

A EXCEPTIONS FOR DATA OR CODE THAT IS CONFIDENTIAL, PROPRIETARY, OR A COSTLY ORIGINAL CONTRIBUTION

The recommendations for journals in 4.2 applies to cases where the data and code are not proprietary or confidential. If data or code is confidential, proprietary, or represents a very costly original contribution, the appropriate replication norms need to be more nuanced. As long as scholars and journal editors make a sincere effort to work through these issues we will be in a much improved situation from where we are now. However, I outline below some specific suggestions.

A.1 CONFIDENTIALITY AND PROPRIETARY DATA

Some data has issues of confidentiality. Protection of human subjects may require that individual respondents cannot be identified from the data. The authors may have agreed to certain restrictions in order to access their data.

There are a range of potential responses, from least to most limiting in transparency. Authors could strive to be as transparent as possible, and explain to journals and readers the limits on transparency that they choose.

Potential responses include, from less to more restrictive:

1. Remove or scramble identifying labels. Often ID labels are unnecessary for analysis, and after removal individuals can no longer be identified.
2. Remove or scramble other identifying variables, such as I.P. address, age, etc...
3. Authors could share the code that they used to carry out the analysis and a dummy dataset. A dummy dataset is a dataset that involves fake data, but otherwise tries to preserve some features of the data. For example, a version of the actual dataset could be provided in which each variable that may help identify individuals is randomly resorted; this would allow readers to investigate questions about the distributions of the confidential variables.
4. Share other non-identifying statistics in the online appendix, such as density plots and covariance matrices.

For proprietary data, the authors could still provide complete replication code and explain how one can acquire access to the data. Often creative solutions exist for restricted data; some military data is not allowed to be shared with the public, but may be shared more restrictively through military libraries. And as with confidential information, as much information could be provided as possible, such as the covariance matrix and a dummy version of the dataset.

A.2 ORIGINAL COSTLY DATA (OR CODE) ACQUIRED BY THE AUTHORS

Our profession does not reward the collection and provision of data adequately. Most scholars who put together new data understandably want to have restricted access to the data for their publication purposes. The issue is similar to that of patent law: as a discipline we want to incentivize scholars to collect valuable new data and our primary instrument for doing so is the opportunity of data collectors to be the first to analyze their data; however, we also want to allow other scholars to access the data to evaluate the claims of the authors and to contribute to other projects. Currently APSA's [Data Access and Research Transparency](#) guidelines (section 6.6) recommends a one year

Online Appendix [A](#) for:

Dafoe, Allan. 2013. Science Deserves Better: The Imperative to Share Complete Replication Files.

embargo period from the time of publication. Note that even with an embargo, journals could still require that complete replication files be provided to the journal before acceptance. Similar embargo possibilities may be permitted for original costly contributions in code.

B REPLICATION AUDIT

The idea of a “Replication Audit” emerged from conversations at the Editorial Committee Meeting at the *Journal of Peace Research*. The group agreed that strengthening replication practices was an important goal. One way to do this would be for journals to guarantee space for replications, since this would (1) reward the practice of engaging in replication, (2) provide a sense for the reliability of results, and (3) incentivize authors to be more careful in their analyses.

However, this kind of exercise is susceptible to publication bias: if editors prefer to publish replications that are “interesting”, then there will be a bias against replications that find support for the original results, and towards those that claim to overturn the original results. This risks promoting a-theoretic fiddling with analyses in a search for “interesting” results and exaggerated interpretation of results. To address this concern we propose a more systematic approach to replication: the Replication Audit. We recommend that journals commit to a periodic Replication Audit of their publications.

The Replication Audit involves randomly selecting a subset of articles from a set of journals, examining them in a systematic and moderate manner, and reporting the findings. The Replication Audit should be guaranteed space in journals to reduce incentives to exaggerate findings. The Replication Audit will consist of a team of researchers, and each article will be replicated by two separate scholars who are not aware of each other. The lack of individual credit for particular findings will reduce the incentives for individual scholars to exaggerate their replication findings, and the redundant replication will increase the accountability and reliability of each replication.

A second concern facing scholars wishing to engage in replication is that the original authors might resent the scholars who examined their work. Our proposal mitigates this problem because (1) the articles selected for replication will be selected randomly, blind to the identity of the authors, and (2) the respective contributions of members of the team will be kept private. Only the combined output of the entire team will be published.

B.1 REPLICATION TEAM

The Replication Team will consist of 3-8 prominent researchers (“advisors”) who have access to competent graduate students. Each advisor will lead a team of around 3-8 strong graduate students to perform the replications. One or multiple advisors will lead the replication team (the “coordinators”).

B.2 PROCESS

A list of relevant articles will be compiled. For now the relevant scope would be all articles employing statistical analysis or computational theory, within the relevant set of journals during a recent span of time (say 5 years). These articles will be assigned a random number to determine their “priority”. The coordinators will go through the articles for each journal, in descending priority, evaluating whether replication files are available. If they are not, an email will be sent to the authors to ask for their replication files, with one follow up email after three weeks. The team will proceed down the list until they have collected N articles with replication files for each journal. Each article will then be assigned to two members of the replication team, so there will be in total $2N$ replication attempts.

B.3 MINIMAL REPLICATION STANDARDS

The Replication Team will answer the following questions for each article.

- Were replication files publicly available? If not, were replication files available upon request?
 - 0 if not available.
 - 1 if available upon request.
 - 2 if publicly available.
- What was the quality of replication files?
 - 0: A dataset (with all variables needed for the analysis) was available. No analysis code was available.
 - 1: A dataset (with all variables needed for the analysis) and some analysis code were available.
 - 2: A dataset (with all variables needed for the analysis) and complete analysis code were available.
 - 3: Most primary datasets, code to create the final dataset, and complete analysis code were available.
- Nominal Replicability?
 - -1: We were not able to replicate the study due to our technical/software limitations, or overly complex code.⁸
 - 0: We were not able to approximately reproduce the main results.
 - 1: We were able to approximately, but not precisely, reproduce the main results.
 - 2: We were able to precisely reproduce the main results.

B.4 MINIMAL ROBUSTNESS STANDARDS

The Replication Team could also evaluate the articles for major technical errors and robustness to arbitrary aspects of the specification.

- Technical Errors?
 - 0: We identified major⁹ technical errors.
 - 1: We did not identify any major technical errors.
- Narrow Robustness? For this question, scholars will examine whether the main results are robust to sensible modifications of arbitrary aspects of the specification and any errors in implementation. We will try to layout a systematic approach to this question, though given the complexity of analysis this may not be possible.

⁸If in doubt, the analyst should spend at least 10 hours trying to replicate the study before coding -1.

⁹A major technical error is an aspect of the analysis that could have had or did have potentially serious consequences, and that most experts would agree should not have been done.

Online Appendix B for:

Dafoe, Allan. 2013. Science Deserves Better: The Imperative to Share Complete Replication Files.

- 0: One or more key results were driven by a technical error.
- 1: One or more key results were not robust to a sensible modification of an arbitrary aspect of the specification.
- 2: Most or all of the key results seemed to be robust to technical errors and sensible modifications of arbitrary aspects of the specification.

The analysis done for this Replication Audit will be completely documented in code. We will try to standardize appropriate parts of the replication activities. Members are encouraged to document their reasons for their coding decisions. When two members of the team have divergent codings, we will have them discuss by email their reasoning. Copies of these conversations will be preserved in our replication files, so that our process is as transparent as possible. Advisors will help adjudicate divergent codings.

C AVAILABILITY OF REPLICATION FILES

The data about the availability of replication files at *APSR* and *AJPS* was collected as follows by Guadalupe Tuñón (Peter Repucci coded AJPS 2009).

- All publications performing quantitative analyses from *AJPS* and *APSR* since 2010 were downloaded and coded.
- **StateAv**: Does the paper indicate that replication files should be available? Yes or no answer. Yes was given to those papers that indicate data is available upon request or provide an explicit reference to where data can be found.

The subset of papers that did not indicate that replication files should be available were coded as “no” for all subsequent questions. The remainder were coded as follows:

- **UpReq**: Does the paper indicate that data is available upon request? Yes or no answer
- **RepRef**: Does the paper provide a reference to a public source (author’s webpage, dataverse, etc) where data can be found? If it did, the reference was included.
- **RepAv**: For those that indicate replication files are publicly available, were they in fact available? Yes if we were able to access the data where the authors claimed it was or after a search in the authors’ webpages. (Publications for which data was indicated to be available upon request take a value of “no”, since data is not *publicly* available).
- **RepFilesAv**: Are replication files available? An undergraduate RA, Peter Repucci, re-examined all publications for which **StateAv**=No. He searched for replication files on Dataverse, on the authors’ websites, and then elsewhere on the internet. If he found replication files he coded **RepFilesAv**=1, if he could not find replication files after 5 minutes he coded **RepFilesAv**=0. Then, in the R code, **RepFilesAv** is set to **RepAv** for all publications in which **StateAv**=Yes.

D SURVEY OF REPLICATION EXERCISES

5/7/13

Survey | Qualtrics Survey Software

This survey is evaluating the extent to which quantitative studies in political science are replicable. Please answer the following questions with respect to a study that you invested at least a few hours in a replication effort. (If you have attempted to replicate more than one, please reload this page when you are done to fill out another survey for each replication effort.)

**Thank you for your input,
Allan Dafoe
Yale University**

Replication Materials

- No replication materials were provided.
- A limited dataset was provided to perform the analysis, but no analysis code.
- A limited dataset was provided to perform the analysis and some analysis code.
- A limited dataset was provided to perform the analysis and complete analysis code.
- Many primary datasets were provided, as well as code to create the final dataset from the primary datasets and code to analyze the data.
- Other: please explain

>>

Please describe the results of your replication efforts (select all that apply)

- I was not able to approximately reproduce the main results.
- I was able to approximately reproduce the main results.
- I was able to precisely reproduce the main results.
- I found one or more major technical errors, though these didn't change the main results.
- I found that one or more key results were driven by a technical error.
- I found that one or more key results were driven by an arbitrary aspect of their analysis.
- I found that one or more key results were fragile in a manner that would lead an impartial scholar to substantially discount the value of the original study.
- Most or all of the key results were robust.

Please describe any other features of your replication exercise that you think might be relevant:

Journal:

Year:

First Author (LastName, FirstName):

Other necessary information to identify the publication:

Respondents were first given the option to select from six characterizations of the availability of replication files. The responses for the full sample are displayed in Figure 2; responses from the PolMeth listserv are similar, but slightly more negative, than those from my class and Gary King’s class. Ignoring the “Other” category, 36% of these respondents reported that complete data files and replication code were available (responses 4 or 5), whereas 64% reported that some important element of the replication files were missing (responses 1, 2 or 3). This corresponds closely to the previous finding that 38% of *APSR* and *AJPS* statistical publications make replication files available.

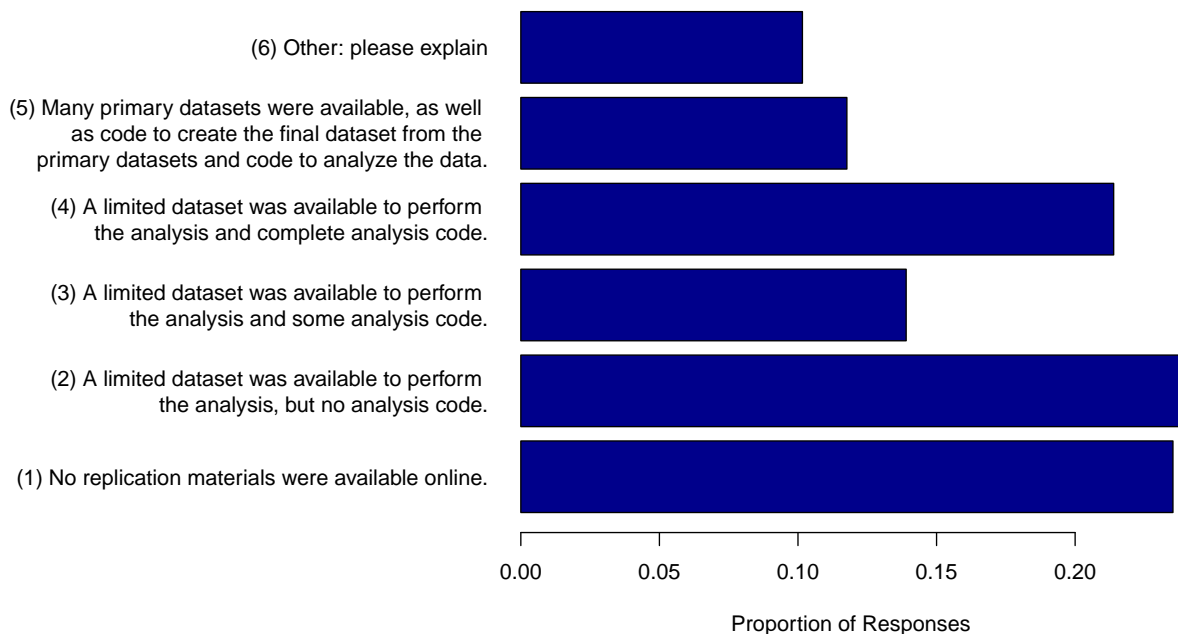


Figure 2: Availability of replication files for respondents.

Respondents were asked to “Please describe the results of your replication efforts (select all that apply).” The responses are displayed in figure 3.

Online Appendix D for:

Dafoe, Allan. 2013. Science Deserves Better: The Imperative to Share Complete Replication Files.

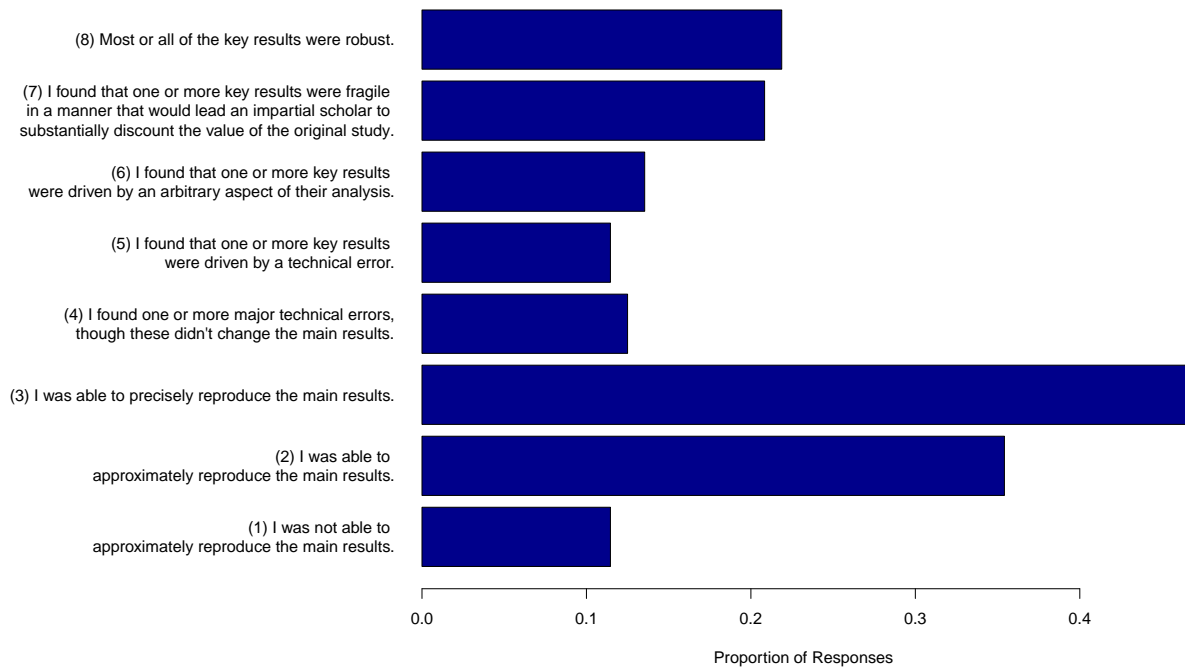


Figure 3: “Please describe the results of your replication efforts (select all that apply).”