Does Blocking Reduce Attrition Bias?

Thad Dunning, Department of Political Science, Yale University
thad.dunning@yale.edu

Several political scientists have recently drawn attention to the merits of blocking in experimental studies.

Experiments allow for unbiased estimation of average causal effects for the experimental study group, as long as attrition and some other threats to internal validity do not arise. However, estimators may be more or less precise. This is where blocking comes in. Here, units are grouped into strata, or blocks, and then randomized to treatment and control conditions within these blocks.

For instance, experimental units may be grouped into pairs with similar incomes, past voting histories, or values of other variables that may predict an outcome. In such "matched pair" designs, one member of each pair is randomly assigned to treatment, while the other is randomly assigned to control.

As Moore (2010) and others have pointed out, blocking can be a worthwhile strategy, especially in smaller experiments. Blocking is most beneficial when units are relatively homogeneous (with respect to the outcome) within blocks and heterogeneous across blocks. Thus, if investigators can identify variables that are good predictors of the outcome, blocking units before randomization may increase the precision of treatment effect estimators.

Matched-Pair Designs and Attrition Bias

Yet, can blocking overcome bias from attrition or non-compliance? Following King et al. (2007), Moore (2010) suggests that "Through blocking, design can anticipate and overcome a frequent field experiment reality: some units may be compromised during an experiment, and they and their blockmates can be excluded from analysis without endangering the entire randomization."

Attrition occurs when some units are lost to follow-up—that is, when the outcomes for some units are not recorded, for a variety of possible reasons. Such attrition can be devastating to causal inference when it is associated with treatment assignment and is a function of potential outcomes (that is, the outcomes that we would observe for each unit if it were assigned to treatment or control).

For example, the subjects most likely to drop out of a medical trial may be those assigned to treatment, for whom the treatment does not appear to be working. In this case, comparing health outcomes of subjects who remain in the treatment group to subjects assigned to control may overstate the efficacy of treatment, since subjects for whom the treatment did not seem to work dropped out of the trial.

Similar issues can arise with non-compliance, which occurs when units assigned to treatment receive the control, or vice versa. For instance, in randomized public policy programs, politicians may be tempted to administer the treatment to control units (King et al. 2007). If we conduct analysis by treatment received—comparing those who receive the treatment to those that receive the control—we risk bias. This is because units that crossover from control to treatment, or vice versa, may be unlike those that don't, in ways that matter for outcomes.[1]

Yet if attrition or non-compliance are functions of potential outcomes, dropping blocked pairs does not eliminate the bias. To see this, imagine that there are two types in the experimental population:

(1) Units that will be compromised (e.g., be subject to attrition) if they are assigned to treatment; and
(2) Units that will not be compromised, whether assigned to treatment or control.

If subjects are randomized in pairs, and we throw out all the pairs in which the treatment unit is compromised, then we know that all of the treatment units that are thrown out are of type (1).

Yet, the control units in the excluded pairs may include both types (1) and (2). For these control units, we don't get to observe the counterfactual response to treatment assignment—that is, whether the control unit would stay in the experiment or not if assigned to treatment.

This can lead to bias, if the propensity to be compromised is related to potential outcomes, because the control group contains types (1) and (2), while the treatment group contains only type (1).

A Numerical Example

A numerical example may help to illustrate this point. Suppose that in a small experiment, there are four units that will be blocked on a single covariate and then randomized to treatment within blocked pairs.

[1] Non-compliance is typically not as harmful for inference as attrition: we can always do intention-to-treat analysis, which estimates the effect of treatment assignment. There are also sensible procedures for estimating the effect of treatment on compliers, using treatment assignment as an instrumental variable.

The table below shows the covariate value for each unit; the potential outcome under control, which is denoted $Y(0)$; and the potential outcome under treatment, which is denoted $Y(1)$. The final column records whether or not the unit will be lost to attrition if assigned to treatment.

Covariate value	$Y(0)$	$Y(1)$	Lost to attrition if assigned to treatment?
2	1	2	NO
2	3	3	NO
3	4	6	YES
3	2	3	NO
Mean outcome:	$\frac{10}{4}$	$\frac{14}{4}$	

The final row of the table shows the mean outcome if all units were assigned to control and the mean outcome if all units were assigned to treatment. The average causal effect is defined as the difference between these quantities.

Here, the true average causal effect is $\frac{14}{4} - \frac{10}{4} = 1$.

Now, suppose we randomize in pairs defined by values on the covariate: for instance, we flip a coin to decide which of the two units with covariate value 2 goes into treatment, and similarly for the two units with covariate value 3.

This implies that there are 4 possible random assignments. The realized outcomes under each assignment are listed in the tables below. The pairs that will be dropped due to attrition are indicated with diagonal slashes.

For each randomization, the estimated average causal effect—that is, the mean outcome in the treatment group minus the mean outcome in the control group, after dropping the pairs in which there was attrition—is listed below each table.

	Covariate value	Realized outcome (control units)	Realized outcome (treatment units)
Randomization 1	2		2
	2	3	
	~~3~~		~~6~~
	~~3~~	~~2~~	
	Mean outcome:	$\frac{3}{1}$	$\frac{2}{1}$

Estimated average causal effect: $2 - 3 = -1$

	Covariate value	Realized outcome (control units)	Realized outcome (treatment units)
Randomization 2	2		2
	2	3	
	3	4	
	3		3
	Mean outcome:	$\frac{7}{2}$	$\frac{5}{2}$

Estimated average causal effect: $\frac{5}{2} - \frac{7}{2} = -1$

	Covariate value	Realized outcome (control units)	Realized outcome (treatment units)
Randomization 3	2	1	
	2		3
	3	4	
	3		3
	Mean outcome:	$\frac{5}{2}$	$\frac{6}{2}$

Estimated average causal effect: $\frac{6}{2} - \frac{5}{2} = \frac{1}{2}$

	Covariate value	Realized outcome (control units)	Realized outcome (treatment units)
Randomization 4	2	1	
	2		3
	~~3~~		~~6~~
	~~3~~	~~2~~	
	Mean outcome:	$\frac{1}{1}$	$\frac{3}{1}$

Estimated average causal effect: $3 - 1 = 2$

The average estimated causal effect is the average over all of the estimates obtained after each of the 4 possible randomizations, that is,

$$\frac{-1 - 1 + \frac{1}{2} + 2}{4} = \frac{1}{8}.$$

This estimator is biased by about -88%, since the true average causal effect is 1. Thus, we have a substantial bias induced by a relatively small association between potential outcomes and the propensity to be compromised. The logic carries through to larger experiments and more complicated examples. For instance, one can construct illustrations in which the average estimate of the average causal effect is negative, even though the individual causal effect is non-negative for every unit and positive for many units.

Notice that dropping pairs in which one unit is subject to attrition also does not recover an unbiased estimate of the causal effect for the first, second, and fourth units—that is, the units not subject to attrition. For these units, the average causal effect is $\frac{8}{3} - \frac{6}{3} = \frac{2}{3}$, so the average estimate of $\frac{1}{8}$ is quite a bit off. Whether the average causal effect defined only for non-attriters is an interesting quantity is a different, perhaps substantive question. My point here is simply that randomizing in pairs, and dropping pairs in which there is attrition, does not recover this effect, in this example and many others.

One can construct additional examples, in which there is a perfect correlation between blocked covariates and the propensity to be subject to attrition if assigned to treatment; in such examples, the strategy of dropping pairs may give unbiased estimates of the average causal effect, defined only for units not subject to attrition. In practice, however, this strategy will depend on a strong assumption about unobservables: we have to assume that all units with the same values of the blocked covariate respond similarly to treatment assignment. Yet, with just one randomization, we do not get to observe the counterfactual response to treatment assignment of control units, that is, whether they would stay in the study if assigned to treatment. So this supposition is not amenable to empirical evaluation.

Returning to the example above, notice that these problems go away if there is no attrition. Indeed, if we include all of the outcomes in the calculations, the average of the four estimates equals 1, just as it should—because without attrition, the difference of means is an unbiased estimator for the average causal effect.

Blocking Increases Efficiency; It Does Not Reduce Bias

It may be worth making one final clarifying point about the role of bias in small, unblocked experiments. As Moore (2010) points out, one of the advantages of blocking is that it can reduce the chances that the treatment and control groups are unbalanced, at least on the blocked covariates. This is especially useful in small experiments, where the luck of the draw implies that there may be substantial imbalances across treatment and control groups on measured covariates.

However, random imbalances that arise between treatment and control groups do not lead to confounding, in the statistical sense—contrary to Moore's (2010) claim that blocking "reduces the bias in causal estimates that comes from comparing a treatment group [with covariate values] of 2, 2, 3 with a control group of 3, 4, 4." It is true that in the data, on any particular draw, covariates such as past voting history may be associated empirically with (realized) treatment assignment. Yet, bias is not the issue here: randomization ensures balance on pre-treatment covariates, in expectation. The issue is instead sampling error—that is, the random imbalances induced by random sampling of the treatment and control groups from the experimental study group—which is a different concept.

In sum, blocking can increase statistical efficiency by balancing the treatment and control groups on blocked covariates, but this is not a means of reducing bias.

Conclusion

Attrition can be quite destructive for experimental inference, when the propensity to be compromised or lost to follow-up is associated with potential outcomes. It would therefore be useful to have a way to adjust this problem away.

Yet, it appears that blocking, and dropping pairs in which one of the pairs is compromised, does not get us around this thorny problem. Blocking can be an effective tool to increase the precision of treatment effect estimators—but whether it reduces bias depends on substantive assumptions that are not justified by randomization or other aspects of the experimental design. Other strategies—for example, investing resources in tracking down units that are lost to follow up, or just a random sample of those units (Geng et al. 2008)—may offer better alternatives for reducing attrition bias.

References

Geng, Elvin H., Nneka Emenyonu, Mwebesa Bosco Bwana, David V. Glidden, and Jeffrey N. Martin. 2008. "Sampling-Based Approach to Determining Outcomes of Patients Lost to Follow-Up in Antiretroviral Therapy Scale-Up Programs in Africa." *Journal of the American Medical Association* 300(5): 506-507.

King, Gary, Emmanuela Gakidou, Nirmala Ravishankar, Ryan T. Moore, Jason Lakin, Manett Vargas, Martha Mara Téllez-Rojo, Juan Eugenio Hernández Ávila, Mauricio Hernández Ávila and Héctor Hernández Llamas. 2007. "A 'Politically Robust' Experimental Design for Public Policy Evaluation, with Application to the Mexican Universal Health Insurance Program." *Journal of Policy Analysis and Management* 26(3): 479-509.

Moore, Ryan. 2010. "Blocking Political Science Experiments: Why, How, and Then What?" *Newsletter of the APSA Experimental Section* 1(1): 3-5.