How (Not) To Reproduce – Practical Considerations to Improve Research Transparency in Political Science

R. Michael Alvarez   Simon Heuberger*

August 25, 2021

Abstract

In recent years, researchers, journals, and professional organizations in political science have been working to improve research transparency. While better transparency is a laudable goal, the implementation of standards for reproducibility still leaves plenty to be desired. In this paper, we identify two practices political science should adopt to improve research transparency: (1) Journals need to provide detailed replication guidance and run provided material, and (2) authors need to start their work with replication in mind. We focus on problems that occur when scholars provide research materials to journals for replication and outline best practices regarding documentation and code structure for researchers to use.

*Alvarez (rma@caltech.edu) is a Professor of Political Science at Caltech, and past co-editor of *Political Analysis*. Heuberger (heuberger.simon@alumni.american.edu) recently received his Ph.D. in Political Science from American University, and has supervised the paper reproduction process currently in place at *Political Analysis* under Jeff Gill's editorship.

1 Introduction

Research transparency has become a central concern in political science (Miguel et al., 2014; Collaboration, 2012; Colaresi, 2016; Freese, 2007; Freese and Peterson, 2017; Gertler et al., 2018; Coughlin, 2017; Clemens, 2017).

Transparency greatly strengthens the quality of research, heightens accountability, and increases trust in the discipline. Data transparency concerns two related, but significantly different, goals: Using the data/code from a published paper to get the same results reported in the paper, and taking the protocol for a study to get the same results with a new or different dataset. The first refers to reproducibility, which verifies published results and code, while the latter refers to replication, which tests the validity of published findings (Shepherd et al., 2017; Plesser, 2018). We are concerned with reproducibility here.

For others to test, analyze, reproduce, and replicate findings from published results, researchers must publish their entire reproduction files (Dafoe, 2014; Eubank, 2016; Lupia and Elman, 2014). Having code and data available also makes it possible for scholars to improve on the methodology used or analyses conducted and thus further advance scientific knowledge (for general discussion see Gleditsch et al. (2003); Gherghina and Katsanidou (2013); Ishiyama (2014); Nosek et al. (2015); specific examples are in Lall (2016), and in the extended debate found in Neunhoeffer and Sternberg (2018); Wang (2018); Muchlinski et al. (2018); Heuberger (2018); Harden et al. (2018)). Archival code and data are also used pedagogically, in particular in graduate methodology courses (Janz, 2016). As Gary King wrote, "[t]he only way to understand and evaluate an empirical analysis fully is to know the exact process by which the data were generated and the analysis produced" (King, 1995, p. 444).

Indeed, the issue of research transparency has become so important that *PS: Political Science & Politics* has published seven papers in a collection called "Opening Political Science." These seven papers all advance important arguments about how political science can improve research transparency (Breznau, 2021; Engzell and Rohrer, 2020; Janz and Freese, 2020; Kapiszewski and Karcher, 2020; Lupia, 2020; Rinke and Wuttke, 2021; Rohlfing et al., 2020). But what is missing from this collection of papers is practical advice for scholars who are submitting their work to journals that have research transparency requirements for publication. What is also missing is guidance for journals how to conceptualize replication guidelines that aid successful reproduction.

Today, many journals request or require authors to submit reproduction material to data archives, which we discuss further in this paper's Supplementary Materials. Some journals also confirm that the materials authors provide to meet research transparency requirements work as expected and that they in fact reproduce the paper's reported quantitative results. Our experience has shown that disorganized and virtually unusable reproduction material is still the norm, rather than the exception, in political science. Major shortcomings include not providing basic documentation (like a README file), setting local working directories, code that does not produce saved outputs of manuscript figures and tables, code that is not commented, and the absence of estimated running times.

For example, out of the dozens of data reproduction sets submitted to *Political Analysis* over a recent 18-month period, all except one suffered from at least one of these shortcomings. The following sections demonstrate these shortcomings with anonymized data from the reproduction work done at *Political Analysis*. All code examples are written in the open-source software `R`. Our aim with these examples is to both showcase what authors can do better and to provide journals with recommendations that can inform their replication guidelines.

2 Shortcomings in Journal Reproduction Materials

In this section, we discuss the most common problems that occur during the process of reproduction review at *Political Analysis*; problems that we suspect are typical for the political science journals with research transparency requirements. We use the example of *Political Analysis* because it has one of the longest-standing policies of research transparency among journals in political science.

We also focus on *Political Analysis* because both authors have had hands-on experience with development of the journal's policies, and their implementation.[1]

Editors of journals with research transparency requirements have the goal that authors provide all of the materials necessary to reproduce (precisely) the quantitative claims made in their soon-to-be published manuscript. This is consistent with the policies on research transparency and and data access of the American Political Science Association (APSA), and of the Society for Political Methodology (SPM). To meet this goal, *Political Analysis* requests that authors provide:

1. A README file that describes the materials the author has provided for reproduction and the computing environment used for analysis.

2. Well-documented, well-named and user-friendly code that reproduces (precisely) and saves the tables and figures in the manuscript.

3. Any software packages or other materials that are necessary to reproduce the results reported in the manuscript.

4. The data needed to reproduce the results reported in the paper.

5. Good documentation that other researchers can use to understand how to run the code to obtain the results reported in the manuscript.

These are relatively simple and innocent requirements[2]. However, many if not most authors fail to meet the minimal requirements when they provide initial reproduction materials for their manuscript. By not providing the appropriate materials initially, authors often cause the journal's editorial team's efforts to use those materials to produce the results reported in the manuscript to fail. This necessitates additional communication with the authors where they revise their reproduction materials and the process iterates until a point where editors can release well-documented, well-organized, and well-behaved reproduction materials on the journal's Dataverse. Authors end up facing delays in the production process for their manuscript — because *Political Analysis* will not send a paper into production until the reproducibility requirement has been achieved. It is concerning that authors often fail to produce usable materials when required to publish their paper, as this implies that the principle of research transparency has probably not been baked into the study from its inception.

This also raises a normative question: should the author or the journal bear the costs associated with the production of research transparency materials that are portable and which a journal's research transparency team can use successfully? We argue that both should bear some costs. On the journal's side, the research transparency team (who we call the "replicators") need to have a generalist's understanding of the primary software languages in use in their field (for political science that is primarily Stata, R, and `Python`); they need to have up-to-date versions of these software languages and operating systems; and they need access to computational resources that can run most larger-scale processes in a reasonable amount of time. But we do not believe that replicators should need to undertake line-by-line code review, or to debug why some provided code only runs using outdated libraries or packages.

We argue that authors should bear most of the costs associated with the production of well-documented, usable code. In fact, we encourage authors to build into their workflows the sort of practices that will produce good code and documentation, and as more authors engage with this we believe that these issues we discuss below will largely disappear. But more importantly, writing code is for many studies a very large component of the research enterprise; as the research conducted is the author's responsibility, we do not believe that it is asking too much from authors to pay the same close attention to their code as they do to collecting accurate data, to providing appropriate citation of previous work, and to following all of the other best practices that generally govern high-quality social science research.

While it is beyond the scope of this paper to go into detail about how authors can build these workflows, we note that funding agencies now require detailed data management plans, which is oftentimes an important part of a research workflow. Second, it is increasingly common to see social scientists using version control and code collaboration tools like Git, Bitbucket, and GitHub. Finally, introductory graduate methods courses are starting to include research transparency and other ethical issues in their curriculum, training the next generation of scholars in these best practices; professional societies can help by providing short courses and other educational materials for those who want to learn how to build research transparency into their workflows.

2.1 Documentation: The Importance of the README file

To be usable by other researchers, reproduction materials require documentation, which means that all reproduction materials need to provide a simple README file. The README is the first file a user opens after downloading the data material and needs to contain all the information a user needs to run the code. While this obviously depends to some degree on the material in question, any README needs to contain five basic sections: (1) A list of all folders, subfolders, and data files contained in the material, (2) the hardware specifications used to run the code, (3) software used, including all packages or libraries (e.g. for `R` or `Python`) and their respective versions, (4) a list of all code files to produce which output used in the paper, (5) the approximate running time of each code file based on the hardware specifications. Figure 1 shows an anonymized example of an insufficient README submitted to *Political Analysis*. It misses all five sections and gives the user no information about what to expect or how to proceed with the data analysis.

```
REPLICATION MATERIAL

AUTHORS: Doe, Jane; Doe, John
ARTICLE: "Using Cool Statistical Methods"
JOURNAL: Political Analysis

## INSTRUCTIONS ##

To replicate the analysis, you have to download all files (besides this README) from the
Dataverse and run them.

## NOTE ##
If you have any further questions, please contact John (doe.john@gmail.com)
```

Figure 1: Example of insufficient README

2.2 Reproduction Data

It is oftentimes assumed for the purposes of article reproduction that authors must provide the complete original dataset — this is untrue. Often an author is working with secondary data, for example, the American National Election Survey (ANES) or the Cooperative Congressional Election Survey (CCES). In these cases, the author needs to provide documentation and code that extracts from these public data sources only the rows and columns used in the published study; that code should also provide details about all processing and manipulation that transforms the original data into the data used in the paper.

Another common problem is proprietary data, i.e. authors who do not have permission to share their data due to copyright restrictions or other legal restrictions on public dissemination of the data. In such a case, authors may provide a percentage sample of the data for reproduction purposes. This usually allows the reproduction of the main findings whilst still protecting and respecting data ownership and privacy.

Finally, a third regularly occurring problem with reproduction data is the inclusion of identifying information in the reproduction materials. A good example that we have seen is reproduction information that may contain the names and addresses from voter registration datasets, or the names and contact information provided as metadata in datasets from manual text processing studies. There are a lot of reasons why it is not good practice to provide any identifying information in reproduction data, even if it is in an otherwise public release file.

2.3 Code and Output

Code and data files should be set up in a self-contained project. In `R`, this should be an `R Project` whilst using the `here` package (Bryan, 2018). This sets the working directory to the `R Project` folder for all script files in the material. Files can then be loaded and saved with relative path files starting at the main replication folder. Local working directories with `setwd()`, as exemplified in Figure 2, do not represent a practical workflow as they only work for one user on one local machine. Self-contained project working directories, on the other hand, work on all machines without any manual user input.

```
setwd("/Users/janedoe/Google Drive/2020/papers/pa/material")
```

Figure 2: Working directories should not be set locally

It is imperative to clearly and consistently name data sources, `R` objects, and output objects. This makes code easily traceable, objects identifiable, and avoids unnecessary confusion. Figure 3 shows an example where two `.csv` source files and three `R` objects are all based on the word "data". Not only does this confuse users who are unfamiliar with the material, but it also overwrites the `R` base function `data()`, which could potentially become problematic.

```
data.rep <- read.csv(file="data.csv")
data.full <- data.rep[,c("var1","var2","var3","var4")]
data <- read.csv(file="data_full.csv")
```

Figure 3: Confusingly named files and `R` objects

The code needs to save output files for every figure and table in the main text and the appendix of the manuscript. Figures should be saved in `.pdf` format and tables in `.csv`, `.tex` or `.html`. Each figure and table should be named according to its number in the manuscript, e.g. `Figure1.pdf`, `Table3.tex`, in order to make the output clearly and easily identifiable. Crucially, the saved output needs to show identical content as the resulting respective figure or table. Table 1 shows an example of saved `R` output in `.csv` form which bears no resemblance to the corresponding manuscript table. While the information presented in the manuscript table may be part of the produced `.csv` file, it is not possible to make out this information in the data's current form.

	X	warstds	predictors.1	predictors.2
1	1	0	0.036354228	9.380534e-03
2	2	0	0.009281192	3.714911e-03
3	3	0	0.017708231	8.841568e-03
4	4	0	0.034197142	3.136079e-03
5	5	1	0.078753771	1.350169e-01
6	6	0	0.025539793	1.478751e-02
7	7	1	0.013429648	1.852500e-01
8	8	0	0.027755230	5.513324e-03
9	9	1	0.012395045	5.857035e-03
10	10	0	0.006440910	1.213052e-03
11	11	1	0.007812825	1.367884e-02
12	12	0	0.002278077	5.956352e-03
13	13	0	0.011487555	6.777899e-02

Table 1:

Civil War	Author 1	Author 2
Angola 2001	0.03	0.04
Burundi 2001	0.03	0.02
Guinea 2001	0.01	0.02
Rwanda 2001	0.05	0.00
Uganda 2002	0.02	0.04
Liberia 2003	0.00	0.02
Iraq 2004	0.03	0.02

Table 1: Code output (left), manuscript table (right)

Finally, we have repeatedly seen examples where authors of papers with simulations or methods that involve sampling or resampling fail to set random number seeds. Failing to use the same random number seed when trying to reproduce manuscript results is problematic, as the reproduction will not generate the exact results reported. Thus, authors should always set the random number seed whenever conducting simulations, or using sampling methods, and document this well in their code.

3 Looking Ahead: Developing Standards and Best Practices for Social Science Research Transparency

Improving research transparency is becoming a higher priority for political science scholars, journals, and professional associations, much work remains. In this paper, we have identified two practices social science should adopt to help resolve the crisis: (1) journals need to provide detailed replication guidance and run provided material, and (2) authors need to start their work with replication in mind.

Many journals have been working to build stronger research transparency requirements into their standards. Earlier in the evolution of these standards, the goal was simply to ensure that all manuscripts making empirical claims provided some code, data, and documentation, without paying a great deal of attention to standardization and the quality of those materials. Today, the top quantitative journals in political science all have strong research transparency requirements and require the provision of data and code prior to a paper's publication. However, a number of these journals do not provide or utilize a permanent and public archive for these materials, and very few of them at this point in time actually confirm that the data and code reproduce the claims reported in the paper. We thus urge all political science journals, to shift their focus from the mere implementation of transparency requirements towards a rigorous evaluation of the quality, executability, and user-friendliness of the research materials. A promising technological development for reproducibility is the use of Docker containers, which we discuss in the paper's Supplementary Materials; journals like *Political Analysis* are moving in this direction, for example with the use of Code Ocean.

We also encourage journals to provide detailed replication guidance to authors in order to aid reproduction efforts. These can range from elaborated bullet points to templates that showcase what is required, as we demonstrate in the Supplementary Materials. Additionally, we urge journals to establish a research transparency team where all members are sufficiently trained to efficiently run Stata, R, and `Python` code and diagnose common problems. With such training structures in place, graduate students are fully equipped to conduct journal reproductions very cost-effectively. Journals also have to make sure that their reproducers have access to computational resources that will reliably run complex simulations, machine and deep learning, and which can handle larger scale datasets.

On the author side, highly disorganized and virtually unusable reproduction material is still very much the norm. Virtually none of the reproduction data sets submitted to *Political Analysis* over the past 18 months ran without producing errors. Scholars need to pay closer attention to the documentation and usability of the research materials they make available to journals and other scholars. They should begin a quantitative study with reproducibility in mind to avoid a mad scramble to collect and document their research late in the publication process — and they can make sure that their material meets transparency requirements when provided to journals.

Notes

1. Alvarez was co-editor of *Political Analysis* between 2010 and 2018, when the journal began requiring that authors provide research transparency materials, and then began the process of validating those materials. Heuberger has been a graduate editorial assistant at the journal under the current editor, Jeff Gill, and he has been in charge of validation of research transparency materials since 2017.

2. For space reasons, the requirements are presented here in rudimentary form. Actual guidance given to authors at *Political Analysis* explains each point in detail.

References

Breznau, N. (2021). I saw you in the crowd: Credibility, reproducibility, and meta-utility. *PS: Political Science &; Politics*, 1–5.

Bryan, J. (2018). Ode to the `here` Package. https://github.com/jennybc/here_here.

Clemens, M. A. (2017). The Meaning of Failed Replications: A Review and Proposal. *Journal of Economic Surveys 31*(1), 326–42.

Colaresi, M. (2016). Preplication, Replication: A Proposal to Efficiently Upgrade Journal Replication Standards. *International Studies Perspectives 17*(4), 367–78.

Collaboration, O. S. (2012). An Open, Large-Scale, Collaborative Effort to Estimate the Reproducibility of Psychological Science. *Perspectives on Psychological Science 7*(6), 657–660.

Coughlin, S. S. (2017). Reproducing Epidemiologic Research and Ensuring Transparency. *American Journal of Epidemiology 186*(4), 393–94.

Dafoe, A. (2014). Science Deserves Better: The Imperative to Share Complete Replication Files. *PS: Political Science and Politics 47*(1), 60–66.

Engzell, P. and J. M. Rohrer (2020). Improving social science: Lessons from the open science movement. *PS: Political Science &; Politics*, 1–4.

Eubank, N. (2016). Lessons from a Decade of Replications at the Quarterly Journal of Political Science. *PS: Political Science and Politics 49*(2), 273–76.

Freese, J. (2007). Replication Standards for Quantitative Social Science: Why Not Sociology? *Sociological Methods & Research 36*(2), 153–72.

Freese, J. and D. Peterson (2017). Replication in Social Science. *Annual Review of Sociology 43*(1), 147–65.

Gertler, P., S. Galiani, and M. Romero (2018). How to Make Replication the Norm. *Nature 554*(7693), 417–419.

Gherghina, S. and A. Katsanidou (2013). Data Availability in Political Science Journals. *European Political Science 12*, 333–349.

Gleditsch, N., C. Metelits, and H. Strand (2003). Posting Your Data: Will You Be Scooped or Will You Be Famous? *International Studies Perspectives 4*(1), 9–95.

Harden, J. J., A. E. Sokhey, and H. Wilson (2018). Replications in Context: A Framework for Evaluating New Methods in Quantitative Political Science. *Political Analysis 27*, 119–125.

Heuberger, S. (2018). Insufficiences in Data Material: A Replication Analysis of Muchlinski, Siroky, He, and Kocher (2016). *Political Analysis 27*, 114–118.

Ishiyama, J. (2014). Research Transparency, and Journal Publications: Individualism, Community Models, and the Future of Replication Studies. *PS: Political Science and Politics 47*(1), 78–83.

Janz, N. (2016). Bringing the Gold Standard into the Classroom: Replication in University Teaching. *International Studies Perspectives 17*, 392–407.

Janz, N. and J. Freese (2020). Replicate others as you would like to be replicated yourself. *PS: Political Science &; Politics*, 1–4.

Kapiszewski, D. and S. Karcher (2020). Transparency in practice in qualitative research. *PS: Political Science &; Politics*, 1–7.

King, G. (1995). Replication, Replication. *PS: Political Science and Politics 28*, 444–452.

Lall, R. (2016). How Multiple Imputation Makes a Difference. *Political Analysis 24*, 414–433.

Lupia, A. (2020). Practical and ethical reasons for pursuing a more open science. *PS: Political Science &; Politics*, 1–4.

Lupia, A. and C. Elman (2014). Openness in Political Science: Data Access and Research Transparency. *PS: Political Science and Politics 47*(1), 19–42.

Miguel, E., C. Camerer, K. Casey, J. Cohen, K. Esterling, A. Gerber, R. Glennerster, D. Green, M. Humphreys, G. Imbens, D. Laitin, T. Madon, L. Nelson, B. Nosek, M. Petersen, R. Sedlmayr, J. Simmons, U. Simonsohn, and M. V. der Laan (2014). Promoting Transparency in Social Science Research. *Science 343*, 30–31.

Muchlinski, D. A., D. Siroky, J. He, and M. A. Kocher (2018). Seeing the Forest Through the Trees. *Political Analysis 27*, 111–113.

Neunhoeffer, M. and S. Sternberg (2018). How Cross-Validation Can Go Wrong and What to Do About It. *Political Analysis 27*, 101–106.

Nosek, B., G. Alter, G. C. Banks, D. Borsboom, S. D. Bowman, S. J. Breckler, S. Buck, C. D. Chambers, G. Chin, G. Christensen, M. Contestabile, A. Dafoe, E. Eich, J. Freese, R. Glennerster, D. Goroff, D. P. Green, M. H. B. Hesse, J. Ishiyama, D. Karlan, A. Kraut, A. Lupia, P. Mabry, T. A. Madon, N. Malhotra, E. Mayo-Wilson, M. McNutt, E. Miguel, E. L. Paluck, U. Simonsohn, C. Soderberg, B. A. Spellman, J. Turitto, G. VandenBos, S. Vazire, E. J. Wagenmakers, R. Wilson, and T. Yarkoni (2015). Promoting An Open Research Culture. *Science 348*, 1422–1425.

Plesser, H. E. (2018). Reproducibility vs. Replicability: A Brief History of a Confused Terminology. *Frontiers in Neuroinformatics 11*, 76.

Rinke, E. M. and A. Wuttke (2021). Open minds, open methods: Transparency and inclusion in pursuit of better scholarship. *PS: Political Science &; Politics*, 1–4.

Rohlfing, I., L. Königshofen, S. Krenzer, J. Schwalbach, and A. Bekmuratovna R. (2020). A reproduction analysis of 106 articles using qualitative comparative analysis, 2016–2018. *PS: Political Science &; Politics*, 1–5.

Shepherd, B., M. Peratikos, P. Rebeiro, S. Duda, and C. McCowan (2017). A Pragmatic Approach for Reproducible Research with Sensitive Data. *American Journal of Epidemiology 186*, 387–392.

Wang, Y. (2018). Comparing Random Forest with Logistic Regression for Predicting Class-Imbalanced Civil War Onset Data: A Comment. *Political Analysis 27*, 107–110.