Proposal for JEPS special issue:
"A Practical Guide to Dealing with Attrition in Political Science Experiments"

October 31, 2021

1 Overview of the problem and our contribution

Our contribution is motivated by the observation that applied researchers have few tools—and no easily accessible overview—to pre-empt, detect and address attrition. Even while a cottage industry has developed around improving other areas of experimental design (Jerit, Barabas and Clifford, 2013; Dafoe, Zhang and Caughey, 2018; Coppock, 2019; Mummolo and Peterson, 2019; Brutger et al., 2022), the issue of attrition—missingness on the outcome variable—has been mostly ignored by practitioners (Gerber et al., 2014; Zhou and Fishbach, 2016). While standards of "best practices" have been set (Gerber et al., 2014, 97), our reading of the state of the field in experimental political science is that attrition is rarely inspected or discussed in published works, despite the issues caused if the missingness reveals something about the potential outcomes of those who drop out of the study (Druckman et al., 2011, 19). To the extent that important advances have been made, they focus almost exclusively on *ex-post* solutions such as double-sampling or extreme value bounds, which, though valuable, do not help with the issue of easily *identifying* attrition that results in threats to inference or *preventing* it from occurring in the first place in the design stage.

Our practical guide to assessing the type, scope and impact of attrition in political science experiments—alongside an accompanying open-source `R` package—provides a helpful method for visualizing attrition across treatment conditions and over-time within the survey instrument, as well as utilizing and comparing balance tests at precise moments in the study and, for those studies suffering from problematic attrition, incorporating estimation and visualization of Manski bounds. Our method will assist in the design and piloting stage by pinpointing when exactly attrition occurs, as well as in the analysis stage by providing a scalable set of "best practices" for experimentalists in assessing whether attrition poses threats to either internal or external validity and transparently presenting relevant data. We aim to make such procedures standard operating procedure for applied researchers as one way of increasing transparency in the field.

2 Attrition in Experiments

Several things are evident from a brief review of the literature and patterns of publishing in political science. First, it is obvious that, conceptually, attrition might pose a problem for inference, but extant work demonstrates that in practice our fears may be justified (even based on reported attrition, which is presumably lower than overall attrition, Musch and Reips, 2000; Zhou and Fishbach, 2016). Second, despite the importance of detecting attrition, it appears as though "ignorance is bliss" for most researchers: Gerber et al. (2014, 88) find that 58% of the experimental articles sampled in their review did not even report the number of subjects in each treatment group for which there is missing outcome data (see also Mutz and Pemantle 2015, 13 and Zhou and Fishbach 2016, 495). Finally, the solutions we have to address it are inadequate on their own since they typically focus on *ex-post* solutions to be implemented in the analysis stage.

Current solutions tend to cluster around two approaches, the first of which focuses on reducing attrition in the design stage, testing and evaluating specific ideas such as using different survey modes (Morrison et al., 1997) or monetary incentives (Göritz, 2014). While much of this work focuses on panel/longitudinal settings, to the extent that the lessons are applicable to online experiments they still require a way to clearly and easily understand when and why attrition is occurring. After all, using incentives to lower attrition from some counterfactual baseline does not solve one's inferential problems if it still occurs and is causally related to treatments. A second cluster of approaches addresses attrition *ex-post*, such as through the use of extreme value bounds (or "Manski Bounds," see Manski, 1995, 2009 and Coppock 2021, 333), inverse probability weighting (Wooldridge, 2007), double-sampling (Gerber and Green, 2012, 236-241) or some combination (Gomila and Clark, 2020; Coppock et al., 2017).

In sum, attrition represents a significant threat to inference and current approaches to addressing it are either *ex-post*—requiring rather strong assumptions in order to estimate treatment effects—or involve implementing design choices that rely on confusing conventional wisdom as to what "works." Our argument is that in both cases, researchers would benefit from a way to understand precisely when and why attrition is occurring in the first place. Our proposed set of solutions, detailed below, aid in the design stage by allowing researchers to work to easily pinpoint and design around problematic questions and treatments—heading off attrition in studies before they are truly fielded—and in the analysis stage by providing an easily understandable method for understanding when attrition is occurring and when it poses threats to inference.

3 Our Contribution

Combining the dictum of Fisher to "analyze as you randomize" with the advice of Coppock (2021) to "visualize as you randomize," our contribution is to offer experimentalists a way to do both. Specifically, our goal is to provide a "holistic approach" to addressing attrition for applied researchers. It will begin with a quantitative literature search of the use of attrition—demonstrating the scope of the problem—to motivate both a checklist of practical solutions for addressing and preempting attrition through piloting and an accompanying `R` package that provides diagnostic visualizations and corresponding tests. Our proposed `R` package will provide a question-by-question over-time snapshot of an experiment along an axis with the corresponding amount of attrition at each of these moments. Researchers need only input (and assign temporal ordering to) their experimental data.

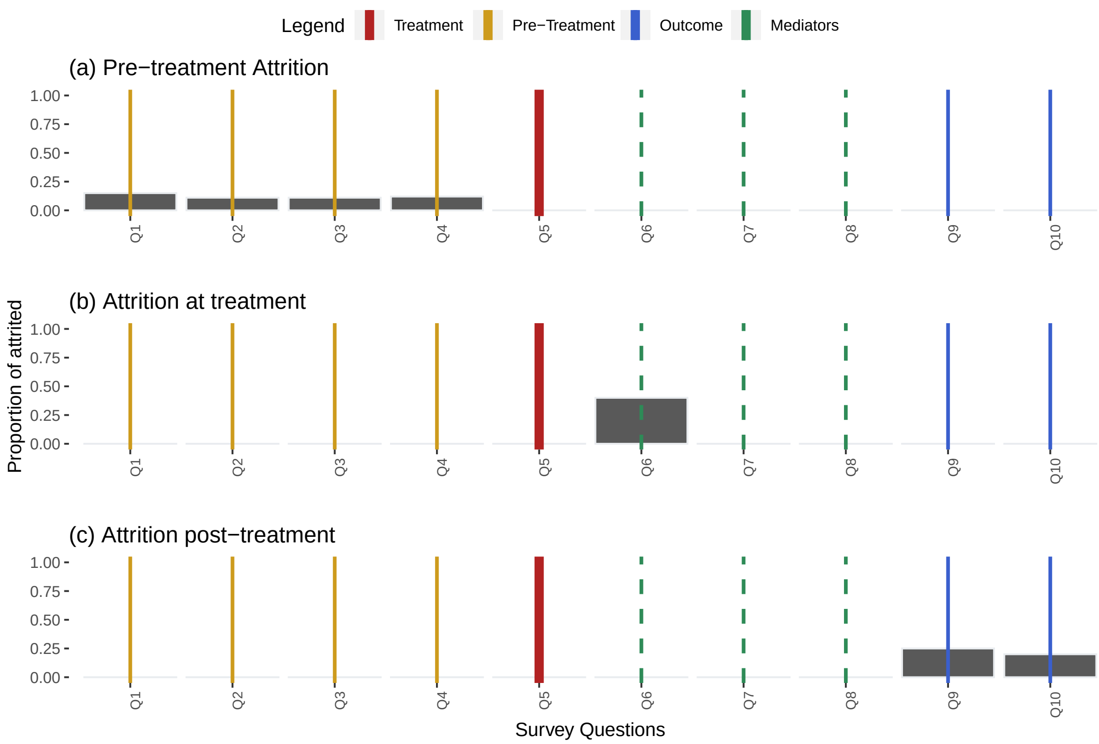

Figure 1: **Attrition timeline plot.** Items in study are temporally ordered from left to right on x-axis; y-axis represents proportion of attrited respondents. Package options also include plotting number attrited, as well as cumulative proportion and number attrited.

Figure 1 presents a toy example to illustrate our approach, with each row representing data from a different survey experiment, each with one treatment and several mediators and outcome questions (multiple treatments can be easily accommodated). X-axis ticks represent items presented to respondents in the order they appeared, with vertical lines indicating pre-treatment, treatment, mediator/moderator, dependent and post-dependent variables. Y-axes indicate the proportion of attrition that occur at that point in the survey. Panel (a) illustrates a study in which attrition occurs entirely pre-treatment, while Panel (b) shows attrition occurring immediately after treatment assignment, and cautions closer inspection around whether missingness may be potential outcome correlated (and thus a threat to inference). Panel (c) likewise suggests missingness that, if left unaccounted for, might lead to biased estimates of an average treatment effect, though attrition occurs later in the experiment.

If the researcher finds evidence of (only) pre-treatment attrition, internal validity is not threatened (since non-attriters can still be randomly assigned to arms, probabilistic equivalence can be achieved), but the remaining sample may no longer reflect the original population. Further checks include verifying—and if need be, adjusting—to ensure that *consent* questions and *pre-treatment questions* are not sensitive or aversive in some way. If attrition mainly appears post-treatment, such as in (b) or (c) of Figure 1, we advise visualization of and detection for differential missingness by respondent-question within treatment arms, which we offer as a secondary visual diagnostic in our accompanying `R` package. A "know-it-when-you-see-it" graphic, the researcher can use a visual check for whether attrition patterns under each treatment arm appear dissimilar.[1] Another such inspection would entail checking for pre-treatment covariate balance across respondents in each arm *at the point of highest post-treatment attrition*. Our package offers balance checks at user-provided "moments" in the experiment to allow exploration of the extent to which attrition is treatment-induced. Finally, we offer estimation of Manski bounds for main treatment effects for researchers interested in ascertaining to what extent attrition might affect estimated average treatment effects, which may influence the order of priority for updating the study to minimize attrition.

Upon diagnosing (1) whether attrition occurs, and if so (2) *where* it occurs, and (3) whether it presents as a threat to inference, researchers can edit and alter experimental designs accordingly for a re-pilot and/or soft launch. (1)-(3) can be repeated as needed and as is practical to budget. For researchers for whom follow-ups may not be possible, our check-list and software may aid in formation of a discussion on the extent to which attrition might affect study results and findings and offers tests of balance and Manski bounds on main treatment effects.

[1] This can be considered a verification of "equal attrition rates" which would satisfy internal validity for respondent subpopulation (see Ghanem, Hirshleifer and Ortiz-Becerra (2019) working paper).

References

Brutger, Ryan, Josh Kertzer, Jonathan Renshon, Dustin Tingley and Chagai Weiss. 2022. "Abstraction and Detail in Experimental Design." *American Journal of Political Science* .

Coppock, Alexander. 2019. "Generalizing from Survey Experiments Conducted on Mechanical Turk: A Replication Approach." *Political Science Research and Methods* 7(3):613–628.

Coppock, Alexander. 2021. Visualize as You Randomize. In *Advances in Experimental Political Science*, ed. James N. Druckman and Donald P. Green. New York, NY: Cambridge University Press pp. 320–337.

Coppock, Alexander, Alan S Gerber, Donald P Green and Holger L Kern. 2017. "Combining double sampling and bounds to address nonignorable missing outcomes in randomized experiments." *Political Analysis* 25(2):188–206.

Dafoe, Allan, Baobao Zhang and Devin Caughey. 2018. "Information equivalence in survey experiments." *Political Analysis* 26(4):399–416.

Druckman, James N, Donald P Green, James H Kuklinski and Arthur Lupia. 2011. An introduction to core concepts. In *The Cambridge handbook of experimental political science*, ed. James N Druckman, Donald P Green, James H Kuklinski and Arthur Lupia. New York, NY: Cambridge University Press pp. 15–26.

Gerber, Alan S and Donald P Green. 2012. *Field experiments: Design, analysis, and interpretation.* WW Norton.

Gerber, Alan S, Kevin Arceneaux, Cheryl Boudreau, Conor Dowling, Sunshine D Hillygus, Thomas Palfrey, Daniel R Biggers and David J Hendry. 2014. "Reporting guidelines for experimental research: A report from the experimental research section standards committee." *Journal of Experimental Political Science* 1(1):81–98.

Ghanem, Dalia, Sarojini Hirshleifer and Karen Ortiz-Becerra. 2019. "Testing Attrition Bias in Field Experiments.".

Gomila, Robin and Chelsey S Clark. 2020. "Missing data in experiments: Challenges and solutions." *Psychological Methods* .

Göritz, Anja S. 2014. Determinants of the starting rate and the completion rate in online panel studies. In *Online Panel Research: A Data Quality Perspective*, ed. Mario Callegaro, Reg Baker, Jelke Bethlehem, Anja S. Göritz, Jon A. Krosnick and Paul J. Lavrakas. Wiley Online Library pp. 154–170.

Jerit, Jennifer, Jason Barabas and Scott Clifford. 2013. "Comparing contemporaneous laboratory and field experiments on media effects." *Public Opinion Quarterly* 77(1):256–282.

Manski, Charles F. 1995. *Identification problems in the social sciences.* Cambridge, MA: Harvard University Press.

Manski, Charles F. 2009. *Identification for prediction and decision.* Cambridge, MA: Harvard University Press.

Morrison, Theodore C, Dennis R Wahlgren, Melbourne F Hovell, Joy Zakarian, Susan Burkham-Kreitner, C Richard Hofstetter, Donald J Slymen, Kristen Keating, Stergios Russos and Jennifer A Jones. 1997. "Tracking and follow-up of 16,915 adolescents: Minimizing attrition bias." *Controlled Clinical Trials* 18(5):383–396.

Mummolo, Jonathan and Erik Peterson. 2019. "Demand effects in survey experiments: An empirical assessment." *American Political Science Review* 113(2):517–529.

Musch, Jochen and Ulf-Dietrich Reips. 2000. A brief history of Web experimenting. In *Psychological experiments on the Internet*, ed. Michael H. Birnbaum. Elsevier pp. 61–87.

Mutz, Diana C and Robin Pemantle. 2015. "Standards for experimental research: Encouraging a better understanding of experimental methods." *Journal of Experimental Political Science* 2(2):192–215.

Wooldridge, Jeffrey M. 2007. "Inverse probability weighted estimation for general missing data problems." *Journal of econometrics* 141(2):1281–1301.

Zhou, Haotian and Ayelet Fishbach. 2016. "The pitfall of experimenting on the web: How unattended selective attrition leads to surprising (yet false) research conclusions." *Journal of personality and social psychology* 111(4):493–504.