

Appendix: Monte Carlo experiment for an RCT with outliers

In this illustrative example, there is a parent population each member of which has his or her own treatment effect; these are continuously distributed with a shifted lognormal distribution with zero mean so that the population ATE is zero. The individual treatment effects β are distributed so that $\beta + e^{0.5} \Lambda(0,1)$, for standardized lognormal distribution Λ . In the absence of treatment, everyone in the sample records zero, so the sample average treatment effect in any one trial is simply the mean outcome among the n treatments. For values of n equal to 25, 50, 100, 200, and 500 we draw from the parent population 100 trial samples each of size $2n$; with five values of n , this gives us 500 trial samples in all; because of sampling the true ATE's in each trial sample will not be zero. For each of these 500 samples, we randomize into n controls and n treatments, estimate the ATE and its estimated t -value (using the standard two-sample t -value, or equivalently, by running a regression with robust t -values), and then repeat 1,000 times, so we have 1,000 ATE estimates and t -values for each of the 500 trial samples. These allow us to assess the distribution of ATE estimates and their nominal t -values for each trial.

The results are shown in Table A1. Each row corresponds to a sample size. In each row, we show the results of 100,000 individual trials, composed of 1,000 replications on each of the 100 trial (experimental) samples. The columns are averaged over all 100,000 trials.

Table A1: RCTs with skewed treatment effects

Sample size	Mean of ATE estimates	Mean of nominal t -values	Fraction null rejected (percent)
25	0.0268	-0.4274	13.54
50	0.0266	-0.2952	11.20
100	-0.0018	-0.2600	8.71
200	0.0184	-0.1748	7.09
500	-0.0024	-0.1362	6.06

Note: 1,000 randomizations on each of 100 draws of the trial sample randomly drawn from a lognormal distribution of treatment effects shifted to have a zero mean.

The last column shows the fractions of times the null that is true in the population is rejected in the trial samples and is our key result. When there are only 50 treatments and 50 controls (row 2), the (true) null is rejected 11.2 percent of the time, instead of the 5 percent that we would like and expect if we were unaware of the problem. When there are 500 units in each arm, the rejection rate is 6.06 percent, much closer to the nominal 5 percent.

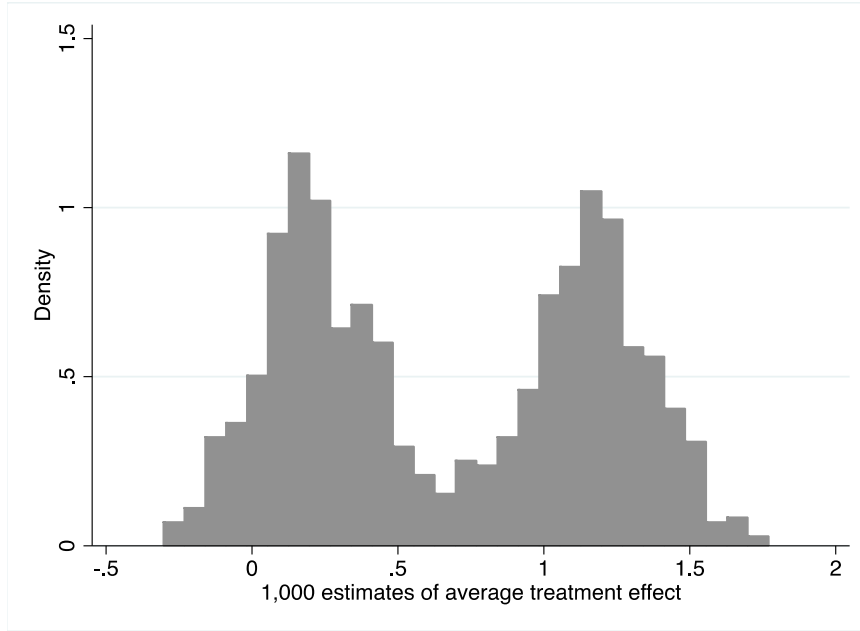


Figure A1: Estimates of an ATE with an outlier in the trial sample

Figure A1 illustrates the estimated ATEs from an extreme trial sample from the simulations in the second row with 100 observations in total; the histogram shows the 1,000 estimates of the ATE for that trial sample. This trial sample has a single large outlying treatment effect of 48.3; the mean (s.d.) of the other 99 observations is -0.51 (2.1); when the outlier is in the treatment group, we get the observations around right-hand mode, when it is in the control group, we get the left-hand mode.