to using R or Stata? .....	2
<b>Impact Analysis Issues</b> .....	<b>3</b>
1. I ran the Stata (.do) or R (.R) program produced by the interface, but the program stopped and did not conduct the analysis. Why did this happen? .....	3
2. Why did the program drop some of my outcomes?.....	6
3. Why did the program drop the weights that I specified for the analysis? .....	6
4. Why did the program exclude some of my subgroups? .....	6
5. Why did the program drop some of my baseline covariates? .....	6
6. Why did the .html tables produce impact estimates but no standard errors or p-values? .....	7
7. Why do my intention-to-treat (ITT) impact results differ depending on whether or not I conduct a complier average causal effect (CACE) analysis? .....	7
8. Why do my results change using the finite-population (FP) and super-population (SP) models? .....	8
9. How does <i>RCT-YES</i> treat missing outcomes, covariates, and subgroups?.....	8
10. How does <i>RCT-YES</i> weight clusters? .....	9
11. How does <i>RCT-YES</i> weight blocks?.....	9

12.	Can I include different covariates for different outcomes in the same domain? .....	9
13.	What is the scale of the impact estimates for binary (dichotomous) outcomes?.....	9
14.	How does <i>RCT-YES</i> handle covariates that are collinear? .....	9
15.	How does <i>RCT-YES</i> handle covariates for matched pair and other blocked designs with SUPER_POP=1? .....	10
16.	Can I assume a finite-population model for matched pair designs? .....	10
<b>Graphing Issues .....</b>		<b>11</b>
1.	Why do I get hung up (stuck) in the R workspace after leaving <i>RCT-YES-Graph</i> ? .....	11
2.	Why does the error “ERROR: [on_request_read] connection reset by peer” appear in the R workspace after exiting <i>RCT-YES-Graph</i> ?.....	11
3.	Why does my graph not appear in the dashboard? .....	11
4.	Is there a way to change the order that the outcome and subgroup variables are displayed in the graph?.....	11
5.	Is there a way to change the order of the lists of outcome and subgroup variables in the dashboard inputs? .....	12
6.	Is there a way to plot outcomes or subgroups from different domains in the same graph? .....	12
7.	Is there a way to change graph colors or fonts? .....	12
8.	Can I save my graph inputs? .....	12
9.	Why is the width of the graphs in the dashboard wider than in the output files? .....	12

## Download, System, and Software Issues

### 1. Is *RCT-YES* supported for Macintosh or Linux?

No. Currently, *RCT-YES* is supported only for Windows.

### 2. When installing the *RCT-YES* software, why does a yellow window titled “User Account Control” appear requiring administrative privileges?

If your system requires administrative privileges, a system administrator will need to run the executable file to install the software.

### 3. To which versions of Windows can *RCT-YES* be installed?

*RCT-YES* can be installed on any Windows machine that supports the .NET Framework v3.5 or greater. The .NET Framework can be installed on a computer with Windows XP or later versions.

### 4. Which version of the .NET Framework is required?

Version 3.5 or greater.

### 5. Is *RCT-YES* compliant with the 1998 Section 508 amendment to the Rehabilitation Act?

Yes. The *RCT-YES* interface, .html output files, and documentation are accessible (508 compliant) for those with disabilities.

### 6. How do I get R and Stata?

You will need to install either the free R software or Stata to use *RCT-YES*. You will also need R if you want to use the *RCT-YES-Graph* application to plot your impact findings. You can get information on how to download R for free from [www.cran.r-project.org](http://www.cran.r-project.org) and how to download Stata from [www.stata.com](http://www.stata.com).

### 7. Which versions of R and Stata do I need?

*RCT-YES*, Version 1 was tested using R Version 3.1.0 through 3.2.4, and Stata Version 13. Because newer versions of R and Stata are typically compatible with previous versions, we expect that *RCT-YES* will run smoothly with future R and Stata versions. However, the program may not run properly

for R and Stata versions released prior to the versions used for *RCT-YES* testing. Thus, we suggest that you install recent versions of R and Stata before you run *RCT-YES*.

## **8. How do I get the software needed for graphing?**

To run *RCT-YES-Graph*, you will need to install R and three free R packages. To install the three packages, type the following line at the prompt in the R workspace:

```
install.packages(c('shiny', 'shinydashboard', 'ggplot2'), repos='https://cran.rstudio.com')
```

Chapter 7 of the User's Manual describes these steps in detail.

## **9. I don't have permission to install R and/or Stata on my computer. What should I do?**

Contact your system administrator, who will be able to assess the best way to install the required software.

## **10. Is there an advantage to using R or Stata?**

*RCT-YES* results should be very similar using R and Stata, with all output files (.html, .csv, and .log) produced similarly. Aside from user preferences or familiarity, there should be no advantage to one over the other.

## Impact Analysis Issues

We urge RCT-YES users to consult the RCT-YES User's Manual to address problems they are having with conducting the impact analysis and producing the impact results. However, to help speed up the process of resolving program issues, we provide answers below to some common user questions.

### 1. I ran the Stata (.do) or R (.R) program produced by the interface, but the program stopped and did not conduct the analysis. Why did this happen?

There are several reasons why RCT-YES will not conduct the impact analysis. Below, we list a few of the primary reasons:

- ***The input data file is not found, is corrupted, or is not in the proper format.*** The input data file must be a .dta file if you plan to use Stata for the analysis and a .rds file if you plan to use R. Appendix A of the User's Manual describes steps for creating these data files from an EXCEL (.csv) file.
- ***RCT-YES program files were moved to another directory, changed, or deleted.*** When RCT-YES is downloaded from the website, the executable file will create program directories on your computer as well as a series of program files that are used to conduct the analysis. If these files are altered, the Stata or R computer programs may not run. If the files cannot be restored, you might need to re-download RCT-YES from the website.
- ***The specifications in the interface screens were changed without re-generating the output files.*** Suppose you previously entered program inputs into the interface screens and generated the output files successfully. Suppose the base name for these files is "my\_rct\_files". Suppose next that you re-enter the interface, read in the previous input specification file called "my\_rct\_files.rctyes", make changes to the inputs, and exit the interface without having successfully re-generated the Stata or R analysis program. In that case, RCT-YES will not conduct the analysis because the input file, "my\_rct\_files.rctyes", will no longer align with the old Stata or R program file that was created earlier. This problem can be fixed by returning to the interface screens and re-generating the output files using the **Generate Output Files** screen to produce updated Stata or R programs that reflect the changes to the program inputs.
- ***The user changed the input specification (.rctyes) file directly without returning to the interface.*** Some advanced RCT-YES users may want to make changes to program inputs directly in the .rctyes file rather than making changes using the interface screens and then generating an updated .rctyes file (and associated files). If the changes to the .rctyes file were not done correctly, the program could crash without reasons provided in the .html or .log files (our program testing

did not address these types of errors, which could be numerous). We recommend that all changes to program inputs be made using the interface screens to avoid these types of errors.

- **Key design variables have missing or invalid values.** RCT-YES requires that key design variables exist in the data file, are available for all observations, and are coded correctly. Specifically:
  - RCT-YES requires that the treatment-control indicator variable (TC\_STATUS) be available for all observations in the data file or the program will stop. In addition, the TC\_STATUS variable cannot have values other than 0 or 1.
  - For clustered designs, the program will not continue if the TC\_STATUS variable varies within the same cluster (that is, has both 0 and 1 values for the same CLUSTER\_ID value).
  - For clustered and/or blocked designs, the program will stop if the CLUSTER\_ID and/or BLOCK\_ID variables have missing values or invalid codes.
  - For clustered designs, if data are provided as cluster-level averages (TYPE\_CLUS\_DATA = 1), the CLUSTER\_FULL variable must be available for all records and must be coded as 1 for records pertaining to the full sample analysis and 0 for records pertaining to each subgroup analysis. Other problems could also emerge for this specification if the data are not structured properly. The section entitled **Data with cluster-level records** in Chapter 4 of the User's Manual discusses the required data structure for this specification.
  
- **All specified outcome variables are invalid.** RCT-YES omits an outcome variable from the analysis for a variety of reasons:
  - (i) **The outcome variable is not found in the input data file.** We strongly recommend that you use the interface to create a *variable list window* that displays the variables in your data file. This window will help you with data entry and minimize variable naming errors. Remember also that both R and Stata are case sensitive, so it matters whether the variable names have lowercase or uppercase letters.
  - (ii) **Small sample sizes.** RCT-YES employs several strategies to help protect personally identifiable information (PII) on individuals and institutions (see Chapter 3 of the User's Manual for a detailed discussion). One key strategy that conforms to industry standards is omitting outcomes that have small numbers of individuals or clusters with available data. The minimum sample size rules used by RCT-YES are somewhat complex, but can be summarized as follows:
    - For Design 1 (non-clustered, non-blocked), by default, RCT-YES will estimate impacts for an outcome variable only if it contains at least 10 treatment and 10 control group members with available (nonmissing) data. The minimum size threshold of 10 can be

changed using the MIN\_NUM option in the interface's **Optional Design and Analysis Parameters** screen, but must be at least 3.

- For Design 2 (non-clustered, blocked), *RCT-YES* applies the same criteria as for Design 1 after removing blocks with small sample sizes. By default, a block is included in the analysis only if it contains at least 2 treatment and 2 control group members with available data so that proper variance estimates can be calculated. If there is concern that too many blocks are being dropped (as reported in Table 4 in the .html file), you can reduce the size threshold to 1 treatment and 1 control group member per block by either specifying the BLOCK\_FE = 1 option or the SUPER\_POP = 1 option with the CATE\_UATE = 0 or 2 option (see Chapter 6 of the Technical Methods Appendix for a detailed discussion of these *different* specifications and their underlying assumptions and estimators).
  - For Design 3 (clustered, non-blocked), *RCT-YES* will include an outcome variable in the analysis if the Design 1 criteria are satisfied (for data provided at the individual level) and if there are at least 2 treatment and 2 control group clusters with available data (the program allows clusters with only one individual).
  - For Design 4 (clustered, blocked), *RCT-YES* applies the same criteria as for Design 3 after removing blocks with small numbers of clusters. By default, a block is included in the analysis only if it contains at least 2 treatment and 2 control group clusters with available data. You can reduce the size threshold to 1 treatment and 1 control group cluster per block by either specifying the BLOCK\_FE = 1 option or the SUPER\_POP = 1 option with the CATE\_UATE = 0 or 2 option (see Chapter 6 of the Technical Methods Appendix for a detailed discussion).
- (iii) **Binary variables have too few records with values of 0 or 1.** To help protect PII, *RCT-YES* excludes binary (0/1) outcomes where there are fewer than 5 observations with a value of 0 or fewer than 5 observations with a value of 1 for either the treatment or control group. The one exception to this rule is the intervention receipt indicator (GOT\_TREAT) variables for the complier average causal effects (CACE) analyses which can have means near 0 or 100 percent.
- (iv) **The outcome variables have insufficient variation.** *RCT-YES* conducts analyses using only outcomes whose values vary across the sample; this condition must hold for both the treatment and control groups. Thus, the program excludes outcomes that have zero variance (that is, that all have the same value).

## 2. Why did the program drop some of my outcomes?

See the answer to the previous question discussing reasons why *RCT-YES* omits outcomes from the analysis.

## 3. Why did the program drop the weights that I specified for the analysis?

*RCT-YES* ignores a weight variable if it contains any missing or invalid values (for example, negative, zero, or character values) for observations with available data for the outcome and subgroup under investigation. It also drops a weight variable if it is not found in the data file. In these cases, the default weights are applied. To fix this problem, the weights in the input data file should be recoded to ensure that they are nonmissing for all observations and have numeric, positive values.

## 4. Why did the program exclude some of my subgroups?

*RCT-YES* conducts subgroup analyses for *categorical* subgroups in which each sample member is allocated to a specific category. To help protect PII, *RCT-YES omits all subgroup categories from the analysis if any category is too small (using the size rules described in Question 1 above)*. This procedure is used because knowledge of the outcomes from the larger subgroup categories and for the full sample can be used to calculate the outcomes of those in the small subgroup categories. In these cases, the small subgroup categories should be combined into larger ones or omitted from the input statements and the analysis should be rerun.

## 5. Why did the program drop some of my baseline covariates?

*RCT-YES* will exclude baseline covariates from the regression models if they are not found in the input data file. Other key reasons why covariates will be omitted from the analysis are summarized below (see Tables 6 and 7 in the .html file as well as the .log file to see the covariates that entered the models):

- ***Too many covariates are specified for the analysis.*** *RCT-YES* imposes rules to limit the number of covariates that are included in the models in order to avoid model overfitting. By default, the program requires that the data file contain at least 5 individuals per covariate for non-clustered designs, 5 clusters per covariate for clustered designs, and 5 blocks per covariate for blocked designs (Designs 2 and 4) with the SUPER\_POP=1 specification. *If these conditions are not met, all covariates are excluded from the analysis.* The default value of 5 can be changed using the OBS\_COV option in the interface's **Optional Design and Analysis Parameters** screen, but we recommend that it should not be set lower than 3.

- ***The covariate has too many missing values.*** RCT-YES replaces missing values of covariates as long as the covariate is missing for 30 percent or fewer cases (the default value) for both the treatment and control groups; otherwise the covariate is dropped from the analysis. The default 30 percent value can be changed using the MISSING\_COV option in the interface's **Optional Design and Analysis Parameters** screen.
- ***The covariate is perfectly correlated with other covariates included in the model.*** This could occur, for example, if a covariate takes on only a single value, if the TC\_STATUS variable is added as a covariate, if binary variables are entered as covariates for each level of a categorical variable (in which case the program will drop one of the binary variables), if block indicator variables are included as covariates for Designs 2 and 4 (because the program includes these indicators), and if subgroup indicators are included as covariates for subgroup analyses (because the program includes these indicators).
- ***The regression model is “too large.”*** In some instances for Designs 2 and 4, if the regression model contains many blocks, subgroups, and covariates, RCT-YES will omit all covariates from the model to avoid system capacity complications and invoke the SUPER\_POP=1 command (see **Methods for blocked designs** in Chapter 2 of the User's Manual for a discussion of this issue).

## 6. Why did the .html tables produce impact estimates but no standard errors or p-values?

This will most likely occur if standard errors of the impacts cannot be estimated due to insufficient degrees of freedom. In these cases, you should reduce the number of covariates included in the model and examine the regression output in the .log file for potential problems.

## 7. Why do my intention-to-treat (ITT) impact results differ depending on whether or not I conduct a complier average causal effect (CACE) analysis?

This could occur due to differences in the samples used for the ITT analyses. *If CACE analyses are conducted, both the ITT and CACE estimates are obtained using the sample with nonmissing data for both the outcome variables and the intervention receipt (GOT\_TREAT) variables.* RCT-YES adopts this procedure so that the ITT and CACE estimates can be compared using a common sample. By contrast, if CACE analyses are not specified, the ITT estimates are obtained using the sample with nonmissing data for the outcome under consideration (which is likely to be a larger sample).

## 8. Why do my results change using the finite-population (FP) and super-population (SP) models?

The FP and SP model results will likely be similar for non-blocked designs (Designs 1 and 3) but could differ for blocked designs (Designs 2 and 4). For blocked designs, the variance estimators for the FP model are based solely on variances of the mean outcomes within blocks, whereas the variance estimators for the SP model are based primarily on the variances of the impact estimates across blocks. Thus, these variances expressions can produce different results. There is no “right” answer as to which estimator should be preferred. Rather, the choice will depend on the specific evaluation context. The *RCT-YES* Methods Appendix provides a detailed discussion of the assumptions underlying each method and issues to consider in selecting the appropriate model.

## 9. How does *RCT-YES* treat missing outcomes, covariates, and subgroups?

**Missing outcomes.** *RCT-YES* estimates impacts using only observations with nonmissing values for the outcome under investigation. This method is known as “*case deletion*.” The program does not impute (fill in) outcomes for those with missing data. Case deletion yields unbiased impact estimates if the missing data mechanisms are random for both the treatment and control groups. *RCT-YES* users can include weights in the input data file to adjust for missing data. If CACE analyses are specified, both the ITT and CACE analyses are conducted using observations that have nonmissing data for both the outcome and intervention receipt variables under investigation to align the impact findings from the two sets of analyses.

**Missing covariates.** For regression analyses, the program replaces missing values of covariates with the covariate average calculated from the sample with nonmissing covariate values, separately for the treatment and control groups. The replacements are done separately for each specified outcome variable (which may have different percentages and patterns of missing data). By default, this imputation is done only if the covariate is missing for 30 percent or fewer cases for both the treatment or control groups separately; otherwise the covariate is dropped from the analysis (see the response to Question 5 above). The program never excludes an observation with available outcome data but missing covariate data.

A somewhat different approach is used for the baseline equivalence analysis. To analyze baseline equivalence for a specific covariate, *RCT-YES* drops observations that have missing data for that covariate. It does not use averages to replace missing covariate values.

**Missing subgroups.** *RCT-YES* excludes cases that have missing values for the categorical subgroup variables. For example, if gender is missing for an individual, *RCT-YES* will exclude this observation from the analysis when estimating impacts for boys and girls (even if that observation has available outcome data).

## 10. How does *RCT-YES* weight clusters?

By default, *RCT-YES* weights clusters equally in the analysis regardless of their size. Weights can be included in the analysis to override these defaults (see Chapter 2 of the User's Manual for a discussion of weighting).

## 11. How does *RCT-YES* weight blocks?

For non-clustered designs, the program weights blocks by the total number of treatment and control group individuals with available data in the block. For clustered designs, the program weights blocks by the total number of treatment and control group clusters in the block. Weights can be included in the analysis to override these defaults (see Chapter 2 of the User's Manual for a discussion of weighting). In cases where the user supplies weights, *RCT-YES* will not additionally reweight the blocks (or clusters).

## 12. Can I include different covariates for different outcomes in the same domain?

*RCT-YES* does not allow this. If you want to specify different covariates for different outcomes in the same domain, you should specify different domains for the outcomes. Note, however, that within a domain, covariates for full sample analyses can differ from those for subgroup analyses.

## 13. What is the scale of the impact estimates for binary (dichotomous) outcomes?

*RCT-YES* outputs impact estimates for dichotomous outcomes in terms of percentage points rather than fractions by multiplying the estimates by 100. This will affect the treatment and control means, impact estimates, and standard errors, but not p-values.

## 14. How does *RCT-YES* handle covariates that are collinear?

If you supply a set of covariates that are collinear by construction, *RCT-YES* will drop one of the collinear variables. For example, suppose you want to control for grade level by including a set of mutually exclusive covariates for grade: grade 1, grade 2, grade 3, and grade 4. If all four indicators are entered, *RCT-YES* will drop one from the impact analysis to avoid collinearity. Note that the program will not provide a warning that a covariate is dropped. However, you can view the .log file to see which covariates were dropped. In addition, Table 6 in the .html file will show squared partial correlations of "1.00" for the set of collinear covariates. Note that all covariates will be used in the baseline equivalence tests.

**15. How does *RCT-YES* handle covariates for matched pair and other blocked designs with SUPER\_POP=1?**

*RCT-YES* estimates impacts and standard errors in two stages for these designs. In the first stage, impacts are estimated for each block. If there are sufficient degrees of freedom, this model will include covariates (centered at the block means). However, for matched pair designs and other blocked designs with small numbers of units per block, covariates will be excluded from the first-stage model. In the second stage, the block-level impact estimates from the first stage are regressed on an intercept and block-level covariates (if specified). The .log file will show the covariates that were included in each stage. In addition, Tables 6 and 7 of the .html file will indicate model covariates from the second-stage regression.

**16. Can I assume a finite-population model for matched pair designs?**

*RCT-YES* estimates impacts and standard errors for matched pair designs assuming a super-population framework. This is because each block (pair) contains only 1 treatment and 1 control group observation, and thus, there are not sufficient degrees of freedom to estimate the default finite-population specification (which requires at least two treatments and two controls per block). However, you can use the BLOCK\_FE=1 option to include block indicators in the model but not block-by-treatment interaction terms. For this specification, you must set the MATCHED\_PAIR input to 0.

## Graphing Issues

### 1. Why do I get hung up (stuck) in the R workspace after leaving *RCT-YES-Graph*?

This can happen if you do not properly exit *RCT-YES-Graph*, but for example, minimize the dashboard instead. You can fix this while stuck in the R workspace by pressing the esc key or the **stop** button in the toolbar, which should allow you return to the interactive prompt in R.

### 2. Why does the error “ERROR: [on\_request\_read] connection reset by peer” appear in the R workspace after exiting *RCT-YES-Graph*?

This can happen if you use Internet Explorer because *RCT-YES-Graph* uses browser features that Internet Explorer does not have. These errors can be ignored and do not affect the graphs or output files.

### 3. Why does my graph not appear in the dashboard?

There are several possible reasons for this:

- *The .csv file for the graph produced by the R or Stata computer program is not specified or is corrupted.* *RCT-YES-Graph* must use a .csv file produced by the *RCT-YES* computer program that was used to conduct the analysis. A .csv file in a different format will not work. Thus, you should make sure that the input file you specify was produced by a successful run of the R or Stata computer program that created the associated .html file.
- *You did not specify at least one outcome variable and at least one subgroup for the graph.* This can occur if you change the graph type or change the outcome domain, in which case the outcome and subgroup variables are reset to blank and need to be re-specified.
- *You did not click the Submit button.* You must click this button to produce the graphs. You only need to click this once; subsequent changes to the graph will automatically update in the dashboard.

### 4. Is there a way to change the order that the outcome and subgroup variables are displayed in the graph?

No. The graphs are always ordered by outcome variable first and then the subgroup variable.

**5. Is there a way to change the order of the lists of outcome and subgroup variables in the dashboard inputs?**

No. The lists of outcomes and subgroups in the dashboard follow the order that was specified in the RCT-YES interface screens.

**6. Is there a way to plot outcomes or subgroups from different domains in the same graph?**

No. You can only graph outcomes and subgroups from the same domain.

**7. Is there a way to change graph colors or fonts?**

Not currently, but color and font options may be added in future versions.

**8. Can I save my graph inputs?**

No. Once you exit the program, the graph inputs will be lost. You will need to reenter the inputs when you go back into the program.

**9. Why is the width of the graphs in the dashboard wider than in the output files?**

This is because the page sizes differ in the dashboard screen display and in the .png and .pdf files. When downloading the plots to a file, it is a good idea to first open the .png or .pdf file before saving the file to make sure that the layout of the graph is as expected.